PROCEEDINGS from the international meetings on

Automated Data Processing in Hospitals

in Elsinore, Denmark, april - maj 1966

CONTENTS Page

FOREWORD

By Erling Dessau, Cand. Polit., Mathematical and Statistical Department, I/S Datacentralen af 1959, Copenhagen, Denmark.

During the months of April and May 1966 a number of international meetings were held in Elsinore, Denmark, concerning medical data processing. The interest in using computers in connection with medical procedures has been steadily increasing since the beginning of the 1960'. In several ways the development has been led by medical research institutions and hospitals in U.S.A., but during the last years a vast amount of work has begun within the medical data processing in Europe and perhaps especially in England and Scandinavia.

It was on the background of this increased interest in the medical data processing subjects that a group of persons in Sweden and Denmark, led by Professor, Dr. Med. A. Tybjaerg Hansen and the undersigned, took the initiative of holding an international symposium. We were so fortunate that the NATO Science Committee sympathized with our thoughts and as part in their support to international advanced study institutes, they granted us a substantial award. Professor Tybjaerg Hansen and I, who were directors of this study institute, are very grateful for the support granted. Especially after having experienced the enormous attendance at the symposium we are happy that this could be carried into effect, and hereby we want to express our most heartfelt thanks.

At the same time as the plans for the NATO Advanced Study Institute were worked out, a co-operation was established with the International Institute for Medical Electronics and Biological Engineering in Paris, whose director, Dr. John Davis, also had worked with the plans of holding an European conference in medical data processing. Then it was proposed to hold this conference in connection with the symposium and a committee of organization got appointed with Professor Tybjaerg Hansen as chairman. It was decided that the international conference on "Automated Data Processing in Hospitals" was to be held on the 20th to the 23rd April, 1966, while the study institute was to be held on the 25th April to the 4th May.

In connection with the conference an exhibition of various data processing equipment was held from the 19th to the 27th April. The leader of this exhibition was engineer Ernst Borg.

The conference and the arrangement of the exhibition were carried throug by means of support from various parts. To be mentioned are the support from the Council of International Organizations of Medical Sciences, an organization subsidized by WHO and UNESCO, (grant proposed by the International Federation for Medical and Biological Engineering), and the extensive assistance rendered by Messrs. Birch and Krogboe, Consulting Engineers, in connection with the carrying through of the arrangement of the exhibition.

Dr. John Davis, The International Institute for Medical Electronics and Biological Engineering, is also responsible for the fact that the conference was carried through with so great success, and Mr. Gillis Claus, Director of SJURA, Stockholm, has been a very great help as well in connection with the planning of the practical arrangement of the conference as especially by placing his knowledge and his organization at disposal by publishing this publication. Finally it is of importance for me to mention Mr. Jørgen Hilden, M.D., who carried a substantial and invalueable burden as well in connection with the practical carrying through of the meetings as in connection with the editing of the conference papers.

The amount of participants in the conference exceeded the most optimisti al prognoses and inclusive invited guests and those who exhibited, about 380 from 21 countries participated in the conference, while about 100 persons participated in the symposium.

In continuation of any conference it is the very great wish of the organ izers as well as the participants to have the papers and discussions printed to be used for further studies. Unfortunately, it is so much easier to wish than to have these wishes carried through as a well-edite book. After several discussions it was decided to publish most of the papers only in a mimeograped booklet. In order to cut down the amount of pages, unfortunately, we have had to leave out that part of the papers which were not delivered as a manuscript, just as a few very long contributions have been abridged. We hope that the authors will forgive us these omissions and restrictions. Time, money and the extent have necessitated these interferences.

At the opening of the symposium I inquired about a model for the medical information system. If papers and discussions had been co-operated better, the answer to this inquiry would probably have been more clear. However, it might have been too early to have this model defined. Still, we do not know enough about the medical information to be able to lay down the interaction. Though, the impression is that the Elsinore conference has helped us very much to understand the content and the interactions in the medical information.

The medical data processing is in an epoch age. Earlier age's groping and limited experiments must now be replaced by a determined work. Doctors, nurses and the national health administration have a very great demand of the data processing technique, and it must be the job of the EDP-system people to fulfil the demand which is claimed from these so important activities, seen from a social point of view.

The Elsinore meetings clarified much for us. But they did not give us any occasion to rest on the achieved results. The manuscripts published give us a survey of our knowledge of to-day, and we are grateful to the persons who through work and intelligence have brought us so far. Now we are, however, all going to work to obtain the more perfect results.

Hereby my thanks to all who participated.

Copenhagen, August 1966.

Erling Dessau.

ED/bs.

PROGRAM I

AUTOMATED DATA PROCESSING IN HOSPITALS

An International Conference on the Interface Problems

April 20th-23rd 1966

Hotel Marienlyst and Hotel Prins Hamlet, Elsinore, Denmark.

The Theme of the Conference:

the interface problems in hospital data processing - a study of the special needs of the patient, the doctor and the hospital administrator defining the role of the computer in clinical medicine.

The Presidium:

President	Prof., Dr. Med. A. Tybjaerg Hansen, University of Copenhagen, Denmark
Vice-president	Prof., Dr. Homer Warner, University of Utah, Salt Lake City, USA
Vice-president	Dr. A. Rémond, Université de Paris, La Salpêtrière, Paris, France
General Secretary	Mr. Erling Dessau, Cand. Polit., department head "Datacentralen", Denmark
Exhibition Secretary	Mr. Ernst Borg, Mssr. "Birch & Krogboe", Copenhagen, Denmark
Assistant Secretary	Dr. Jørgen Hilden, research fellow, University of Copenhagen, Denmark

Organization Committee:

Dr. A. Tybjaerg Hansen	Professor of Internal Medicine, Rigshospitalet, Copenhagen Ø, Denmark
Mr. E. Dessau	Head of Mathematical and Statistical Department, Datacentralen, Copenhagen S. Denmark
Dr. J. Mosbech	Medical Consultant, Statistical Section, National Health Service of Denmark, Copenhagen
Dr. P. E. Andersen	Senior Physician, Odense County and City Hospital, Odense, Denmark
Dr. V. Oram	Head of Surgical Department, Århus Municipality Hospital, Denmark
Mr. Åke Pernelid	Director-in-Chief, National Swedish Rationalization Agency, Stockholm, Sweden
Mr. Gillis Claus	Director, SJURA, Stockholm, Sweden
Dr. P. Hall	Senior Physician, Karolinska Sjukhuset, Stockholm, Sweden
Dr. T. Hauen	Senior Physician, National Board of Health, Oslo, Norway
Dr. J. F. Davis	Director, International Institute for Medical Electronics and Biological Engineering, Paris, France

<u>Wednesday 20th April</u>

Opening of the conference
The president, Prof. A. Tybjaerg Hansen.
Opening address by the Danish minister of education: Mr. K. B. Andersen.

The role of computers in the hospital - the interface problems in medical and administrative patient management.
Chairman: E. Dessau

A. Tybjaerg Hansen
Professor, Rigshospitalet, Univ. of Copenhagen

Aksel Marchmann
Vice-director, Copenhagen County Hospital Administration.

On the Requirements for Information Systems in Hospitals
Charles D. Flagle
Professor, Johns Hopkins Univ., Baltimore, USA.

Time-Sharing a Computer for Processing Patient Data direct from Transducers at the Bedside
Homer Warner
M.D., Ph. D., Univ. of Utah, Salt Lake City.

The Role of Computers in the Hospital
Lee B. Lusted
Professor of radiology, Univ. of Oregon, Portla

The Computer in the National Health Service
Henry Yellowlees
M.D., Ministry of Health, London

<u>Thusday 21st April</u>

Analog-Digital Data Processing; Hybrid Systems.
Chairman: Professor H. Warner

Antoine Rémond
M. D., La Salpetrière, Paris

A Digital Computer for the Management and Study of the Critically Ill
Max H. Weil
M.D., Ph. D., Univ. of Southern California, Los Angeles

Hybrid Approaches to a Hybrid World
Josiah Macy jr.
Professor, Yeshiva University, New York.

Analyses of ECG and EEG.

Spectral Analysis Techniques and Pattern Recognition Methods for Electroencephalographic Data

W. Ross Adey
Professor, Univ. of California, Los Angeles

Computer Analysis of Medical Signals in Hospital Practice
C. A. Caceres
M.D., Public Health Service, Washington, D.C.

Computer Classifications of Electrocardiograms
Hubert Pipberger
Professor, V. A. Hospital and Georgetown Univ., Washington, D.C.

Friday 22nd April
Chairman: Dr. A. Rémond

A Computer System for Automatic Processing of Hospital Laboratory Services
William Kirkham
M.D., National Institute of Health, Bethesda

Data Processing in the Clinical Laboratory
Gunnar Jungner
M.D., Univ. of Gothenburg

Computer Assisted Radiation Treatment Planning
J. van de Geijn
M.D., H. Joannes de Deo Hospital, The Hague

Some Potentialities of the Computer in the National Health Service
E.D. Acheson
M.D., Univ. of Oxford

Computer Diagnosis of Congenital Heart Disease and Brain Tumors
Kosei Takahshi
M.D., Univ. of Tokio

Man-Machine Communications in a Hospital
G. Octo Barnett
M. D., Massachusetts General Hospital, Boston

The Information Problem in Medicine
G. Wagner
Professor, M.D., Deutsches Krebsforschungszentr
Heidelberg

Saturday 23rd April
Chairman: Professor A. Tybjaerg Hansen

Computer Aid for Hospital Planning
James Sounder
M.D., Bolt Beranek and Newman Inc., Van Nuys

Input and Output Hardware
J. F. Davis
M.D., International Institute for Medical Electronics and Biological Engineering, Paris

The Feasibility of Automated Data Processing in Hospitals
Erling Dessau
Cand. Polit. Mathematical and Statistical Dept. Datacentralen, Copenhagen

The Computer in our Modern Hospitals
Paul Hall
M.D., Karolinska Sjukhuset, Stookholm

Closing of the conference.

PROGRAM II

INTERNATIONAL ADVANCED SYMPOSIUM ON DATA PROCESSING IN MEDICINE

April 25th - May 3rd 1966

Hotel Prins Hamlet, Elsinore, Denmark.

Themes:

Computer Assisted of Bio-medical Information

Cybernetics in Medicine

Computers and Patient Management - The Use of Operations Research Methods

Analog Information

Man-Machine Communication in the Hospital

Education through Computers

Instruction in ADP for Hospital Personnel.

Organized by: Professor, Dr. Med. A. Tybjaerg Hansen
Rigshospitalet, University Hospital of Copenhagen

Cand. Polit. Erling Dessau
Head of Mathematical and Statistical Department
Datacentralen, Copenhagen

Subsidized by: NATO Science Committee, Paris.

Monday 25th April
Chairman: Dr. Antoine Rémond, La Salpêtrière, Paris

Information Processing in Medicine

a. What are we aiming at by the Information Processing
Procedures: A Statement
Cand. Polit. Erling Dessau
Datacentralen, Copenhagen

b. Processing and Evaluation of Hospital Laboratory Data
Professor Donald A.B. Lindberg
University of Missouri, Colombia, Mo.

c. The Digital Computer as a Tool for Analysis o
Physiological Systems
Professor Homer R. Warner
University of Utah, Salt Lake City, Uta

d. A Clinician's Wishes and Expectations of the
Automated Information Processing
Professor A. Tybjaerg Hansen
University of Copenhagen, Denmark

Tuesday 26th April
Chairman: Professor C. Caceres, Public Health Service, Washington, D.C.

1. Hybrid Systems - Pattern Recognition

a. Professor Josiah Macy, Jr.
Yeshiva University New York

b. Professor D.H. Bekkering
Medisch-Fysisch Institut, Utrect, Holla

2. Intensive Observations and Care Units

a. Dr. Erik Sandøe
University Hospital of Copenhagen

b. Professor Max Harry Weil
University of Southern California, Los Angeles

Discussions: Dr. Mogens Jørgensen, Copenhagen County Hospita
Professor Wesley Clark, Washington University, St. Louis, Missouri

Wednesday 27th April
Chairman: Professor A. Tybjaerg Hansen, University of Copenhagen

Analog Data Acquisition and Processing

1. Automated Analysis of ECG and EEG

a. Professor Hubert V. Pipberger
Georgetown University Hospital, Washington, D.C.

b. Docent O. Arvedson
University of Umeå, Sweden

c. Docent Ingemar Petersén
Sahlgrenska Sjukhuset, Gothenburg

Engineer E. Kaiser
Kaiser's Laboratory, Copenhagen

d. Dr. Chr. Guld
University of Copenhagen Neurophysiological Institute

2. Pattern Recognition

Professor Lawrence Stark
University of Chicago, Illinois

3. e. Demonstration of On-line time sharing;
Storage retrieval and updating of large Data bases (Santa Monica - Elsinore)
by Dr. Anne Summerfield, SDC

<u>Tuesday 28th April</u>

Chairman: Dr. Illkka Väänänen, Helsinki, Finland

1. The Patient Record

a. Dr. Paul Hall
Karolinska Sjukhuset, Stockholm

b. Dr. Julius Korein and Dr. Leo Tick
New York University Medical Center, New York

c. Dr. A. E. Bennett
St. Thomas's Hospital, London

d. Dr. Johs. Mosbech
National Health Service, Copenhagen

e. Engineer Werner Schneider
Uppsala Data Center, Uppsala, Sweden

f. Dr. P. Dragsted
Hjørring Centralsygehus, Denmark

2. Cybernetics in Medicine

a. Professor Aldo Masturzo
International Society of Cybernetic Medicine, Naples

b. Dr. Joseph Wartak
Computation Centre of Polish Academy of Sciences, Warsew

c. Dr. Kosei Takahashi
University of Tokyo, Japan

3. Computers in Medical Teaching

Professor J.C. Pages and
Professor Grening
University of Paris, France

<u>Friday 29th April</u>

Chairman: Professor Arne Jensen, Technical University of Denmark

1. Operations Research - A Survey of Methods and Procedures

Docent Tore Dalenius
University of Stockholm

2. Operations Research in Medical Work

Docent Tore Dalenius
University of Stockholm

3. The OR-unit in a Hospital

Professor Charles D. Flagle
Johns Hopkins University, Baltimore, MD.

4. Operational Research and the National Health Service

Mr. J. B. Cornish
Ministry of Health, London

5. Cybernetics in Medicine

Professor Lee B. Lusted
University of Oregon Medical School,
Portland, Oreg.

Saturday 30th April
Chairman: Cand. Polit. Erling Dessau, Datacentralen, Copenhagen

1. Models in Medicine

Dr. Julian Bigelow
Institute of Advanced Study
Princeton, New Jersey

2. How to manage Health Care! - A Panel Discussion on Systems Work, Operations Research, and Cybernetics in Health Care Procedures

The Panel: Chairman: Prof. Arne Jensen, Copenhagen
Professor Charles D. Flagle, Baltimore
Mr. J.B. Cornish, London
Professor Florian Serbanescu, Pisa
Docent Tore Dalenius, Stockholm
Vice-Director Aksel Marchmann, Copenhagen
Head Physician Paul Hall, Stockholm
Dr. Poul Marke, Copenhagen
Head Physician Gunnar Jungner, Gotenburg

Monday 2nd May
Chairman: Vice-Director Aksel Marchmann, Copenhagen, Hospital Board

1. Planning for an Extensive Automated Medical Information System

a. Mr. Armando Lassus
Puerto Rico Medical Center, San Juan

b. Mr. Jordan J. Baruch
Bolt Beranek and Newman Inc., Cambridge, Mass.

c. Mr. Stephen J. Siegel
State University of New York

d. Mr. P.J. Budd
Veterans Administration, Dept, of Data Management, Washington, D.C.

e. Mr. R. Edwin Hawkins
El Camino Hospital, Mountain View, California

2. Discussions

Chairman: Erling Dessau

3. Learning by Computers

Professor J.C. Pages
IBM France, Paris

Tuesday 3rd May
Chairman: Professor Rud. Keiding, Århus County Hospital

1. Automated Data Processing in the Clinical Laboratory

a. Dr. W. Kirkham
National Institutes of Health, Bethesda, Md.

b. Dr. J. Guigan
The Guigan Laboratories, Paris

c. Dr. Gunnar Jungner
University of Gothenburg

2. Bio-medical Telemetry

Professor C. Caceres
Public Health Service, Washington, D.C.

Chairman: Dr. Paul Hall, Karolinska Sjukhuset, Stockholm

1. Systems Work and Data Bases

a. Vice-Director Aksel Marchmann
Copenhagen County Hospital Board

b. Dr. Anne Summerfield
System Development Corporation, Santa Monica, California

2. Summary

a. Cand. Polit. Erling Dessau
Datacentralen, Copenhagen

b. Professor A. Tybjaerg Hansen
University Hospital, Copenhagen

3. Closing of the Symposium.

THE ROLE OF COMPUTERS IN THE HOSPITAL - THE INTERFACE PROBLEMS IN MEDICAL AND ADMINISTRATIVE PATIENT MANAGEMENT

By A. Tybjaerg Hansen, M.D. Professor, University of Copenhagen, Denmark

Although the computer has been underway for more than a 100 years, it has been quickly developing only the last twenty years through the subtle ineraction between developing technology and developing demand for the services that the computer has to offer.

The automated data processing has already turned out to be of immense practical significance in banking, business, administration, process control in factories and in big science as atomic research, astronomy, space research etc.

In hospitals and in the health services the new technology has come rather late probably because its capabilities outside the field of pure calculations have not been fully realized.

Now when it has become apparent that the health services and particularly the hospitals may have some of their most pressing problems solved by applying computer systems a rapid development is to be expected.

As an introduction to our transactions I shall try to sketch the outlines of the theme we are to deal with. I shall do it as a medical doctor i.e. in this context as an informed layman - so I hope - reminding you that it hinges on the doctor's attitude and depends much on his understanding of the essential problems whether automated data processing will be unduly delayed or for that matter unduly hastily introduced in the hospitals and in the health system. That the computer eventually will find its proper place I do not doubt.

Actually I think nobody does - only one can look at it as a pessimist viewing all the real and all the imaginary dangers or as an optimist only perceiving the real bright sides and in addition some fata morganas.

INTERFACE PROBLEMS IN MEDICAL AND ADMINISTRATIVE PATIENT MANAGEMENT

By Henry Yellowlees, Dr. Senior Principal Medical Officer, Ministry of Health, London, England.

It is my purpose in this paper to examine the role of computers in hospitals from the point of view of a central administrative authority responsible for the Public Health in its widest sense and inevitably what I have to say is concerned primarily with the problems which have arisen in the British National Health Service. However, as my main theme is the need for more relevant and more accurate information for medical and administrative purposes at local level, as well as at central level, and as a large part of the information required has originated in individual hospitals I believe that our experiences and our plans in the United Kingdom will be of interest to this international conference.

Prof. Flagle discussed the long term planning needs in terms of "surviva in the institution", and whereas he gave due prominence to the needs for this institution to fit it to the community background which it served, one had the feeling that the accent was here on the individual hospital, and of course in the central department in a nationalized health service the institution to me is a national health service, and it might well that, as a result of data which I have centrally, I would wish to abolis entirely one particular individual institution. Now again quoting Prof. Flagle: he mentioned a list of points under his long term planning section. He wanted to know about population projections. He needed to know about morbidity, to have some information on future technico-logical advances, on the social environment in which his hospital would have to operate, to know something about economy and something about the trends in medicine. Now, for us in the British Health Service, the population projections we have, and we have these broken down on an area basis whi will fit the areas of the hospitals in which we are interested. We can get advice on some technological advances, and we know in general the

NB Special charts 1 and 2 at pages 46- 47

kind of social environment in which our service must grow. What we do not know, and what we are very anxious to know, is the true picture of morbidity in the community as a whole; this involves of course not only the position of patients who have been admitted to hospital but also of those who have not. Personally I think that the morbidity state in a country as a whole is the only way in which you can satisfactorily estimate the true demand for medical care. Therefore, we are very intereste in morbidity, and we have to collect information to find out what is going on.

Secondly a word on the economy. Our hospitals of course are not busines concerns in the way that they have to provide their own finances and ca decide that they will increase the amount of money they have by raising their charges or by getting extra sources of income. They must of cours have detailed costing figures available so that they can have some idea of what is going on, but they are not able to vary their own income which is allocated to them. Therefore, they have to plan, I imagine, in a way which must be really quite radically different from the way in which hospitals in the United States plan.

Well, I have said that I was speaking from the central position, and th objectives to the National Health Service in Britain are that no citize should be in need of medical care and unable to get it because of lack of money, and secondly that adequate medical treatment is available to all who need it. But the tax payer pays, and it is very unlikely that the proportion of the gross national product allocated to the health services will greatly increase; therefore, our hospital service knows pretty well the way it has to go for a long way ahead and that it has fit within its budget which it cannot alter itself.

What kind of information do we need for management centrally? We need increase management efficiency to get better value for money which will enable us to use the money saved, elsewhere in the service. We need to increase doctor efficiency. Doctors must undertake and are increasingl undertaking a management function of the hospital; it is not just administration, it is part of the content of the doctor's work. We must, therefore, increase doctor efficiency by enabling them to retrieve information and having recorded their knowledge and experience to retrie that knowledge, so that they can improve their clinical skill. We must

HEALTH SCREENING

By Gunnar Jungner, M. Dr. Associate Professor, University of Gothenburg.

There are many different opinions about health screening, about its medical value, how it should be performed, the aim and goal. However, one thing is true and that is what Erling Dessau once claimed in a discussion: No health screening without a data processing machine.

1. What I am going to talk about is, in fact, the task for a computer in so-called health screening.

1.0 Definition

It is then necessary to define what is meant by health screening.

Screening is an attempt to find by limited resources what by conventional means would have been impractical, costly and personnel-demanding. As Breslow (1959) states: "Screening is the presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. A screening test sorts out apparently well persons who probably have a particular disease from persons who probably do not have the disease. It is not intended to be diagnostic ...".

Before we are going into the practice and discuss some experiences I suggest that we try to get to an agreement, at least temporary for this discussion, about a few things.

1.1 First, about identification of a person. Let us assume that this problem is solved and that there is standard mode of identification. But, I would like to stress a few points and I assume that this identification contains at least these things: birth date, possibly in or together with some number (3,4 or 5 digits) which at least give sex and the region. We should also have the name with surnames and initials, the address and area code and telephone number, but also occupation, title, relatives, community code, some educational code and possibly social-medical code expressing certain special features and medical notes of the person. This means a complicated information. It is obvious that in health screening we have some pertinent wishes, more than for instance is generally assumed for hospital wards. The reason is that we need an identification and a characterization much more and in many other medical fields.

1.2 Another important detail that I would like to assume that there is an agreement on is the medical records. As you know this assumption is not at all true but nevertheless I feel that it is important - in order to make this discussion short but fruitful - to have that problem with following assumptions. The medical record is in principle the complete amount of medically important facts, from the birth to the actual time. This means necessarily that the medical records contains many different sections and that the amount of information is unwieldy and impossible to supervise and overlook. Let us assume that all practical difficulties are over and that we can get the details back when needed, we then can use a summary of the most pertinent data from each section of the medical record and that the gathered summaries reflect the medical experience for the person in question.

1.3 The importance of having an idea of the wards is shown by a drawing where the life time of an individual is represented. We then find that there are quite a few types of medical information. At birth - the only occasion when it is customary to be healthy and yet be in a hospital - health and sick care meet. The first identification comes in and also the first medical findings. Later we have for instance school examinations, we may have examinations during military service, at taking of life insurances as well as pre-employment examinations, when going abroad etc. Already before so-called health examinations are undertaken we have

---

NB References at pages 60 - 61
Diagrams and special figures at pages 62 - 71

SOME POTENTIALITIES OF THE COMPUTER IN NATIONAL HEALTH SERVICES

By E.D. Acheson, Dr, Director, Oxford Record Linkage Study, University of Oxford.

In most communities the health services include a number of hospitals, agenoies, clinics, and authorities which may be responsible for differ- ent aspects of the same person's health, either concurrently of in succession. In this paper I will discuss the problem of communication of data about patients between such agencies and will refer specifical- ly to the use of modern data processing methods in this context.

Let us examine three common-place situations in which it is necessary to communicate data between health agencies. First, there is the situation where information is required in hospital B about a patient who has previously been admitted to hospital A. Under these circum- stances hospital A has to write a letter or telephone the other hospital, giving full particulars of identification about the patient. In hospital B a search of an index of names will have to be instituted to find the appropriate records, which are usually filed under a number specific to the hospital concerned. It is important to point out that this clerical barrier to communication between the hospitals will remain even if the two hospitals concerned have automated medical records systems unless they share a common system for the identification of patients. A different situation arises where morbidity data are required for the whole community. Let us suppose for example that it is necessary to obtain information about every case of bronchogenic carcinoma for the whole community. One might suppose it was sufficient to ask each of the agencies to report the number of cases and that the incidence could be determined simply by adding these up. Unfortunately such a method would result in a substantial overestimate, because during the course of their illness such patients may come into contact with several different hospitals or clinics. The usual solution to the problem (as in Denmark)

NB References at page 78

is to set up a community cancer register with its own master index of names. Each time a case is reported to it, the master index is searched to see whether this case has already been reported to another agency. Once again the hospital or clinic which reports the case must provide full identifying data because its own filing number will be meaningless to the cancer register.

There is a third situation which is even more common than the other two. This is where it is necessary to send data about a patient from a hos- pital or clinic to the doctor who referred the case. Again, in a com- puter system which is based on a single hospital it is likely that the data will be filed on magnetic tape under the personal number used in that particular hospital. To the general practitioner in the field this number is meaningless, and again unless the full indentification data and the address of the patient are recorded on tape with the other data, a manual clerical operation will have to take place to communicate in- formation from the hospital system to the family doctor.

In Oxford, a community of 100,000 persons, I know of twenty-four manual indices of names, each dealing with a different aspect of the health of the same people. Each of these creates a barrier to communication, and each requires a separate clerical effort to be maintained up to date. Furthermore each new procedure which is developed in the future brings the threat of another isolated system of medical data (data isolate) which will require a manual or clerical procedure to effect communica- tion with the remainder of the medical data in that community.

RECORD LINKAGE

Record linkage, a term which was introduced by H.L.Dunn in 1946, is concerned with breaking down such barriers of communication. Dunn described record linkage as follows: "Each person in the world creates a book of life. This book starts with birth and ends with death. Its pages are made of the principal events in life. Record linkage is the name given to the process of assembling the pages of this book into a volume". Later H. Newcombe in Canada extended this concept from the person to the family. Thus personal record linkage is concerned with bringing together the successive items of data about an individual; and family record linkage is concerned with assembling the cumulative

Many of the advantages of record linkage are realized only when the system becomes national. Contemporary western societies are so highly mobile that a follow-up system functioning on a regional or local basis will have limited value. The arrival of the electronic computer has made it reasonable to envisage and to plan for national systems of integrated medical data, accessible to medical men in the service of the patient and for research.

REFERENCES

ACHESON, E.D. (1964), Brit. J. Prev. Soc. Med., 18, (1), 8-13.

DUNN, H.L. (1946), Amer. J. Publ. Hlth, 36, 1412.

HOBBS, M.S.T. & ACHESON, E.D. (1966), The Lancet, i, 462-3.

NEWCOMBE, H.B., KENNEDY, J.M., AXFORD, S.J. & JAMES, A.P. (1959), Science, 130, 954.

NEWCOMBE, H.B., & KENNEDY, J.M., (1962) Comm. of the A.C.M., 5, 563-566.

NEWCOMBE, H.B. & RHYNAS, P.O.W. (1962) Proceedings of the Seminar on the Use of Vital and Health Statistics for Genetic and Radiation Studies, U.N., New York. 135-154.

NEWCOMBE, H.B. (1966), Brit. J. Prev. Soc. Med., 20 (2), 49-57.

COMPUTER PROCESSING OF MEDICAL DATA IN MINIMALLY STRUCTURED NATURAL LANGUAGE

By Julius Korein, M.D., and Leo J. Tick, Ph.D., Department of Psychiatry and Neurology, New York University Medical Center, and the Data Processing and Computation Laboratory of the School of Engineering and Science of New York University.

INTRODUCTION

The primary goal of the research reported was to develop a technique for processing narrative (natural language) medical data by computer for the purpose of analysis, storage and retrieval of these data for medical record keeping and patient care, and clinical research. The present report deals with a statement of the problems and the methods used in attempting to resolve them. Finally, some results will be presented with detailed examples and suggestions as to the direction of future work will be indicated.

DISCUSSION OF THE PROBLEMS

The requisites of medical narrative data processing may be considered under two headings. One, is that of the input and its relationship to the physician utilizing the technique. The other, is the possible types of analyses that are of significance to the physician.

Supported in part by grants from the Health Research Council of the City of New York (U-1329 and I-470) and the Neurological Research Fund (/-/8-0164-708) of New York University Medical Center, and the National Institutes of Health.

NB References at pages 85 - 86
Legends for Figures and Tables at page 87
Figures at pages 88 - 92
Tables at pages 93 - 95

QUALITY CONTROL IN MEDICINE

By , Professor, Dr., Deutsches Krebsforschungszentrum, Heidelberg, Germany.

Although we all know very well that nobody can work without making errors (remember the latin proverb "errare humanum est"!), in our daily practice we tend to forget this knowledge, and admit, e.g., results from the laboratory as exact, as long as they do not exceed a certain range of expectance. But a long experience in controlling data derived from the bedside, from clinical examination, or from the laboratory has shown us, that these data are much less exact and reliable than doctors generally believe.

Considering the fact that the accuracy and reliability of clinical data is varying within a wide range, Pipberger and Freis (4) in 1960 proposed to distinguish between "hard", "relatively hard", and "soft" data. This terminology, originally borrowed from the jargon of engineers, has since then been widely accepted also in the medical documentation and data-processing literature. Among the most dubious data that we have to handle are, without any doubt, the anamnestio statements of our patients, especially those refering to dates.

A most instructive example for the unreliability of patients' statements has been published by Lilienfeld and Graham (3). These authors asked 19 consecutive male patients if they were circumcised or not and checked the answers with the findings of the following clinical examination: As Fig. 1 shows, there were no less than 66 (=34.4%) wrong answers.

Trying to get an idea of the reliability of the measurements of the patients' height, as laid down in the records of our clinic, we compared these data in the first and second record of 575 adult patients, having

NB References at page 99
Figures at pages 100 - 102

been admitted twice within 3 years (8). Fig. 2 shows the result: the differences between the two "measurements" represent a broadly scattered normal distribution with only a few extreme values, most of which turned out to be slips of the pen. Much more serious is the unexpectedly high peak in the middle of the curve, representing the rate of those patients in which the figures of both statements were the same. As we could detect, in most of these cases (nearly 30% of the whole collective) there had been ne second measurement at all; the figures had merely been copied from the older record. The unreliability of these data is therefore not only due to errors in measurement but is also charged with the problem of falsification.

Fig. 3. showing the distribution of 300 automatic measurements of the radium-control-time of a dosimeter, proves that also machines can make errors. In this case it is the acting in unison by the different elements of the construction that is responsible for the deviations from the true value (5).

Perhaps even higher than the error-rate in medical findings and measurements is that concerning the judgements of patients and doctors. In a study mentioned by Gruber (2) 11 patients got two different placebos, but were told that the red tablets were a strong hypnotic while the white ones were placebos. Fig. 4 gives the result of the trial: 3 patients, influenced by the words of the doctor, had slept much better after medication of the red tablets. Three others, apparently opposite-minded, had experienced quite the contrary; the white tablets had been much better than the red ones. Only 5 of the 11 patients (=45.5%) had felt no difference between the two placebos.

The problem of observer error resp. observer variation is so well known that I may restrict here to only one example, taken from a study of Groth-Petersen, Lövgreen, and Thillemann (1). The Danish authors compared the judgements of 3 experienced radiologists in reading 2.200 photofluorograms of the lung (see Fig. 5). With regard to sequelae of pleurisy only 14.3%, as to pulmonary findings without calcification even only 6.3% of the expert opinions were in accordance. Similar experiences had been made by Yerushalmi and his group of Californian radiologists (9).

The frequency distributions of data-collectives can occasionally be

Agreement and Disagreement on Different Diagnoses.

Readers	Sequelae of pleurisy no.	Sequelae of pleurisy per cent	Pulmonary findings without calcifications no.	Pulmonary findings without calcifications per cent	Hilar calcifications no.	Hilar calcifications per cent	Pulmonary calcifications no.	Pulmonary calcifications per cent
A + B + C	7	14.3	13	6.3	3	9.4	15	22.7
A + B	2		7		4		14	
A + C	1	16.3	23	15.1	2	18.8	2	30.3
B + C	5		1		0		4	
A	24		59		6		9	
B	2	69.4	12	78.6	14	71.8	10	47.0
C	8		90		3		12	
Total	49		205		32		66	

Figure 5

Frequency distribution of the end-ciphers 1-0 in leucocyte-counts

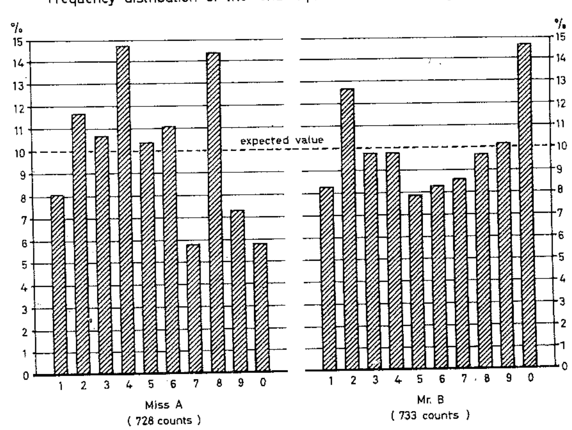


Figure 6

A TIME-SHARING COMPUTER SYSTEM FOR PATIENT CARE ACTIVITIES

By G. Octo Barnett, M.D.+, Director, Laboratory of Computer Science, Massachusetts General Hospital, Boston, Massachusetts and Paul A. Castleman, Project Director, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

The digital computer has been used for a number of years in the accounting departments of major hospitals; more recently it has found increasing use in the medical research laboratory. For the most part, these research uses have been either the statistical analysis of large amounts of data, using batch processing techniques, or the highly specialized processing of analog signals.[1,2] However, increasing attention is now being paid to the possibility of using computers to improve patient care by expediting the communication and the information storage and retrieval procedures that form an essential part of medical practice.[3,4,5,6,7,8]

The collection, storage, analysis and retrieval of a variety of types of information are of great importance in providing good patient care. The Medical Records Room at the Massachusetts General Hospital (a 1,000 bed, acute care hospital) now stores over 1,600,000 records dating from 1821. Approximately 50,000 new patient records are started each year. A patient record consists of many different forms of written data, such as the patient's medical history and physical examination, reports of laboratory tests, reports of medications and treatments given, reports of operations, diagnoses, progress, etc. Some of the record is "hard" densely coded information, but most of it is "soft" medical prose. The number of transactions that occurs in a major hospital during the course of a day's activities is overwhelming. The various laboratories in a single major hospital perform over two thousand tests each day. It is estimated that at least 7,000 different doctor's orders are written daily and that the nursing service administers over 60,000 drugs and treatments. It is probable that at least 100,000 separate

+) Established Investigator, American Heart Association

NB Bibliography at page 111. Figures at pages 112 -115

items of information are entered into the patient records each day.

At present, almost all of the data collection and storage is done by writing on slips of paper. Most of the data transmission is done by carrying carbon copies of these slips of paper from one area of the hospital to another. Almost all of the data retrieval is effected by laboriously examining vast numbers of records containing innumerable different types of information (often in vague language or illegible script).

In order to explore the feasibility of automating some of these information processing functions, an experimental timesharing, remote-access computer system is being developed.[9,10] A considerable portion of the research effort is concerned with the definition of the information processing needs involved in patient care, the exploration of how best to deal with the data which should be stored in the patient's medical record, and the development of a computer system that has terminals and methods of communication which are optimal for use by the hospital personnel. The research project is carried out by the Laboratory of Computer Science of the Massachusetts General Hospital and the Information Technology Division of Bolt Beranek and Newman Inc. The hardware configuration of the computer system includes a modified DEC PDP-1 computer, a sixty million character Univac Fastrand drum, two magnetic tape units, and a small four hundred thousand character swapping drum. Half of the 24,000 word computer core is occupied by the Time-Sharing Executive routine. The remainder of the core contains special-purpose interpreters, subroutines, and 4,000 words for the running user program itself. The small drum is used for program storage while multi-programming; the large drum and magnetic tapes are used for bulk storage of data and library programs. Over one hundred Teletypes serve as input-output devices; at any one time, sixty-four of these may be active. (See Figure 1)

The research project has four main areas of concern: a) the definition of the hospital functions under consideration, b) the development of programs that can be used by non-technical individuals, c) the implementation of an interpretive communication system using a time-shared computer and d) the evaluation of the acceptability and reliability of the total system in actual hospital operation. The remainder of the

paper will be concerned with a detailed discussion of these areas.

A. DEFINITION OF THE HOSPITAL FUNCTIONS

The difinition of the existent hospital information system requires considerable effort and is complicated by the fact that most of the present techniques evolved in a relatively unplanned fashion over many decades. The hospital operation is characterized by a wide diversity in the dypes of doctor's orders, by the large number of individuals involved in implementing these orders, and by the dynamic pace of patient care procedures. Even a broad classification of the variety of doctor's orders would include over twenty different groups such as medication, activity, dietary, physiotherapy, consultation, laboratory test requests, nursing therapy, etc. In each group, there can be as many as several thousand unique orders. The implementation of these orders requires the coordinated activity of a number of professional and nonprofessional individuals located in various and, frequently, widely scattered parts of the hospital. The implementation also involves a number of different functional units such as the diet kitchen, the various chemical and clinical laboratories, the x-ray department, housekeeping, central supply, blood bank, the surgical operation rooms, etc.

Any computer system which is to be effective as a communication system must be capable of rapid, continuous and reliable operation because of the nature of patient care activities. A medication for a given patient may be ordered while the patient is on one ward, administerd while the patient is in the recovery room, and charted (recorded in the patient's medical record as having been administered) when the patient is on another care unit. In the Massachusetts General Hospital, over 500 patients may be in transit during a typical day.

We have felt that the system under development should simulate closely the present form of hospital operation, rather than superimpose a new and artificial form of operation. For example, we have retained the somewhat cumbersome but familiar method of patient identification by a seven digit unit number, rather than assigning a smaller but artificial number. The exception to this rule is that whenever possible, advantage is taken of the total data base in computer memory to carry out some operations which are impossible without a computer system. For example, the hospital may wish to specify maximal limits for the size of dose of a given drug. The system under study would permit the maximal limit to be specified directly to the computer system. If this limit is exceeded when the drug is ordered through the Medication Order Program, an appropriate comment is returned. (See Figure 2)

Medical practioe and hospital administration are by nature conservative; both are relatively resistant to modification from within and highly resistant to any change imposed from without. The description of information processing problems in the hospital is not a function that can be completely assumed by an outside consultant. The many facets and implications of why procedures are done in a certain way can only be understood by one who has had extensive experience observing and participating in actual hospital operation. For example, it would be erroneous to make the simplifying assumption that the medical administration could specify a reasonable absolute maximal **dose** for a drug. In all cases, the physician must be allowed the final decision in the choice of therapy; the computer system merely reminds the one using the Medication Order Program of the usual upper limits of the dose of the drug. Any prescription that exceeds these limits must be entered in a slightly modified fashion. (See Figure 2)

B. DEVELOPMENT OF PROGRAMS FOR THE COMPUTER SYSTEM

A computer system which is to be used on the patient care units in a hospital requires careful human engineering, since it must be useful and acceptable to individuals who are primarily concerned with patient care and have little knowledge of computers. Because the system is designed for use by such individuals the programs appear to the user in a "conversational mode", wherein the oomputer types a question and the user enters a response. All user programs provide abundant checking of both semantic and syntactic errors, and also provide error messages, facilities for correcting errors easily, and rapid verification of encoded entries. If in doubt about how to respond, the user may inquire "HOW" to receive a brief statement of the required syntax.

Figure 3 illustrates a few of these techniques. (For illustrative purposes, the information entered by the user is underlined.) Question number one is first asked in an abbreviated or short form. The user then

entered a special code which caused the computer program to ask all following questions in a long, or unabbreviated, mode. In question four, the user typed "HOW" in order to obtain further information about the required syntax. In the first response to question four, the user made an error in typing, and the computer rejected the answer and repeated the question. The user then entered a unit number which was rejected by the computer system because that particular number had previously been assigned to another patient. In question eleven, the user corrected a spelling error using a speoial key which erases the last character entered.

The basic procedure for using the system is as follows.
From a teletype terminal, a user calls the desired program from the library of programs stored in the Fastrand drum. From then on, the user carries on a diaglogue with the computer, with the user-program acting as intermidiary. The computer system is highly problem-oriented in that each user program is concerned with a definite function. As is illustrated in Figure 3, the interaction with the user is predesigned into the program through the use of specific questions to control the input of information.

C. AN INTERPRETIVE COMMUNICATION SYSTEM

In a communication system the basic unit of information is a message. In designing an interpretive communication system, we have choosen to think of messages as the whole range of information entered by any hospital user, whether it be addressed to a hospital location, to storage files, or to the message processor itself. An interpretive communication system is a system in which the routing of any message or modification thereof is not uniquely defined by a set of rules, but is in part determined by the contents of the message and by the state of the data base (including those standing orders that are from time to time entered as part of the data base by the user.) For example, in the admissions program, illustrated in Figure 3, the admitting office specified the patient care unit and the bed assignment. The computer program then examines the data file and permits the bed assignment only if that bed is currently unoccupied. The computer system also automatically generates a message to the appropriate care unit notifying the nursing staff of the admission and supplying appropriate information on the patient. (See lower part of Figure 3). Such a general form of message manipulation is necessary because of the complex interactions which constitute communications in a large hospital; there is also the consideration that the inter-oommunication links must be allowed to change with time. Thus, a user program is available to re-specify the composition of a patient care unit in terms of name, number of rooms, number of beds, and teletype line number.

The interactive and interpretive nature of the computer system is illustrated in Figure 2 in the form of a medication order and a medication listing. In response to question 2, the entry of the room and bed location results in the verification of this code by the computer system returning the name, age, sex and unit number of the patient. In response to a"Y" on question 4, the computer lists all the active medication orders on this patient. In question 5.2, the computer fails to find an exact match for the spelling of the drug name entered and types back a list of drug names which have a similar phonetic spelling. In the same question, the computer notes that the usual maximal dose for this drug has been exceeded. The user over-rides this formulary limit by entering the dose, preceded by an asterisk. In question 5.3, the computer returns the generic name for a drug which has been requested by a trade name, and in addition, types information which the hospital pharmacist had entered in the hospital formulary about this particular drug.

In present hospital operation, there are a number of procedures and different types of cross-sectional listings that are used to implement a doctor's medication order. In order to have patient X receive drug Y at time T; the nurse performs a considerable amount of time on clerical duties; on a single patient care area, she might sort and read medication cards more than a thousand times in an eight-hour shift. In the computer system, this sorting and listing is done automatically, and the computer generates each hour on each patient care unit a list of the medications which should be administered at that hour. This is illustrated in the bottom part of Figure 3, where the medication list for this care unit contains those medication orders to be administered at three o'clock.

Figure 4 is an illustration of a history of test results on one patient for one week. These results are automatically organized in a format designed by the hospital to display tests in groups, such as serum electrolytes, hematology tests, etc. There are over fifty different patient-care functions for which similar programs have been written. Many of these,

although useful for demonstration, are too primitive or unreliable for daily operations on a patient care unit.

D. EVALUATION OF ACCEPTABILITY AND RELIABILITY

Programs now in use at the Massaohusetts General Hospital on a research basis provide for:

1. the admitting, transfer and discharge of patients;
2. a census of patients and an inventory of available beds;
3. the recording of medication orders and the automatic printing of up-to-date lists of drugs to be administered;
4. the charting of medications given and the preparation of summaries of past drug orders and drug administrations;
5. the entering of laboratory test results and the printing of any selected sample of laboratory test results.

The computer system is being developed and evaluated on nine patient care areas (about 300 patient beds) and three laboratories at the Massachusetts General Hospital. However, the programs have been operational for less than a year and in most cases are in an evolutionary stage. Much of the effort is still limited to a smaller number of patients bed (10 to 40 on a ten hour/day, five day/week basis. The system is not being used for actual service operation; that is, it is not actually replacing any hospital function. However, it is being tested by duplicating the present hospital operating procedures on a parallel basis. At this point, we have a very impressive demonstration system, but transition to a 24-hour/day, 7-day/week operational system is not an easy or a rapid process.

There are numerous questions of acceptability and reliability that have not been answered. For example, one of the major concerns of the hospital has to do with the actual identification of the personnel who would directly use the system: whether the Teletype operator would be a nurse, a doctor and/or a specially trained clerk. This answer is dependent on the sophistication of the particular terminal devices selected; the computer system and the design of the programs can be adapted to a variety of such devices. Initially, we have used keyboard entry and are experimenting with a matrix input device and with oscilloscope display routines

The reliability problem is related to the degree of flexibility desired, to the size and sophistication of the hardware, to the development of back-up procedures both in the computer system and in the hospital, and to the use of the computer as a research tool as well as a service system. There are also factors involving record secrecy and legality which we are only beginning to consider.

Three factors do appear obvious, however. First, it must be re-emphasized that the definition of the problems and the specification of appropriate medical and nursing algorithms are vital first steps; there is little previous experience in this area and the developmental efforts are plagued with frustration and misconceptions. The importance of critical, imaginative, and knowledgeable planning cannot be over-emphasized. Second, the problem is more complex and difficult than was originally conceived; the desired system must be both powerful and flexible; designed not simply to grow, but also to change in rather fundamental ways. Third, the preliminary experience we have had thus far strongly indicates that a well-designed communication and information-storage-retrieval system can definitely lighten the paper-work requirements of the professional staff and improve the methods of hospital administration. It is hoped that such a time-sharing system can facilitate medical research and, more important, make possible a higher standard of patient care.

BIBLIOGRAPHY

1. STACY, R.W., WAXMAN, B. (Editors): Computer in Biomedical Research. New York, Academic Press, 1965.

2. LEDLEY, R.S.: Use of Computers in Biology and Medicine. New York, McGraw-Hill, 1965.

3. RIKLI, A.E., ALLEN, S.I. and ALEXANDER, S.N.: Study Suggest Value of Shared Computer. The Modern Hospital. 106: 100, 1966.

4. CAMPBELL, C.M.: Akron Speeds Information System Slowly. The Modern Hospital. 104: 118, 1965.

5. COLLEN, M.F.: Periodic Health Examinations Using an Automated Multitest Laboratoy. J. Amer. Med. Assoc. 195: 830, 1966.

6. SLACK, W.V. et al: A Computer-based Medical-History System. New Engl. J. Med. 274: 194, 1966.

7. LINDBERG, D.A.B.: Electronic Processing and Transmission of Clinical Laboratory Data. Missouri Med. 62: 296, 1965.

8. VALLBONA, C., BLOSE, W.F. and SPENCER, W.A.: System for Processing Clinical Research Data. 6th IBM Medical Symposium Proceedings. 437, 1964.

9. BARNETT, G. O. and BARUCH, J.J.: Hospital Computer Project, Memorandum Eight. Boston, Massachusetts General Hospital, 1965.

10. BARNETT, G.O. and BARUCH, J.J.: Hospital Computer Project, Memorandum Nine. Boston, Massachusetts General Hospital, 1966.

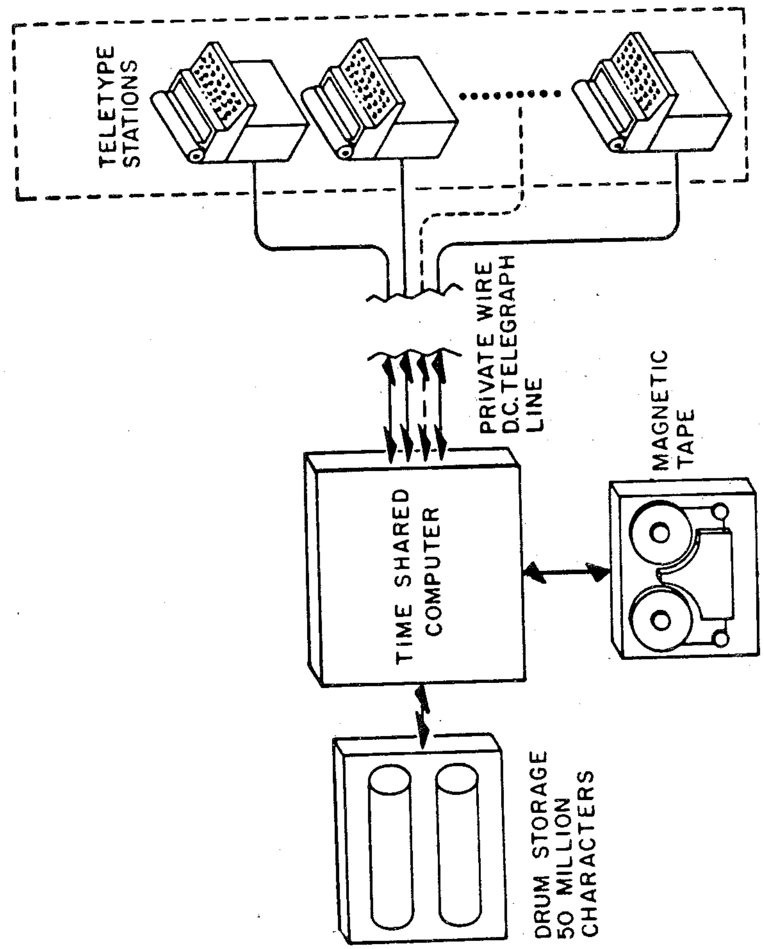


Figure 1

Hardware configuration of the experimental time-sharing computer system.


```
>
1:04 PM 3/24/1966   TH2F(26)   TH1
CALL MEDORDX   BJT

 MEDICATION ORDER

1  D  -/
1  DATE  T
  1.1  TIME  T
2  DOCTOR  BARNETT,O
3  PATIENT  102B DAVIS, JANE      54  F   843-32-45
4  LIST  Y
     2  ASCORBIC ACID,50MG,PO,BID,I
     1  TETRACYCLINE,250MG,PO,4ID,5DAY
5  ORDER
  5.1     ASPIRIN,0.6GM,PO,Q3H,5DAYS
  5.2     TYROID,150MG,PO,3P,I
...DO YOU MEAN...    THYROID

  Y
USUAL MAXIMAL DOSE IS   100.000 MG
DOSE ?  *150MG
THYROID,*150MG,PO,3P,I
  5.3     KYNEX,500MG,PO,4ID,7 DAYS
SULFAMETHOXYPYRIDAZINE
CONSIDER SHORT-ACTING SULFONAMIDE
  5.4
6  INITIALS  BJT

1  DATE  --

-X-
```

The typescript of the program used in ordering a medication and an example of a medication listing. For illustrative purposes, the information entered by the user is underlined.

```
TH1
MED LIST FOR 3:00 PM 3/24/1966


TH102B      DAVIS, JANE      54F       843-32-45
     3 ASPIRIN,0.6GM,PO,Q3H,5DAYS
     4 THYROID,*150MG,PO,3P,I

TH103        MARBLE, CURT      25M        999-45-B5
     6 TETRACYCLINE,250MG,PO,9A 3P 9P 3A,10DAYS

2:30 PM 3/24
```

Figure 2

```
1:01 PM 3/23/1966   TH2F(26)   TH1
CALL ADMSHRD  BJT

TH1

SHORT ADMISSION FORM

1   DOA  ←/
1   ADMISSION DATE  3/21/66
2   PATIENT NAME  DAVIS, JANE
3   ADDRESS  45 CROWN WAY, LINCON, MASS.
4   UNIT NUMBER  ←HOW  E.G. 131-64-17  1111-11-11   FIX
4   UNIT NUMBER  888-87-56 ALREADY IN USE  FIX
4   UNIT NUMBER  843-32-45
5   DIVISION  BM
6   SERVICE  MED
7   DATE OF BIRTH  01/02/12
8   AGE  54
9   SEX  F
10  VISITING DOCTOR  BARNETT, O
11  ADMITTING DIAGNOSIS  SIKK\\CK
12  INSURANCE  BC
13  LOCATION  TH109  ILLEGAL CU AND/OR ROOM NUM  FIX
13  LOCATION  TH104 ALREADY IN USE
13  LOCATION  TH102B
14  INITIALS  BJT

1   ADMISSION DATE  -

-THANK YOU-
```

The typescript of the program used in the admission of a patient to the hospital and the notice that is automatically generated on the appropriate patient care unit.

```
NOTIFICATION OF NEW ADMISSION

DAVIS, JANE      54      F      843-32-45
TH102B
SICK        BARNETT, O

-THANK YOU-
```

Figure 3

FOR ILLUSTRATIVE PURPOSES, THE INFORMATION ENTERED BY THE USER IS UNDERLINED.


```
4:19 PM 3/25/1966   TH2F(26)   TH1
CALL LABREX   BJT

  1.1   PAT  ←/
  1.1   PATIENT  TH102B  DAVIS, JANE
  1.2   BEGINNING DATE FOR REPORT  3/10
  1.3   ENDING DATE FOR REPORT  3/17

  2.1   PATIENT  -
ADJUST PAGE; HIT EOM TWICE WHEN READY.
```

DAVIS, JANE
843-32-45   TH102B

LABORATORY REPORTS
MAR 10 - MAR 16

DATE SPECIMEN COLLECTED

	MAR 10	MAR 11	MAR 12	MAR 13	MAR 14	MAR 15	MAR 16
HEMATOLOGY REPORTS							
HEMATOCRIT %	38			34	36		PEND
SERUM ELECTROLYTE REPORTS							
SODIUM MEQ/L		132	143				
POTASSIUM MEQ/L		3.5			PEND		
CALCIUM MG%	10.6				PEND		
CHLORIDE MEQ/L		97	92				
REPORTS ON BLOOD OR SERUM METABOLITES							
UREA NITROGEN (BUN) MG%	16	20			PEND		
GLUCOSE, FASTING MG%			108				
	MAR 10	MAR 11	MAR 12	MAR 13	MAR 14	MAR 15	MAR 16

Figure 4

Example of a laboratory report showing how the report is requested on the patient care unit.

A CONCEPT OF DATA PROCESSING FOR MEDICAL RECORDS

By A.E. Bennet, M.D., S:t Thomas Hospital Medical School, London, England.

The past six years have seen numerous reports of automatic data processing of the whole or parts of the medical record. An early paper by Schental et al.[1] in 1960 clearly recorded the need for improved techniques in the handling of clinical and research data, and proposed that the routine use of a digital computer was feasible for this purpose. In this and their two subsequent papers[2] the described techniques which they had developed and tested. Their concept, however, was based on recording as much of the information as could be put into computer translatable form. Other workers, Korein et al.[3], Lipkin and Woodbury[4] published accounts of work which was also based on this concept. Although all succeeded in gaining their stated objectives, they clearly recognized the limitations and difficulties encountered and were prepared to critically re-examine their work.

Handling much more uniform data arising from autopsies and laboratory investigations, Bahn et al.[5] and Smith and Melton[6] adopted an approach orientated towards the needs of research. Thus at one end of the scale we have attempts to record the whole clinical history, examination and special data which would fulfil the need for medical care and also for medical research; at the other end we get see a limited approach fulfiling the need for research in a restricted field. The all embracing approach has proved expensive, impractical and I would suggest unsatisfactory: the narrow approach is perfectly satisfactory but has inherent disadvantages in a limited design.

This suggests the need for definition of purposes. Are we attempting to fulfil the needs for patient care and so replace medical notes as they exist at the moment; or are we attempting to extract and process selected

---

NB References at page 122

Notes at page 123

data for research programmes using the search abilities of the computer to produce collected fact and allow associations between different factors to be uncovered? Or is there a common path which might ultimately lead to success in both fields?

Some part of the present contents of the individual record will have to b retained. X-ray plates, photographs, electrophoretic patterns, tracings etc. cannot be stored in a computer, at least not economically at present. So it would seem that some system of manual filing must remain for the forceeable future and, therefore, I suggest our ojbective should be the modifioation and simplification of present methods. Accepting this, success of computer applications in this field depends on how well any system devised augments and simplifies the present organization rather than attempts to replace it. Whatever is done must be practical in the context of individual therapeutic medicine; it must be simple, convenien and require no additional practical skills. If we cannot meet these requirements then I consider that the results of our efforts will be severely limited and the value thus largely destroyed.

At present our case notes exist with the clinical history, examination and continuation notes written as a longhand narrative in a partially organized sequence. They are completed on blank forms and include relevant positives, some only of the relevant negatives and are of such length or detail as is considered necessary by the clinician. At the close of a period of patient care either by discharge from the ward or from attending an out-patient clinic, a summary of the case is prepared. A copy of this is retained in the notes and a copy sent to the patient's referring physician. We believe that further development of this summary would serve the needs for future medical care and allow accumulation of data for research. To go further and record all information assumes that it is all equally valuable and this we cannot accept; nor is it required for research and administration purposes. Much clinical data is subjeot to the biases of observer variation and this destroyes a large measure of its usefulness in research.

Initially a study of the handling and content of case notes was undertaken. This was reported[7] in 1963 and although other such studies must hav been undertaken there is to my knowledge no other published study for comparison. The method employed was to identify a random sample of the direction this work was first started and still continues.

REFERENCES

1. SCHENTAL, J.E., SWEENEY, J.W., NETTLETON, W. (1960) J.Am. med. Ass., 173, 6
2. - - - - - (1961) ibid. 178, 267
   - - - - - (1963) ibid. 186, 101
3. KOREIN, J., WOODBURY, M.A., GOODGOLD, A.L. (1963) J.Am. med. Ass., 186, 132
4. LIPKIN, M. and WOODBURY, M.A. (1962) Communs Ass. comput. Mach., 5, 532
5. BAHN, R.C., SCHMIT, R.W., YOUNG, G.G. (1965) Proc. Staff Meet. Mayo Clin., 39, 835
6. SMITH, J.C., MELTON, J. (1963) Meth. Inform. Med., 2, 85
7. HOLLAND, W.W., De BONO, E., GOLDMAN, J.A. (1964) Lancet, 1, 819
8. COCHRANE, A.L. CHAPMAN, P.J., OLDHAM, P.D. (1951) Lancet, 1, 1007
9. FAIRBAIRN, A.S., WOOD, C.H., FLETCHER, C.M. (1959) Br. J. prev. soc. Med., 13, 175
10. SMYLLIE, H.C., BLENDIS, L.M., ARMITAGE, P. (1965) Lancet, 2, 412

NOTES ON THE PROCEEDINGS OF AN INFORMAL DISCUSSION GROUP ON THE ELECTRONIC DATA PROCESSING OF MEDICAL RECORDS

Friday, 29th April, 8.30 - 10.0 p.m.
Chairman: Dr. A.E. Bennett.

Following the opening remarks of the Chairman, Dr. Jungner expressed h views on the need for the development of procedures which would allow recording of serial information on individuals. Record linkage was disc sed and all members present agreed on the benefits that would be deriv from this form of operation and the work of Newcombe (Canada) and Acheson (England) was mentioned. Expense was briefle discussed but the problem of a positive identification for each individual emerged most clearly. Dr. Korein suggested that the fingerprint could provide a uni identification, convenient and readily typeable, but the political impl cations of such a method of routine identification were not acceptable to some participants. Other methods based on a unique alpha-numeric num ber and Soundex coding of names with other identifying details were de- bated.

The discussion progressed to considerations of the clinical record. Mr. Baruch with Drs. Korein and Tiok maintained that processing of all the data contained in the case-notes was desirable to serve firstly the needs of medical care and then research. Professor Dalenius replied tha the lack of comparability and uniformity in the data contained in case-notes, particularly in the narrative parts, made the value extremely lo for any form of retrospective research. This was upheld by other member and the point was made that large-scale collection of data in the hope that it might prove useful was an expensive, wasteful process. This was denied by Mr. Baruch who maintained that the process of sifting data captured by an information system was more expensive than storing it in case of future need. This was accepted by all members with the proviso that the value of such data was very low and that the true role of the computer for medical research in this field lay in its capability of pro cessing data which was uniform and valid and collected with the defined objective of testing a hypothesis.

SOME ORGANIZATION AND ADMINISTRATION PROBLEMS OF HOSPITAL INFORMATION MANAGEMENT SYSTEMS

By Lee B. Lusted, M.D., Professor of Radiology, University of Oregon Medical School Senior Scientist, Oregon Regional Primate Research Center.

Automatic data processing joins the medical and hospital family during a tumultuous era highly charged with potential opportunities for automatic data processing. In the United States the cross-currents of changing times are heaping up pressures which seem likely to alter many established patterns of medical practice and health care. One observer(1) sees the situation essentially as a struggle between the medical schools and the medical practitioners for the control of medical education and the basic structure of medical care. The recent passage of the Medicare program has raised difficult questions about methods of financing the costs of medical and hospital care and over the entire health field hovers the specter of a shortage of doctors, nurses and paramedical personnel.

In commenting on the shortage of doctors, Dr. Philip R. Lee, (2) Assistant Secretay of Health, Education and Welfare, recently said: "Doctors are just going to have to work harder". However, Dr. David D. Rutstein, (3) Professor of Preventive Medicine at the Harvard Medical School, and Dr. William H. Stewart, (4) Surgeon General of the U.S. Public Health Service, share a different view. They say in essence that better use of the physician's time is the key to future medical care and they propose the widespread use of medical assistans, automation and the computer to conserve the physician's time. But this is not easily accomplished

Work supported by the National Institutes of Health, Grant Fr 00163, Contract No. PH43-64-1163, and the Medical Research Foundation of Oregon.

NB Summary at page 134
References at page 135
Figures at pages 136 - 139

as Dr. Stewart(4) subsequently noted because more and more the physic is functioning as a manager - a job for which he is not specifically prepared either by training or tradition. He must direct the efforts and energies of a growing number and variety of health workers.

Can automatic data processing help bring improved organization to thi confused situation? I remember hearing a story several years ago, whic was perhaps apocryphal, about how the computer brought peace to the airframe manufacturing industry. It seems that some of the bitterest intercompany warfare was taking place among several airframe manufacturers in the early 1950 era. These companies eventually discovered that there were great advantages in airframe design to be gained by using electronic computers. However, computers were very expensive and besides there weren't many computers available. So after much soul searching the airframe companies decided that they should share a computer facility. Which they did and thus began an era of cooperation an better understanding. Who knows what time-sharing computers and region hospital systems may yet do for the health field.

My interest in the subject of automatic data processing in hospitals extends back over the past fifteen years and during the last ten years my research activities have dealt with one or another aspect of biomedical computing. In this paper I want to discuss some of the organiz tion and administration problems which arise when an automatic data pr cessing system is used in a medical center or hospital. In presenting this discussion I am drawing on a variety of experience from the past several years: from my professional work as a clinical radiologist, fr work with the National Institutes of Health Advisory Committee on Computers in Research and from my experience for the past four years as project director of the Primate Information Management Experiment (PRIME), Oregon Regional Primate Research Center.

Since a primate center may seem an unlikely place in which to undertak a computer-based information management experiment some explanation seems in order. Two factors were particularly important in choosing th Oregon Regional Primate Research Center. First, the Oregon Center was to house approximately 1000 subhuman primates which were to be used for a wide variety of biomedical research projects and each primate was to have an individual record similar to a patient's hospital recor

High standards of health are maintained among the primates and consequently there is a constant flow of reports on routine laboratory examinations, x-ray examinations, surgical procedures and medication. Second, the information management system although somewhat similar to a hospital record system would not have the service pressures of patient care which are present in a hospital.

Experience with the Primate Information Management Experiment has shown that the computer administrator, who must gain the cooperation of all participants, is faced with a dilemma. He must be influential in changing and guiding the work habits of men who do not work for him and try to gain their cooperation even though they may not feel that they will benefit from the experience. At the same time, he must convince the administrative supervisors of these men that they, the administrators, will benefit from a central record system and promise that it will not unduly interfere with the normal activities of the employees. The organizational position of the administrator of the computer facility is important to the general success of the project if it tends to cross traditional lines of authority. If he is lower or equal in rank to the administrative heads of participating departments, it is not likely that he will receive adequate cooperation.

I want to proceed now to comment on some administrative interface problems of hospital data processing systems.

I   <u>The Administrative Position of the Computer System Director is Important.</u>

Some years ago the Harvard Business Review[5] published the results of a survey which showed that in companies claiming the best results from computer usage the computer executive was no more than two administrative levels below the company chief executive. I doubt that a similar survey has been made for hospital computer projects but I think that the observation from the business field holds a good deal of validity for the medical center or the hospital. In the medical center or hospital this means that the computer system director should be responsible directly to the dean of the medical school or the hospital administrator. Such a statement may seem a truism of good administrative practice but it frequently has been ignored in setting up biomedical data processing

systems. We have seen recently one or two instances in which the medical center automatic data processing system was started as a department project and was subsequently transferred for administrative purposes to the medical center director's office. The result was, of course, to make the automatic data processing system function as an interdepartmental activity. I don't wish to give the impression that by assigning the computer administrator to the hospital administrator's office that we shall be able to solve all of the problems but perhaps some of the stickier ones will be less difficult to deal with.

A.   <u>Hospital computer systems are interdepartmental in function.</u>

A hospital computer system by its very nature performs functions which affect several departments or services in a hospital. This is true whether the system is small and does only laboratory work or is large and does the medical record processing. The medical center or hospital may already have an administrative mechanism for dealing with interdepartmental matters but the mechanism may not be adequate to deal with questions raised by an interdepartmental automatic data processing system.

For instance, in a hospital with a medical record automatic data processing system, how should such questions be decided as (1) what data should be collected and who should record it; (2) what procedures should be used for checking the accuracy of the data; (3) what people should have access to what kinds of data. If an interdepartmental committee is set up to advise on these matters the hospital administrator is in the position to appoint the committee. An important point which is frequently overlooked is that the persons responsible for the data processing, i.e., the director of the computer system and programmers are in no way responsible for deciding what kinds of information should be generated by the system.

B.   <u>Hospital staff resistance to the hospital computer system may be a problem.</u>

Computerized science in general still receives a mixed reception and since computers are relatively new to the hospital field they are frequently regarded by medical researchers and physicians with a good deal of skepticism.

Spinrod[6] who heads the Computer Systems Group at Brookhaven National Laboratory, recently reported the results of a survey of biomedical research workers which showed that 23% foresaw more rapid computer growth in their field, 19% thought such growth would lag behind its current slow pace and 58% believed biomedical computer use would keep pace with science in general. These findings could hardly be interpreted as enthusiastic support for biomedical computing and it might be possible that for any particular hospital only a small proportion of the staff would be in favor of a hospital computer program. The enthusiastic support and participation of hospital personnel is needed for a hospital computer project and I suggest that a sine qua non of success is that the computer project director should have a medical degree because the support of the medical staff is more readily won by such a person. A vigorous education program must be a part of any hospital computer project and should be organized to reach all segments of hospital personnel.

I want to consider next some questions about hospital computer system costs and whether the expense of the automatic data processing system can be justified in terms of improvement in the quality of patient care.

II. Hospital Computer Systems Are Expensive. Is the "Show Worth Price of Admission?"

A hospital administrator who confronts the decision of whether he should or should not embark on a hospital computer system project is confronted by a range of opinions which are likely to leave him like Hamlet in an agony of indecision. One point of view is expressed by the opinion of Dr. J.M.A. Lenihan,[7] regional physicist of the Western Regional Hospital Board, England, that "Every large hospital now being built or planned should include electronic computing and data handling facilities. The costs in money, space and staff are trivial in comparison with the benefits to be gained."

A different point of view is expressed by Dr. Robert M. Farrier,[8] Associate Director of the Clinical Center, National Institutes of Health. He says that it is not a question of what computers can do as it is of

what we want them to do -- and whether it is worth the price of doing the things that way. He questions whether the doctor needs to know all that he can know, e.g., continuous temperature recording of every patient, and suggest that we need to be quite critical about whether the "show is worth the price of admission." Dr. Farrier's advice is that most hospitals should wait until hospital computer systems are fully tested because few hospitals can afford **extremely expensive errors** in judgement and no hospital can afford to rush into a program on which life depends.

The administrator who searches beyond opinions for facts will find that there is little quantitative data available on the amount of time spent, personnel required, or the error rate in the present methods of hospital function which are being duplicated by the hospital computer system. Some hospital computer projects have been in progress in the United States three years and a few as long as five years but the number of publications on the progress of the projects is still quite small.

The recent status report[9] on the Massachusetts General Hospital Computer Project points out that some of the limitations of the present hospital computer system are present because the system is experimental and programming activity is constantly in progress. This would not be the case on a computer system in service operation. The Massachusetts General Hospital investigators conclude that in the final analysis, the evaluation of the relative merits of the computer system will have to be based to a major extent on the critical judgment of the professional staff of the hospital. A conclusion which I think may leave the hospital administrator in a rather unhappy state and still wishing that he had more objective data.

Two recent articles which discuss the automatic data processing of clinical records do contain interesting and useful information. The first article reports the computer handling of patient records in the Diabetes Clinic, Western Reserve University Hospitals, Cleveland, Ohio.[10] The investigatore point out that the cost to use the automatic data processing system is approximately $2.00 per patient visit and that they could find no difference in the actual physician time spent with each patient whether or not the computer-based record system or the conventional record system was used. However, the participating

Summary

Many problems confront the person who wishes to develop an automatic data processing system in a hospital. These problems have been arbitrarily divided into the two general categories of "people problems" and technical problems for purposes of discussion in this paper. Of the two categories, the "people problems" are more difficult to manage and they demand more attention if the hospital automatic data processing system is to be accepted as an integral part of the hospital. Some problems of administering the hospital computer system are presented and some comments are made on costs.

The complex problems of medical record processing and communication in the hospital make it obvious that attempts must be continued to develop hospital computer systems. Time-sharing computer facilities with remote input-output units may be available in the future as the basis for hospital communication systems. At the present time such systems are not available and it seems appropriate to develop less comprehensive and less expensive hospital computer-based systems.

An example is given of such a unified computer-based medical record system (PRIME) which has been completely functional for over a year.

REFERENCES

1. SOMERS, A.R. Some socioeconomic observations on medical education and the practicing profession. J.Med.Educ. 40 (2): 258, 1965

2. U.S. News and World Report, page 38, Mar. 7, 1966

3. Medical Tribune, page 16, Mar. 5-6, 1966

4. Am. Med. Assoc. News, Feb. 28, 1966

5. Top Management and Computer Profits. Harvard Business Review 41:6, 1963

6. Medical Tribune, page 11, Feb. 5-6, 1966

7. Medical Tribune, page 16, Jun. 26-27, 1965

8. Medical Tribune, page 14, Jan. 1-2, 1966

9. Hospital Computer Project. 1966 Status Report. Memorandum Nine., Massachusetts General Hospital, Boston, Mass.

10. LEVY, R.P., CAMMARN, M.R., and SMITH, M.J. Computer handling of ambulatory clinic records. J.A.M.A. 190: 1033, 1964

11. BAIRD, H.P. and GARFUNKEL, J.M. Electronic data processing of medical records. N.E.J.M. 272: 1211, 1965

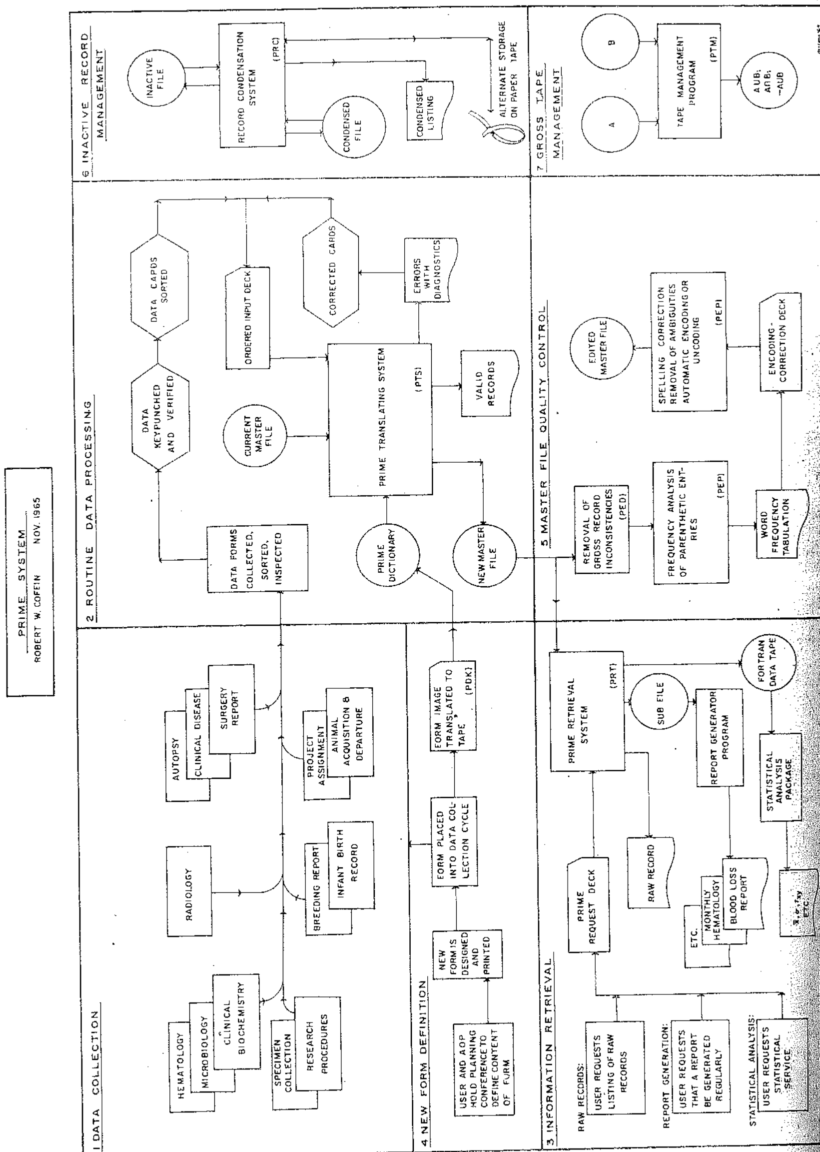
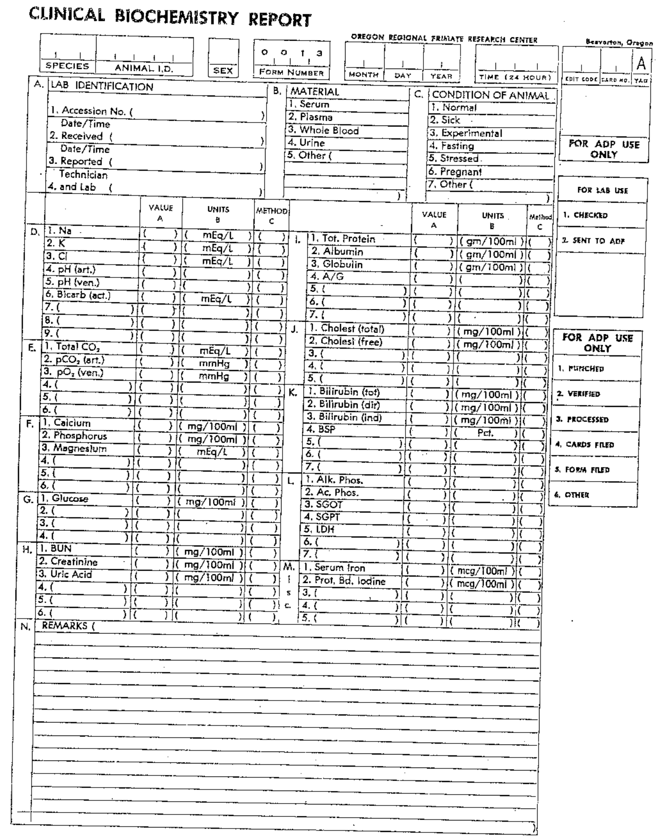


Figure 1



CLINICAL BIOCHEMISTRY REPORT

OREGON REGIONAL PRIMATE RESEARCH CENTER — Beaverton, Oregon

SPECIES | ANIMAL I.D. | SEX | 0 0 1 3 FORM NUMBER | MONTH | DAY | YEAR | TIME (24 HOUR) | EDIT CODE | CARD NO. | A TAG

A. LAB IDENTIFICATION
1. Accession No. ( )
Date/Time
2. Received ( )
Date/Time
3. Reported ( )
Technician
4. and Lab ( )

B. MATERIAL
1. Serum
2. Plasma
3. Whole Blood
4. Urine
5. Other ( )

C. CONDITION OF ANIMAL
1. Normal
2. Sick
3. Experimental
4. Fasting
5. Stressed
6. Pregnant
7. Other ( )

FOR ADP USE ONLY

		VALUE A	UNITS B	METHOD C
D.	1. Na	( )	(mEq/L)	( )
	2. K	( )	(mEq/L)	( )
	3. Cl	( )	(mEq/L)	( )
	4. pH (art.)	( )	( )	( )
	5. pH (ven.)	( )	( )	( )
	6. Bicarb (act.)	( )	(mEq/L)	( )
	7. ( )	( )	( )	( )
	8. ( )	( )	( )	( )
	9. ( )	( )	( )	( )
E.	1. Total CO_2	( )	(mEq/L)	( )
	2. pCO_2 (art.)	( )	(mmHg)	( )
	3. pO_2 (ven.)	( )	(mmHg)	( )
	4. ( )	( )	( )	( )
	5. ( )	( )	( )	( )
	6. ( )	( )	( )	( )
F.	1. Calcium	( )	(mg/100ml)	( )
	2. Phosphorus	( )	(mg/100ml)	( )
	3. Magnesium	( )	(mEq/L)	( )
	4. ( )	( )	( )	( )
	5. ( )	( )	( )	( )
	6. ( )	( )	( )	( )
G.	1. Glucose	( )	(mg/100ml)	( )
	2. ( )	( )	( )	( )
	3. ( )	( )	( )	( )
	4. ( )	( )	( )	( )
H.	1. BUN	( )	(mg/100ml)	( )
	2. Creatinine	( )	(mg/100ml)	( )
	3. Uric Acid	( )	(mg/100ml)	( )
	4. ( )	( )	( )	( )
	5. ( )	( )	( )	( )
	6. ( )	( )	( )	( )

		VALUE A	UNITS B	Method C
I.	1. Tot. Protein	( )	(gm/100ml)	( )
	2. Albumin	( )	(gm/100ml)	( )
	3. Globulin	( )	(gm/100ml)	( )
	4. A/G	( )	( )	( )
	5. ( )	( )	( )	( )
	6. ( )	( )	( )	( )
	7. ( )	( )	( )	( )
J.	1. Cholest (total)	( )	(mg/100ml)	( )
	2. Cholest (free)	( )	(mg/100ml)	( )
	3. ( )	( )	( )	( )
	4. ( )	( )	( )	( )
	5. ( )	( )	( )	( )
K.	1. Bilirubin (tot)	( )	(mg/100ml)	( )
	2. Bilirubin (dir)	( )	(mg/100ml)	( )
	3. Bilirubin (ind)	( )	(mg/100ml)	( )
	4. BSP	( )	(Pct.)	( )
	5. ( )	( )	( )	( )
	6. ( )	( )	( )	( )
	7. ( )	( )	( )	( )
L.	1. Alk. Phos.	( )	( )	( )
	2. Ac. Phos.	( )	( )	( )
	3. SGOT	( )	( )	( )
	4. SGPT	( )	( )	( )
	5. LDH	( )	( )	( )
	6. ( )	( )	( )	( )
	7. ( )	( )	( )	( )
M. Misc.	1. Serum Iron	( )	(mcg/100ml)	( )
	2. Prot. Bd. Iodine	( )	(mcg/100ml)	( )
	3. ( )	( )	( )	( )
	4. ( )	( )	( )	( )
	5. ( )	( )	( )	( )

FOR LAB USE
1. CHECKED
2. SENT TO ADP

FOR ADP USE ONLY
1. PUNCHED
2. VERIFIED
3. PROCESSED
4. CARDS FILED
5. FORM FILED
6. OTHER

N. REMARKS ( )

Figure 2

ØRPRC ANIMAL BLØØD LØSS REPØRT								
FRØM 12/ 1/55 TØ 2/15/56								
ANIMAL ID	TØTAL							
137 2111 M	17.0	1/25/56	15.0	12/23/55	2.0			
137 2112 M	4.0	1/25/56	2.0	12/23/55	2.0			
305 2117 F	27.0	1/31/56	11.0	1/ 5/56	16.0			
305 2118 F	22.0	1/ 5/56	22.0					
305 2119 F	37.0	1/31/56	17.0	1/ 5/56	20.0			
305 2120 F	33.5	1/31/56	11.5	1/ 5/56	22.0			
305 2121 F	15.0	1/ 5/56	15.0					
305 2122 F	28.0	1/31/56	6.0	1/ 5/56	22.0			
305 2123 F	39.0	1/31/56	17.0	1/ 5/56	22.0			
305 2124 F	31.0	1/31/56	16.0	1/ 5/56	15.0			
305 2125 F	29.0	1/31/56	5.0	1/ 5/56	22.0			
305 2126 F	34.0	1/31/56	13.0	1/ 5/56	21.0			
305 2127 F	22.0	1/ 5/56	22.0					
305 2128 F	12.0	1/ 5/56	12.0					
305 2129 F	34.5	1/31/56	11.5	1/27/56	12.0	1/ 5/56	13.0	
305 2130 F	22.0	1/ 5/56	22.0					

Figure 3

CLINICAL LABØRATØRY RESUME

TEST DETERMINATIØN: SØDIUM

ANIMAL CØNDITIØN: CATEGØRY NØRMAL

AGE: ALL

SEX: F

METHØD:

DATES: 01/01/60 TØ 02/03/65

SPECIE	<= 130	131 132	133 134	135 136	137 138	139 140	141 142	143 144	145 146	147 148	149 150
305	4	1	0	2	7	9	16	13	21	49	70
283	0	0	0	0	0	0	0	0	0	3	0
300	2	0	0	0	3	3	3	6	12	15	10
392	0	0	0	0	0	0	0	0	0	0	0
88	0	0	0	0	0	0	0	0	0	0	1
301	0	0	0	0	0	1	0	1	4	2	2
308	0	0	0	0	0	0	0	1	0	1	1
85	0	0	0	0	0	0	0	0	0	0	0
233	0	0	0	0	0	1	0	0	0	0	0
TØTAL	6	1	0	2	10	14	19	21	37	70	84

STATISTICAL SUMMARY

SPECIE	MEAN	STD ERRØR MEAN	STD DEV
305	151.31762	0.55779	11.19761
283	152.88889	1.52247	4.56740
300	149.01111	0.76397	7.24762
392	170.00000	0.00000	0.00000
88	150.00000	0.00000	0.00000
301	149.69231	2.13453	7.69615
85	151.00000	0.00000	0.00000
233	140.00000	0.00000	0.00000
ALL	150.89675	0.45672	10.44475

Figure 4

INTEGRATED INFORMATION SYSTEM FOR AN INTEGRATED MEDICAL CENTER

By Armando Lassus, Manager, Information Processing Division, Puerto Rico Medical Center.

INTRODUCTION

The Puerto Rico Medical Center began to take shape in 1945, when the Puerto Rico Planning Board developed a preliminary plan for an integrated hospital and public Health facilities program - a vast and imaginative project designed to bring better health to hundreds of thousands of my fellow Puerto Ricans.

Ten years later, in 1955, Puerto Rico's Department of Health achieved agreements with a number of institutions sponsored by Federal, State, Municipal, and private agencies. Two years later, the services of a firm of architects were engaged for the development of the master plan of physical facilities. In July, 1966, the Puerto Rico Medical Center is due to be fully operational as an Integrated Medical Center.

The fact that the Medical Center is still in the process of being born has been both a blessing - and a complicating factor. We were not bound by an existing way of doing things. On the other hand, we had absolutely no basis of experience on which to build our plans.

In my presentation I expect to cover the following topics:

1. Origin and organization of the Medical Center
2. Problem formulation and specific objectives
3. Discussion of each of the applications comprising the initial phase
4. The approach followed to initiate the actual planning of the system
5. The plan that was established for the development of the system
6. The implementation plan, which takes into consideration a gradual implementation of the total system

7. The several problem areas which we have encountered from the very beginning of our planning work and the way that these problems were partially or totally solved.
8. Value of the Integrated Information System in relation to the total operation of the Medical Center.

ORIGIN AND ORGANIZATION OF THE MEDICAL CENTER

The Puerto Rico Medical Center is the largest and most extensive of three medical centers that are being established as the highest echelon of medicine in each of three health regions in which the island has been divided for the Health Care Program.

In the chart (Chart 1) you can see some of the estimated work load projections.

CHART NO. 1. PUERTO RICO MEDICAL CENTER, WORK LOAD PROJECTIONS

Activity	Daily Volume	Annual Volume
Inpatients:		
Admissions	170	45,000
Discharges	170	45,000
Births	35	9,200
Outpatients:		
Patients	--	60,000
First visits during the year	225	60,000
Re-visits	900	240,000
Total visits	1,125	300,000
Emergencies	240	87,000
X-Ray:		
Examinations	450	130,225
Plates	990	285,000

PLANNING A COMPREHENSIVE HOSPITAL INFORMATION SYSTEM

By P.J. Budd, Chief Data Management Director, United States Veterans Administration.

It is my intention to share with you the planning, problems, and experiences which we at the United States Veterans Adminstration have had in seeking to take full advantage of the potential of data processsing in the field of medicine.

After providing you with a limited organizational and historical background, I intend to: (1) relate to you our approach to, and our progress in the use of electronic data processing in our medical activities. I then particularly want to describe for you: (2) the several determinations which we shall be making in the course of our present studies -- and indicate for you (3) the potential benefits and the problems which we believe are inherent in the use of electronic data processing methods in hospital operations.

THE VETERANS ADMINISTRATION

I will begin by stating that the United States Veterans Administration is an agency of the U.S. Federal Government. The top agency official, the Administrator of Veterans' Affairs, is appointed by, and is responsible directly to the President of the United States. He is charged with the implementation of national legislation and administrative policies pertaining to the benefit programs authorized for our veteran population.

There are approximately 26,000,000 veterans in the United States, and this total is continually increasing. The benefits which the citizens of the United States have made available for our veterans are diverse in nature. While most of my presentation today will be concentrated on the VA Medical Program, I will briefly mention the other principal benefit programs. I do this for a specific purpose -- for it was in the benefit areas that the Veterans Administration achieved initial success in the

utilization of electronic data processing methods. The cumulative impact of that work, the continuing technological innovations, and especially the knowledge generated by the efforts of others have instilled an intense interest in us to use automatic data processing in our hospital operations.

In addition to the medical program our major benefit programs are: (1) the payment of compensation to ex-servicemen disabled while in military service; (2) the payment of pensions to discharged servicemen who are in need and have a nonservice-connected disability; (3) the monetary assistance to veterans for education and training upon their return to civilian status; (4) a low-interest financial loan program to assist veterans in acquiring private homes or in the establishment of private business enterprises, and (5) the provision of life insurance coverage which affords participating veterans and their dependents economical protection for financial security.

Prior to the advent of the general purpose electronic computer, the continued effective management of these large programs, and especially the medical program, loomed ominously before us.

This is precisely why the Veterans Administration responded positively and speedily when general purpose electronic computers arrived during the 1950's. Since that time, we have automated a vast array of what were formerly manual and time-consuming business procedures in virtually every program area -- and we are now conducting research and development efforts to extend and refine our present automated procedures.

The primary purpose of the VA Medical Program is to provide a high quality of medical care to our veterans. We consider medical research and the professional training of medical personnel as corollary medical program objectives. To facilitate the achievement of these goals, the VA supports an extensive medical research program and actively conducts various professional training programs in conjunction with 78 schools of medicine, 41 schools of dentistry, 69 schools of nursing, 68 universities awarding doctoral degrees in psychology, and equally substantial programs in social work, physical and occupational therapy, and other medically related professions.

MEDICAL TIME-SHARING SYSTEMS

By Jordan J.Baruch, General Manager, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, USA.

I would like to talk today about what is going on inside a couple of the computers, and the types of programs that we are using at the Massachusetts General Hospital. I would like to talk about the most frequent questions that we get about time-sharing systems, talk a little bit about a system which is running now at the Mass. General Hospital, and then spend most of my time talking about a somewhat larger system that has not been put together at the current time. Our current plans are to be operating with one or more experimental hospitals in 1967. The main reason for talking about the larger system is that it is a good way of illustrating many of the things that we have learned in the five years that we have been monkeying with small, medium, and large time-sharing systems.

The basic hospital problem has to do with the capturing of data getting that data into a data base which is stored some place in the computer's memory and then being able to manipulate that data either on command or on program command. This is what people do in the hospital essentially now, it is what people will do when they have a computer.

A SYSTEM CONFIGURATION

Let me show you a diagram of the system that is being put together at the moment for the purpose of handling hospitals, since it does illustrate the components that go into a moderately large time-sharing system. I labeled three hospitals on the left hand side and in each hospital there is a terminal. As we get on with our discussion we will see that these terminals may be quite different and that some of them may in fact be computers. You will notice that the terminals feed over lines to little boxes in the hospital. These boxes are local computers hence the term LC. The purpose of the LC is to take the information coming in from the terminals which may be at a rather low rate to store it in memory for a short period of time and piping out over the wire to the computer center 1 at the rate of about the same as voice frequency which is about 2000 bits per second. This is not a technological necessity, it purely serves to keep telephone line rental low. So we have these local computers that take the information coming from the terminals, and put it out on telephone lines, they also take the information coming from the computer center over the telephone lines, and distribute it among the terminals. However, since it is a computer it can do certain tasks that we would like it to do. For example it can control the format of the information going out to the terminals. The multiplexers (M) are used again to take the signals from little terminals and put them out over common telephone lines. In this case they come out of two terminals are squeezed into one telephone line up to hospital no. 2 where there is a local computer, and then the local computer treats them as if they came from its own terminals. Hospital no. 1 has associated with its local computer a device labeled D which is a disk. This disk is a storage device for those circumstances where we need a large buffer of information because one of the terminals, like J. Macy's may be puring stuff out like a firehose instead of trickling it, so we have a big sump there where we can soak up the information then spew it out over the telephone line at a suitable rate. Another reason why we use a local disk, however, is that we need a cliché device so that we don't saturate the communication medium. Clichés are voluminous standard messages or displays. It is senseless to transmit all those bits: we can store those clichés on a disk file in a hospital and just transmit the trigger that sets them off. The trigger says where to put it and when to put it, and that takes very little information. All this go over those two lines to a computer center. The computer on the left is assumed to be plugged in. It consists of a CP which is a central proces-ser connected to which is a fairly high speed drum.

THE BALANCED MEMORY

Let us take a look at the memory. There is a set of general registers: these are the arithmetic registers that you are all familiar with, the accumulator the input-output register, and so on, all high speed registers made out of solid state devices. The general registers contain

NB Discussion at page 187

ON THE REQUIREMENTS FOR INFORMATION SYSTEMS IN HOSPITALS

By Charles D. Flagle, Professor, the Johns Hopkins University, Baltimore, Maryland, USA.

PART I - INTRODUCTION

The purpose of this paper is to provide a basis for review and evaluation of work in the field of data processing in hospitals. Hopefully, a model can be constructed within which each of our efforts can be recognized for its contribution to the total, and at the same time we may perceive the gaps that must be filled.

With the possible exception of computer aided diagnosis, there is a paucity of literature on medical and hospital data processing. This is perhaps because the subject is one for empirical development rather than scientific research. The potentials of data processing machinery are so great that their brute force application is sufficient to produce great change in old systems. Perhaps there is another explanation for the absence of technical reporting of studies of hospital communications. It is a messy business and the methods of observation and sampling do not yield studies comfortable to publish in learned journals.

Nevertheless the weight of numerous ad hoc studies indicates an enormous amount of the human effort in hospitals is consumed in communication and information handling processes. Hsieh (4) cites a number of studies, showing that about thirty per cent of nursing time is spent in communication, and half of this can be found in the transcription and other paper work associated with patient medication. Chagrin at this diversion from direct patient care is heightened by evidence that errors occur in 5% to 15% of the procedures.

The enormous cost, the errors and delays in hospital communications are enough in themselves to inspire effort and justify capital expenditure in new data processing systems. Beyond the immediate improvement in support of existing procedures there are potentials for totally new schemes of medical care administration.

To establish perspective for consideration of the broad problem of information flow in a hospital or other health services it is helpful to think of the institution itself as a purposeful system. As a system a hospital is both complex and complicated. It is complex in its multiple goals of patient care, teaching, research and survival as an institution. It is complicated in its several streams of professional and administrative authority, each controlling some of the set of services that combine in success or failure in reaching its objectives.

The dominant central action in the system is patient care, the interaction of physician, patient, and the supporting hospital services in the diagnosis and treatment of desease. Within the hospital a number of these basic physician-patient relationships occur - perhaps hundreds, and if we include outpatient care, thousands, on any given day. The short term problem of hospital administration is to support, with all its many services, the large number of physician-patient relationships that are <u>simultaneously</u> and <u>independently</u> in progress. The system view of both the individual physician-patient relationship and the total hospital as a support of many such simultaneous relationships will be expanded upon later, after an examination of the general problem of system performance and control. The argument to be developed is that communication plays the key role in successful system performance. In the hospital system, it is not far fetched to say that many of the current operating problems manifested as shortages of personnel and facilities or inadequacies and errors in service are in reality evidence of a crisis in communication.

In most western countries the trend in hospital administration has been toward centralization and specialization of ancillary and supporting services, removing them physically and administratively from the wards and clinics. Sterily supply, dietary, pharmacy and housekeeping services are examples of activities increasingly separated from the scene of patient care. While this has freed medical and nursing personnel on one hand, it has created new logistical problems of coordination, inventory,

---
NB References at page 197

physicians' orders, those related to intensity of need for nursing care. In the latter case the universe of events is a group of patients in a ward or clinic, and the decisions are those of personnel and logistical support requirements. An example of the great variability of aggregate needs of the set of patients on a hospital ward is shown by Connor et al (6). First, in work measurement studies, it was shown that the bedfast, intensive care patient requires about five times as much attention as those who are ambulatory and convalescent. Then it was shown that, because of the random, independent aspect of admission or relapse of intensive care patients, their number was quite variable ranging from two to fifteen on a thirty bed ward. This work demonstrated 1) the need for daily observation and assessment of aggregate patient requirements, 2) the need for timely interpretation and transmission of the results of such observations, and 3) the ability of the administrative system to have the flexibility and responsiveness to meet the demands thus forecast. In summary, the study demonstrated the system characteristics of ward administration on a day to day basis and proposed methods of short term decision making. Similarly by consolidation of physicians' orders the aggregate work loads on laboratories and other ancillary services are made available for short term decisions. Techniques for medium and long term decisions are not as fully developed but we can assess their nature to some extent. Basically the information requirements are those for forecasts - one may extrapolate directly from observed trends in resource utilization or attempt to correlate resource utilization with demographic variables whose trends can be reliably estimated.

PART IV. THE STATUS OF INFORMATION SYSTEMS IN HOSPITALS

Many forces have conspired to focus effort in hospital information system development into short term problems. As noted before, existing communications are very costly; there is an attractive opportunity to save time and money. The cumbersome existing procedures have come about through relatively recent centralization of hospital services and logistical support away from wards and clinics, thus creating need for communication and coordination. In addition, short term decisions are related to existing activities and procedures; the problems are tangible and accessible, their solution is not only necessary but apparently possible. Thus researchers have been encouraged to examine the potentials of automatic data processing for the day to day activities of patient care and hospital administration.

No doubt the medium and long term problems are equally pressing and in the long run more important, but the immediate applicability of data systems is less clear. Decision procedures in forecasting and planning are not well formulated. They may be strongly influenced by political and social factors as well as by statistics. Much of the necessary forecasting requires well linked sources of information. Modern data systems are offered here not as an efficient substitute for costly procedures, as in the short term case, but as something new, an added cost rather than a substituted one. However, it must be recognized that poor medium and long term decisions are extremely costly in wasted resources or unmet needs. Therefore, though recognizing the difficulties, system analysts must puch their work beyond the short term decision problems to a creative association with leaders in medical care in search of procedures for planning future health service systems.

REFERENCES

1. COOMBS, C.H., A Theory of Data, John Wiley and Sons, New York, 1964
2. SELZNICK, P., Leadership in Administration, Harper and Row, New York, 1957
3. ABERCROMBIE, M.L., The Anatomy of Judgment, Hutchinson, London 1960
4. HSIEH, R.K.C., Evaluation of Hospital Medication Systems, Doctoral Dissertation, The School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, 1966
5. BUSACKER, R, and SAATY, T.L., Finite Graphs and Networks: An Introduction with Application, New York, McGraw-Hill, 1965
6. CONNOR, R.J., FLAGLE, C.D., HSIEH, R.K.C., PRESTON, R.A. and SINGER, S., "Effective Use of Nursing Resources -- A Research Report", Hospitals, Journal of the American Hospital Association Vol. 35, No. 9, May, 1961.

OPERATION RESEARCH IN MEDICAL WORK

By Tore Dalenius, Professor, University of Stockholm, Sweden.

This paper is based upon a report commissioned by Rådet för sjukhusdriftens rationalisering (SJURA), Stockholm, Sweden.

1. Operationе Research

In the almost 30 years that have passed since the term operations research - hereinafter referred to as OR - was introduced in 1937, OR has been given an even wider meaning. In this memorandum the term OR will be used in the following two senses:

(1) For analysis of how already existing resources can best be used for the solution of a certain task. For this type of OR the term "classic OR" will be used below.

(2) For the analysis of those resources which must be developed for the best solution of a certain task. For this kind of OR the term "systems analysis" will be used below.

OR was first used for the scientific analysis of military problems. However, during the last 20 years the OR method has been successfully applied for the solution of problems in civil life, including problems in the health services.

2. Operations research in the Health Services in the U.S.A. and England.

Both the United States and England have been and are still the leading countries in research into the direct significance of OR in the health services as well as its applications.

Among the U.S. institutes, which are active in this field, mention will be made here of the two following:

(i) Operations Research Division of the Johns Hopkins Hospital, Baltimore, Maryland

(ii) Systems Development Corporation, Santa Monica, California.

In addition, research relevant to the application of OR within the health services is being carried out at several universities, including the Massachusetts Institute of Technology (M.I.T.), The University of Michigan, Ann Arbor, Michigan, and the University of California, Berkely, California.

English institutes which have been and are active in the field considered here include the Nuffield Provincial Hospitals Trust, which deserves particular mention. Both on its own initiative and on research contracted by universities this institute has successfully developed the application of OR in health services.

3. Operations Research in the Health Services in Countries other than the U.S.A. and England

Here it should be mentioned that to-day there are many cases of the application of classic OR in health services in countries other than the United States and England.

4. Some Surveys

Surveys of various kinds have played an important part when compiling this memorandum. A number of important surveys are given below:

References:

BAILEY, N.T.J.: Operational research in medicine, Operational Research Quarterly, 1952, p.24

FELDSTEIN, M.S.: Economic analysis, operational research, and the national health service. Oxford Economic Papers, 1963, p.19

FLAGLE, C.D.: Operations research in a hospital. Contained in Flagle et al: Operations research and systems engineering. Johns Hopkins Press, Baltimore, 1960

DATA BASES AND SYSTEMS WORK

By Aksel Marchmann, Vice-Director, Copenhagen County Hospital Board, Copenhagen, Denmark.

THE PURPOSE OF DATA:

Collection of data serves one main objective, namely documentation. The documentation has above all a message for the human being within the organization.

Consequently, the matter of creating data and systems work is primarily dependent upon the knowledge of the whole structure of the organization and its information system, and on the knowledge of the methods of information, the theory of decisionmaking, and, finally, upon the placing of the human being in the total system.

So, by way of introduction, I suggest we get a closer view of the organization.

THE ORGANIZATION TO-DAY:

We are all familiar with the organization plans on which we by a graphic diagram draw up the hierarchy as a pyramid, with the board top-most, then follows the management. On an equal level we have then the functional departments which again divide themselves into several lower levels to end up with the executing levels. In between the management may also be some staff assistants who act in a purely advisory capacity.

Depending on the structure of the activity, we come back to the departments divided by objective or the functional departments.

How many levels must we have between the bottom-most ones and the topmost departments or vice versa? A number of "golden rules" apply. For the bottom-most level the maximum is six reporting departments and twenty executing ones. In a wider sense, it is a question of a "stratif cation of the organization plan" by the planning, the administrating an the executing levels.

REGULATIONS:

A supplementary part of the organization is the so-called regulations which we can consider as "standing orders". Now, how do these come into the picture?

Let us illustrate this by an example. We choose a business which has two departments only, namely a sales department and a production department. On the front door there is a sign saying "tailer's shop". The firm is divided into two separate departments which are headed by the same manager. Each department is run by a branch manager, a department manager, under whom there are one or two intermediate links and, finally we have "those who carry out the work".

Let us suppose we enter the shop (i.e. its sales department) and order a suit. The shop assistant will measure you and then hand over the order to the tailer's workshop. Suppose we had gone through "the actual line of command", it would then have been necessary for us to contact the manager himself. However, as the procedure for this kind of business has for once been decided upon, the standing orders are laid down in regulations. In other words, the regulations provide that "the shop assistant" has supervision of the tailor's workshop. Is this identical with the organization plan we drew up? The answer is no.

COMMITTEES:

In order to bridge a number of activities, to co-ordinate development and to tackle special applications we set up a series of committees. The members of the committees represent assuredly a collection of the most suitable persons. (Very often the same persons may be on all committees). Do these committees fit in with the "lines of command" we drew up? No.

CRITICISM OF THE ORGANIZATION PLAN:

Is the procedure we followed in the drawing up of the organization plan

MATHEMATICAL MODELS IN MEDICINE

By Julian Bigelow, M.D., the Institute of Advanced Study, Princeton, New Jersey, USA.

The ideas which I shall discuss are mathematical models which may be of interest and may be of use in the medical biological work. These mathematical models may need the requirements for calculated solutions using computing machines because other ways they may be too difficult or too troublesome but this is not the main point that we intend to convey. In fact in this regard we emphasize that computing machines do not necessarily solve problems in the field of medicine or indeed in any other field - they merely make possible such solutions if intelligent human beings properly proposes the problems and if they properly interpret the process by which these solutions are evaluated. There is very little question in my mind that mathematical procedures, computational procedures, if they can be brought into effective contact with the fundamental underlying problems in medicine and biology, can and will represent important tools. I would like to make a side-comment: there is a serious question in my mind as a matter of personal judgement as to how many medical people ought to get concerned with computing machines, there is a serious question for instance as if one were to take some arbitrary fraction of the people who were busy now in medicine and divert their energy into computation whether the field of medicine would be further ahead in years from now or whether it would not.

The model is an explicit quorum scheme for representing some process or phenomena of nature so that they can be better studied or understood. I think that all well-trained scientists and many who are not so well trained use models to help think about their work although very often the models are intuitive or even subconscious. Each of us by direct every-day experience forms models of the objects, the space, the physical laws of mechanics, the bahavior of fluids, etc. and even the more complicated phenomena such as the growth of plants and animals. Regarding these every-day unformalized models we cannot be very demanding as to the

accuracy and completeness with which they describe objects and processes, until the situation arises in which some detail makes a great deal of difference. For instance all of our bills are roughly alike except that the 10$ bill differs from the 1$ bill in a small detail, hence the accuracy and detailed precision we demand of a model depends upon exterior consideration apart from the actual model itself. That exterior consideration is the significance or importance of these details in connection with the exterior usage of the model. We try and build into the models those features and details we believe to be important to its function and to leave out those details which for the present contemplated use seem not to be important. In general no model is ever perfect, no grain of sand is exactly like another and the only valid model of a specific human being is the original of that human being, thus we go about our lives personal or scientific, equipped at each moment or time with a suitcase of approcimate but workable models, each cut down in precision at some arbitrary cut-off point where it did not seem worth while for us to go further in accuracy and detail, but with a bag full of models that are not very accurate there is a tendency in ordinary life and a need in progressional research to try the models as explanations for situations that are new in the sense that when they arrive the standard kit of models don't quite fit, their incompleteness becomes strikingly noticable and newer modified models become necessary. Seeing the difference so strikingly by comparison of the new situation of the old model one can often manage to construct an improved model that will closer represent the new requirements. We assert that it is upon this kind of process that scientific research and the progression of human knowledge is based.

I would like to take a moment to digress again in connection with the use of computing machines: one of the important things the computing machine can do for you is to insist that you address it by preparing an explicit statement of what you want it to do. It is both a disadvantage of the computing machine an inherent curse of the computing machine and a very great advantage and a very great power of the computing machine that the things we can say to it do not consist of loose general statements, they are highly specific, one must address the machine in a fashion as to tell it what to do from each moment to the next in terms of computation, and this represents a certain kind of an explicit commitment to a hypothetical procedure which you believe will be effective in carrying out a study in the direction of application you have in mind. There are some interesting sort of side-examples one oan cite where people have attacked problems whioh were not known to have solutions explicitly in mathematioal form and in a course of sitting down and writing instruotions for the oomputing maohine explicitly have solved this problem in a caloulated sense, the discipline of carrying out the programming has in fact taught the mathematician the answer to the problems, so he never needed to put it on a computing machine. I stress again and again this fact that the advancement of knowledge and information is of this curious sort that unless one is willing to sharpen the hypothesis of the question of the moment sufficiently sharply to be able to make an explicit statement of what appears, then one does not know where the fuzzy edges of the decision are, one does not know where the model is incompletely covered, but as soon as this is done a number of things which can test whether or not a model is sufficient and point to its feedback to come to the surface. This occurs interacting with computing machines, it occurs interacting with mathematics, it occurs interacting with carefully designed experiments, in the real world, in biology and elsewhere, but I think that the fact that the computers are forcing people in other fields, and forcing people in medicine to consider really what it is they are doing at each moment, it is one of the valuable side products of this movement. It may be that the actual computations are indeed less important then the recognition of the obligation to reduce to explicit statements what one believes one is doing at a particular decision point in investigating some pieoe of research.

We are now going to speak about not general modelling, not intuitive modelling (i.e. mathematical modelling in a far more abstract sense), but about mathematical modelling as it appears in literature involving sophisticated symbols and oomplicated equations representing prooesses and objects. I would like to roughly desoribe a kind of cases in which mathematical modelling has proved to be of some help in bio-medical work and I would like to disouss some features about them which in my opinion are responsible for the contribution that has been made.

Let me first speak of a model whioh I am sure has oome up at this meeting, namely, the problem of evaluating the state of stability, the state of health, the state of reserve of a patient who is in critical condition. Suppose we assign the patient at any given moment a point in multi-

dimensional Euclidean space, coordinates representing condition descriptors such as neurological state, cardiocirculatory state and so on. We can now try to define a safe region surrounded by fatal regions, and the problem of understanding and improving by iteration upon our knowledge of them a patient under stress is crossing over some borderline, consists exactly in taking a trial set of variables seeing whether or not one can form some kind of function depending upon this trial set of variables and collate that with actual experience in terms of survival and in terms of other medioal knowledge (so we have to establish both other variables correct or suitable) and secondly what function of them represents the best the separation of the space of survival from the space of non-survival. The only way one knows how to do this is to set up hypothetical sets of variables, set up a function which you think expresses these. (Some of them may be independent of the others and some of them may not: dependence may be very complicated). It is strictly by a cascade of events that one of the functions fall below a certain level which then enforces a second function etc. but the heart of the question of making a model is to set down some explicit hypothesis of this sort and see whether or not by interaction between the physicians and somebody who can bring the results of their observations into mathematical form we can boost up an expression which represents the survival function. It is perfectly clear that the survival function will not be a constant from one person to the other.

It is perfectly clear also that one has to entertain probabilistic measures that under a given experiment the function may have turned out to be somewhat erroneous in its description so that one has to apply to this a measure and sometimes some kind of probability invariance to take care of those things we have not got into the model. But the explicitness of the model and the fact that it stands there and either is confirmed or rejected quite radically if we are far from choosing the right kind of representation or right kind of variables, these are the keys to the process of repetition and iteration that which one oan hopefully achieve in a model of this sort which would be useful say in helping the anaesthetist who is monitoring various functions during an operation.

The representation we will have at any one minute or time or any one period of time will always be somewhat inadequate. There should always be a corrective term for outside oansative features which we have not built into the model. Therefore, the actual behavior in a given circumstance has to be tempered with some kind of judgement and experience in its use but to the extent which we can reduce it to a scheme of this sort, to the extent we can test the scheme, we can pick out a sample part of the process which is represented in a explicit way and then we can work with that sample part in actual experience in the area in which research and further development applicable along these lines should proceed.

And in fact that point is the main theme of the message I wish to convey, namely, the mere fact that the models at a given stage are inadequate is not a reason for not using them, in fact the model should be expected to explain some fairly large, some fairly significant fraction of the phenomena in question and the purpose of these models is to give you a basis for starting out a study and to study further differences.

I would now like to shift to a different kind of question regarding models. It is very well known to anybody in mathematics who tries to do anything in connection with biology that biology is an extremely difficult and complicated field for mathematics. I would like to explore for a moment the kinds of stages of evolution that the construction of mathematical models can be expected to go through. In the first place there are various kinds of complexities, one kind of complexity is that in a given situation one has a very large population of events taking place which can be treated relatively uniformly in some sense. Examples of this are examples where you have the biochemical changes taking place due to enzyme action inside the cell itself, I believe that it is correct that there are single cell bacteria in which the evolutionary period from one generation to the next is shorter than something like 20 minutes and in which it can be demonstrated that the actual number of chemical reactions that are taking place of transforming one molecular entity into another in that period of time are of the order of 108 - these figures are of course appallingly large. The kinds of mathematical models that we have to consider to carry out such processes as these and to examine them and study them, seem at first glance to be absolutely terrifying but in point of fact the existence of large populations of complex molecules taking successive straps of chemical reactions may in fact be representable by a very much simpler model in that the highly parallel nature of what is happening may be a parameter which one can represent by a single

parameter system with the model in question and one may be able to represent such an evolution and developing a very very large number of molecular objects by what is called in applied mathematics and other fields a lump representation in which you use the statistical average of the process going on in each stage in place of any attempt to follow the individual processes involved.

The role of the observations you can make is critical and one has to constantly iterate and feed these back and compare them piece by piece with the actual biological facts and the actual medical facts taking place or one is sure to get off on side-tracks from which the sense of reality is lost, but in many cases where the variety of complexity are occuring in biological processes is a complexity consisting of many parallel things which in some sense can be thought of as a large population of events, then the statistical procedures, the procedures of statistical mechanics and the laws of mass action which are useful in physics and physical chemistry and elsewhere can come to one's aid and one can write with what essentially amounts to families of ordinary differential equations in terms of the reagents present and to examine the ordinary differential equations not distinguishing one family of reactions indise the bacteria or inside the cell or over a population of cells which is going into an evolution from another but simply talking about the average one or the average reaction rate. Typically, studies have been done along these line with successfull applications which I believe may be valuable. I happen to do with things like the diffusion of chemical reagents and biological reagents introduced into humans, introduced into animals and the problem of representing the behavior of the total animal with regard to its experiment as being one of a partition set of reservoirs with diffusion processes between them and the diffusion processes between the reservoirs through membranes into the outside environment and back again - these things which are called mass actions kind of phenomena indeed are very complex but it can very well be that the important question is the questions of the statistical parameters involved rather than in the details and in many cases the statistical parameters and the statistical observables can give you what you want to know and the problem is treatable as one in physical chemistry or indeed in the statistical mechanics of gases and fluids. There are of course cases where this is not true, but in order to attack those cases where we cannot represent large populations of evolving and dependent processes by averages over those processes, it is clear that further mathematical contribution to the field in terms of the value of constructive models will be dependent upon something deeper in terms of interrelationship between the mathematics concerned and the subject field. That deeper thing in my opinion will be the devising of new and different methods of making observations and measurements. It may involve the questions of providing new instrumentation and new classes of transducers,which may not now be recognized by medical people, but which have properties of being able to sort and separate in different ways. Therefore, the second point I wish to make is that there appear to be large sample problems which certainly can be expressed in terms of the biological framework in which the explicit specificity of the reactions involved need not be a key subject under investigation but can be replaced by something like a rate constant and in which the comparison of the results of the model would be comparison of experiment in real life made possible by something which is related to the mathematical idea of the central limit theorem of probability, which roughly speaking says that even though the individual species involved in a large population of events may differ in some detail, if the population is large enough and observations are made on its average many of the detailed characteristics of the individual average out in the process of doing the operation as a measurement and of doing the operation as a computation and to this extent one can treat the thing as if they were uniform. In fact, if there were not phenomena as these one would not be able to make an estimate of the number of grains of sand in a shovelful by taking out small samples of grains of sand and make an inference of their size and distribution. In fact one can do many things in the physical world of this sort in which the detailed distribution of the individuals combined in such a way in the over all measurement that they average out to a constant which represents the typical one, and one in fact deals with these statistics. Now, of course that is only the beginning, the very most elementary foot in the doorway with regard to the processes which clearly go on in biology. Clearly, in biology there are long sequences of processes in which very explicit and very highly discriminatory reactions take place, and for these it is perfectly clear that one is dealing with problems in which not the statistical aspect, not the mass action aspects of ordinary procedures are occuring but in which one needs to trace a long sequence down, something like a topological branching tree of events and to introduce combinatorial statements about the outcome of an individual sequence of reactions.

For these lots of difficulties lie in wait for both the mathematician and the biologist or the medical person who is working on such topics, in particular the fundamental difficulty which occurs again and again is that one does not have simple ways of making observations on the individuals of the population which is undergoing changes, one does not have a way of tagging them and, therefore, one does not have a way of examining sequences of events which one may call long range order or long range sequential dependents. The lack of the experimental ability to examine a specific sequence without disturbing the conditions under which one is operating is directly related to the fact that one is in a very weak position in constructing models. The reason for this is that in fact mathematics, although it is too weak a science in some sense to be as helpful as one would like in biology, has also the property that it is too much a science, namely, if one sets down the wrong set of initial assumptions after what is going on in a process one can generate literally enormous numbers of families of possible consequences. There is no inherent way that mathematics tells you which of these consequences to select as the right one to represent the problem in mind. In general what one has is something much more like a steady state. From the steady state of distribution one has to try and refer back to see what the processes which could have produced that distribution are. This is a very much more difficult problem.

I think that the requirements which such problems impose upon the use of mathematics and the construction of mathematical models are closely linked to the requirement that the mathematician and presumably the biophysicist and presumably the medical man in question have to get together and live together for a considerable period of time in order to evolve over all the classes of phenomena to pick out in those cases where they can dream up or invent new techniques and measurements which are going to reveal relevant parameters in respect to long range or organizational kinds of problems in the subject field, but I would like to say that, however frightening a task this is, however, much may appear to be an impediment in the progress of biology and medicine to contemplate detail for detail explicit models of this sort. I personally am convinced that this road must sooner or later be followed in many important areas of work where biologists and medical people now proceed. I am convinced because the essential thing which is confronting those who wish to construct models and to make hypotheses which are capable of verification is related to the organization of the objects in biology and in medicine and that the detailed examination of this organization is the very key to understanding the process itself.

Furthermore, however weak mathematics is at representing highly complicated successive stages of organization and processes which go on for many many different kinds of sequential transformations (and as I said before it is weak because it is too rich because you can invent too many things and you can invent things which are not real). However, weak it is, it is the only science which really concentrates on this kind of process itself. I think that, however difficult it is, it is inevitable, and if one looks at literature 10, 20, 30 years from now one will gradually see more and more cases where an inload has been made, where some means of identifying and increasing successive stages has been devised which can experimentally be carried out, and where the model itself has been formulated mathematically and shown to coincide with the actual phenomena involved in such a large number of places that the probability that it is far wrong can be demonstrated to be quite small indeed.

One final remark: the role of computation in such procedures: here I would like to return to some of my first remarks, namely, that the role of computation is extremely important, but it is important possibly primarily because by carrying out computations on a model the research team in question begin to learn what the implications on the model they pose really are. On the other hand, in biology as in many areas such as the area of chemistry, the computations which nature is able to carry out, when one starts from a given set of conditions, are extremely powerful, fast, and accurate. One can produce over and over again complex reactions both organic and inorganic which have remarkable ability to pass from many stages to reproduce themselves very accurately, and in which the end products can be described with an incredible degree of precision, given that the initial conditions are specified, this means in fact that the computing machine and its operating team is likely to lag behind imperical discovery, imperical invention and imperical treatment by some considerable period of time. This does not mean that it is not making a contribution. The contribution which it can be expected to make, is a contribution to consolidating our understanding of prior wellknown phenomena which have been observed, and it is only the consolidation of our understanding of those phenomena that one can in fact hope to proceed to the investigation of new classes of phenomena.

DATA PROCESSING IN THE CLINICAL LABORATORY

By Gunnar Jungner, M.D., Associate Professor, University of Gothenburg, Sweden.

Computers have great and important tasks in the hospital laboratory. In the first place, interest has been concentrated on the clerical part of the work, as the increasing number of analyses involves problems that are difficult to overcome by ordinary laboratory operation.

A second phase of development can be distinguished, in which the computers are connected more or less directly to the analytical procedures. This was a natural step, since the instruments became increasingly aimed at automatic registration. Although the different stages of development are not clearly delimited, it seems as if still another stage must be envisaged, in which the computer intervenes to such an extent in the operative system of the laboratory that new ways of managing it must be found. The motive power is the effort to increase the production of analyses, to make the work more accurate and easier and, perhaps, above all to reduce the costs with respect to both staff and money.

Here, some fundamental viewpoints will be presented in connection with an account of some projects aiming at a high degree of automation, and in which the analytic robot and the computer are essential components. To survey the possibilities of getting help by the computer in the hospital laboratory, it is necessary to balance the laboratory requirements and the potentialities of the computer, in order to achieve a system which is flexible, effective and easily handled.

First of all, let us consider the demands of the hospital laboratory, when they are presented as the tasks required of the computer (Fig. 1). This presupposes that a considerable part of the analytic activity is automated.

NB Figures at pages 248 - 251

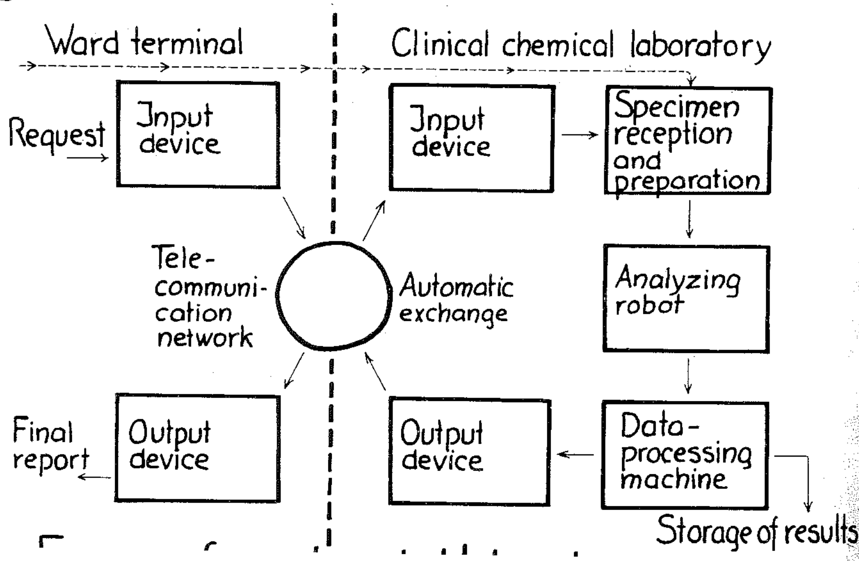
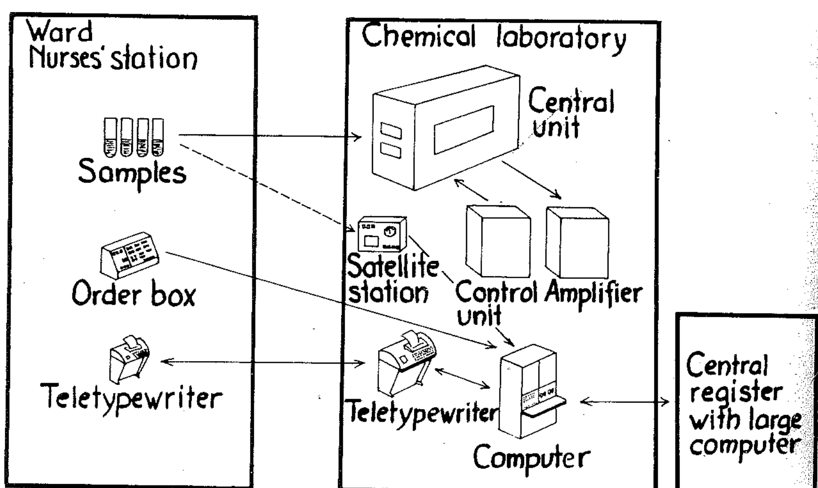


Figure 4



AutoChemist system

Figure 5

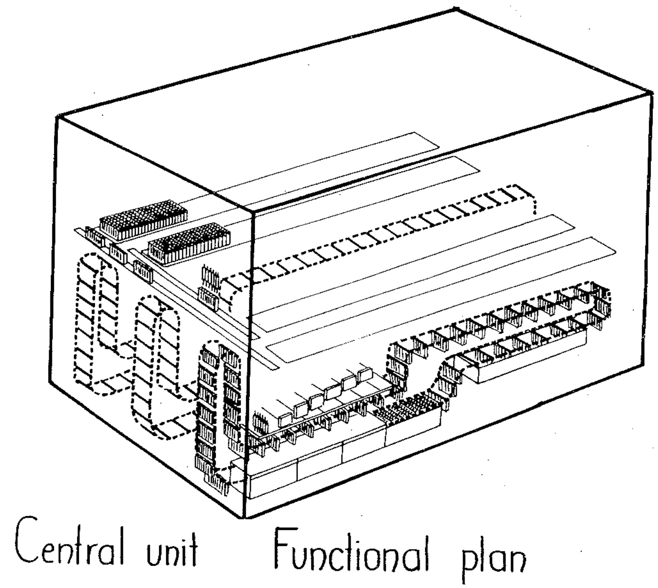
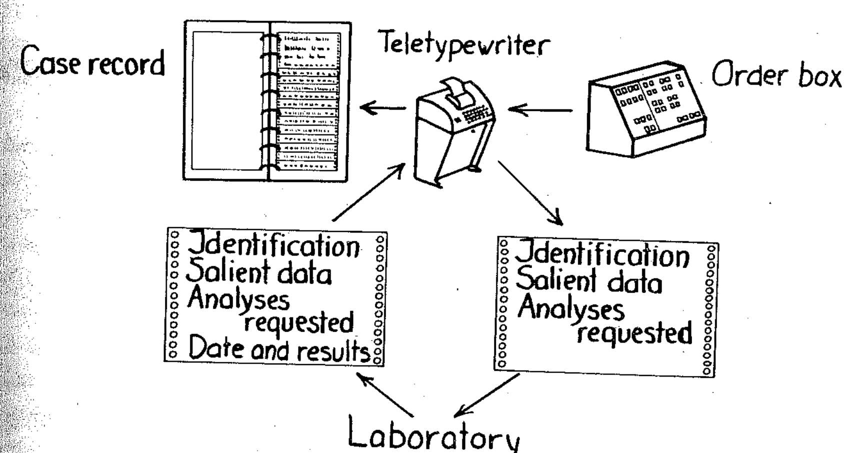


Figure 6



System for automatic laboratory request and report

Figure 7

LABORATORY AUTOMATION AND COMPUTERS

By J. Guigan, M.D., Laboratory Director, Paris, France.

(After showing a film that demonstrated how his automated laboratory in Paris works Dr. Guigan explained his plans regarding future reforms. The contribution was given in French and translated by Mr. Nielen.

I am really not too sure that I have all the details. I think Dr Guigan decided to start all over again. If you look at the last 10 years you find that actually the automatic analyzing machinery was developed not with a computer in mind (it could not have been). It was developed on the tracks of the human observer. The automatic analyzing equipment helped the human to make an easy observation before the computer was able to read what these analyzers show.

Now Dr. Guigan has made the observation that as the computer is indesistent and is on-line you should actually develop a completely different and new laboratory with different analyzing equipment, because a digital computer actually can do much better than detect peaks in curves. Dr. Guigan proposes to develop new analyzing machines each specialized for one particular analysis and designed completely from scratch. They are not necessarily auto-analyzers in the present sense of the word.

You enter the sample and the request for analysis to the left in Fig. 1. Dr. Guigan is not concerned with the trajectory before this because the samples come in his laboratory with identification and with the request for analysis. The request immediately goes into the computer. The analyzers would be on-line with the computer and not only on-line in the sense that the computer reads the analysis but also that the computer could direct the analyzer to do things. Now the sample is put on a chain at A

NB Figure at page 254

and is put into the garbage at B or into the archives which is the same thing. The chain moves in the direction shown and the computer controls the analyzers to take the sample from the chain as it passes by. I presume there is a possibility to lead the sample along C in case you want an analysis repeated or in case the result of one analysis indicates that you have to go to another analyzer that was earlier in the line. (Of course, the computer may be programmed to submit the sample to tests not requested by the doctor, depending upon the outcome of the preliminary tests.) Maybe there is a switching yard as shown at A because in order to expedite the analysis it would perhaps be more clever to sort of divert the samples to different analyzers and have them work in parallel. (Of course we must be very careful in this new type of laboratory to give the samples an absolute place in the chain and not a relative place. You see, it is very dangerous to have samples that are identified only by their relative position in some sequence. It is actually mandatory that you should have your samples in an absolute place in the chain.)

I think it would be very feasible to have the chain run at a relatively high speed. Here again the computer is essential. The computer would know the position of the chain at all times just as the computer knows the position of the printing chain in the printer to the nearest millisecond or perhaps the nearest microsecond.

Finally the computer would give out the reports and you would have a completely automated laboratory.

You would always have to do something that is not routine enough to build into this system. If the computer decides that a certain analysis which is not in the chain would be done then of course you can divert the sample at B and have it go to some special other analyses just at will.

DISCUSSION

Kirkham: We should like to work with you on this, we think that with some of the digital read-out devices the flame photometers now can be automated and run maybe 150 samples per hour. I think that there are a lot of companies that are interested in developing such special purpose automated analyzers the output of which can be put into a computer on-

DUIGAN

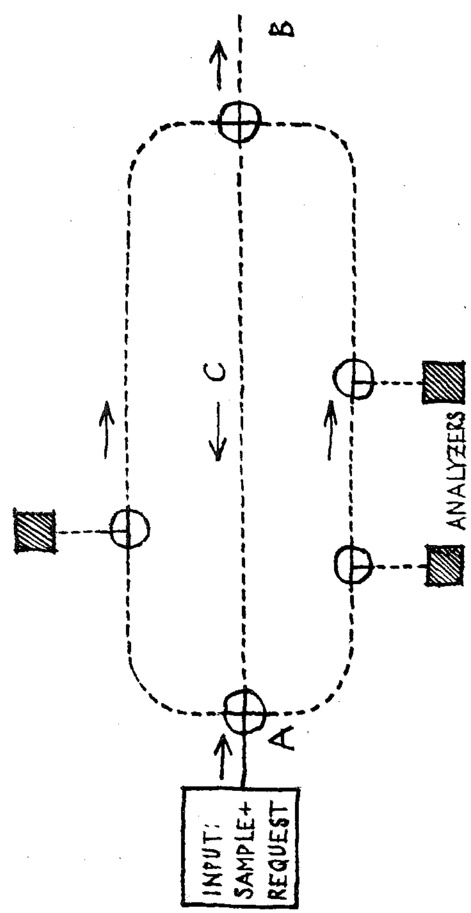


Figure 1

AUTOMATION OF HOSPITAL LABORATORY SERVICES AS PART OF A MODULAR APPROACH TO HOSPITAL AUTOMATION AND DATA PROCESSING

By William R. Kirkham, Ernest Cotlove and George Z. Williams, Department of Clinical Pathology, Clinical Center, National Institutes of Health, Bethesda, Maryland and
Scott I. Allen, Office of the Director, Clinical Center, National Institutes of Health, Bethesda, Maryland, USA.

Two techniques, automation of health services and automated data processing have great potential for assisting in patient care and extending to greater numbers of people the skills of modern medicine. Most computer installations in hospitals have been oriented toward administrative or business functions, a few have been oriented to information communication and patient care functions. The orientation of a computer system in a hospital is to a great extent governed by the interests and hospital background of the persons responsible for the development of the system. A few hospital automation programs with orientation toward patient care have been developed by hospital laboratory staffs seeking to apply new techniques to their problems. We will briefly review the several approaches to laboratory automation and data processing in the United States; the hospital laboratory automation program at the Clinical Center, NIH, and emphasize the factors to be considered in a hospital automation program. The hospital automation systems reviewed are considered only in respect to the interaction of the system with the clinical laboratory operations. There are other important aspects of hospital operations in each of these systems which are not pertinent to this review.

Several groups have attempted to plan and implement complex realtime hospital systems with input/output terminals on all nursing stations, usually utilizing one computer for all processing and communication. A few community or regional computer centers which will serve multiple hospitals are at various stages of development in the United States. At

NB Figure and references at page 266

REFERENCES:

1. RIKLI, A.E., ALLEN, S.I., and ALEXANDER, S.N. Real-Time Hospital Information Systems: A Survey. To be published.
2. PEACOCK, A.C., BUNTING, S.L., BREWER, D., COTLOVE, E., and WILLIAMS, G.Z. Data Processing in Clinical Chemistry. 11:595-611, 1965
3. SOBOTA, J.T., BREWER, D., CUMMINGS, M.E., FELD, M. and WILLIAMS, G. Z. A Data Prooessing System for Hematology: Am. J. Clin. Path. 44: 502-511, 1965
4. WILLIAMS, G.Z., COTLOVE, E., YOUNG, D.S., and KIRKHAM, W.R. A Computer System for Automatic Prooessing of Hospital Laboratory Services, a paper presented at International Conference on Automation in Hospitals, Elsinore, Denmark, April 23, 1966.

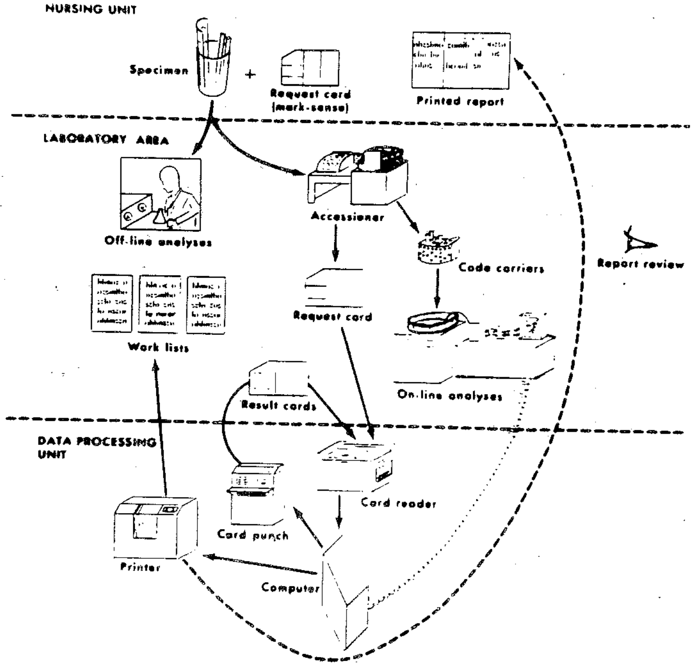


Figure 1

THE COMPUTER AND THE HOSPITAL

By Donald A.B. Lindberg, M.D. Associate Professor, University of Missouri, Columbia, Missouri, USA.

This paper will illustrate the multiple aspects of hospital medical practise which are impinged upon by a computation or information processing system. Specifically, we shall examine briefly a computer system for

1. processing hospital laboratory data,
2. the analyses which are possible from these data and their implications for laboratory quality control and physiological research, and
3. the uses of such data in education.

MISSOURI LABORATORY DATA SYSTEM

The results of clinical laboratory determinations in this institution are transmitted to an IBM 1410 digital computer within the hospital via 1092/93 densely coded matrix keyboard transmitting units located in the laboratories. The system has been presented in detail elsewhere.[1] The computer system evaluates and contributes to the results and then routes the reports through a transmission system to the remote printer which is nearest the patient' s ward or outpatient clinic.

The results of determinations in the bacteriology, hematology, and chemistry laboratories are entered on the matrix keyboards by depressing one or more of the 260 buttons provided. The meaning of each button is

---

NB Supported in part by U.S.P.H.S. grants AI-03679, HM-00374, and CD-00234.

Summary at page 271
References at pages 272 - 273
Tables at pages 273 - 275

indicated by a printed English language description appearing next to the button. The descriptions are drawn on a sheet of opaque plastic which overlies the keyboard. The meaning of each of the 260 buttons can be redefined according to which plastic overlay rests on the keyboard; the latter device senses with microswitches the notch coding on each overlay.

At the computer site, the information received from the laboratory appears in punched cards and is processed by the computer. Here a complex evaluation of the data takes place. The evaluation includes a consideration of such factors as age, race, and sex of patients; previous patient diagnoses; relation of new test results to earlier ones; expected normal range of values; relation to the frequency distribution of results at this medical center; biological properties of bacteria and antibiotics; and the pattern of multi-tests results. Valid results are transmitted directly and without human review to the hospital ward station. Individual reports are rejected by the system and are sent, along with explanation, back to the laboratory. Here the director reads the explanation and the report, informs himself of the unusual character of the report, takes whatever action is necessary to validate the report or to proceed further with the laboratory investigation of the patient's problem, and then responds to the computer system by either releasing or cancelling the report in question. All records are stored for one day on random access storage and are then sorted and merged on magnetic tape with records of previous laboratory determinations.

Through this system, the quality of the information reaching the clinician is improved and transmission of the information is more rapid than with manual systems. For example, in a three month period 389 invalid patient unit numbers or invalid specimen accession numbers were detected in the bacteriology system before reports were issued, through the use of the Modulus Eleven self-checking number system[2]. Six hundred and twenty-five invalid patient numbers were detected in hematology. On the other hand, most reports which we cause to be rejected by the computer system as dangerous or suspicious are not, in fact, erroneous. During the same three month period, 123 reports from the bacteriology laboratory were rejected as dangerous on strictly medical grounds, and 1278 from hematology. All these rejected reports were investigated by the appropriate laboratory directors and 109 of the 123 in bacteriology and 1234 of the 1278 in hematology were approved and released for reporting. In nine out of ten cases, therefore, the computer system merely served as an exception reporting device to alert the laboratory physicians to unusual circumstances.

LABORATORY QUALITY CONTROL

The advent of multi-channel analytical devices in clinical chemistry has accentuated our need for better quality control techniques. These are possible only with computer systems. A simple beginning of such controls was provided by studies with Van Peenen and Couch.[3,4] Here it was noted that the combinations of high, low, and usual values of the serum electrolytes did not all occur with equal frequency. The statistically rare or non-occurring patterns of these values as well as the analogous rare pattern of blood urea nitrogen/serum creatinine[5,6] and a variety of liver function tests[7] are included in our computer programs as "forbidden patterns." Their occurrence causes a report of multi-channel analysis to be rejected and/or the determination to be repeated. We are aware that a theoretical basis for rejecting these patterns is still lacking but rejoice that we at least have a mechanism to prevent erroneous reports from being rendered. The more subtle association of patterns of results and their physiological meanings is being investigated.

Another area with implications for quality control and medical administration is the analysis of bacteriological data. Tables I, II, and III present part of an analysis of the computer processed reports of the bacteriology laboratory.

Tables II gives the relative frequency of certain common pathogens as they appear in our laboratory reports. As might be expected the staphylococci, _E. coli_, _A. aerogenes_, _Pseudomonas_ species, and the _Proteus_ species constitute the commonest laboratory findings.

Seven thousand and eighty-four (69%) of specimens for routine culture yielded no growth. An even greater percentage of specimens for acid fast culture yielded no growth: 97% of the 1,865 specimens indicated in Table I. It would be desirable to have comparable figures from other institutions with which to compare these yields. We can now analyze our own laboratory usage as it changes in time but cannot compare with others. Amidst the frequent accusations of abuse of the laboratories by unnecessary requests, one seeks in vain for a statement of the acceptable number of specimens one should be expected to culture without growth in order to detect viable tubercle bacilli in one subsequent specimen. It has seemed to us that in our general medical institution, where known tuberculous patients must be refused admission, we do not consider it an abuse of the laboratories to have processed 97% of acid fast cultures without growth. The overall yield also seems acceptable but should be compared with other institutions.

Table III presents data which may serve as a guide to medical students and medical technology students in achieving an initial general orientation to cultural specimens and expected results. E.coli remains primarily a urinary pathogen, with occasional excursions into the respiratory tract. Pseudomonas sp. are about equally divided amongst urine, sputum, and wounds. We had had a subjective sense that Pseudomonas sp. were associated with burns in a higher percentage of cases than our data suggests. Clostridium perfringens is truly "where one finds it; " the use of anaerobic culture technique has a more profound influence on the pattern of recovery than the types of specimens. A statistical table cannot, of course, communicate the importance and the drama which the few isolations of this pathogen from wounds actually represented. However, the unit numbers of the patients associated with such isolations can also be easily recovered by the computer system and the clinical situation surrounding these laboratory problems can be recreated by reviewing their charts.

Looking specifically at the patients whose specimens yielded Mycobacterium tuberculosis, one finds that 16 patients yielded 52 isolates. The 16 tuberculous patients also were the source of 48 specimens which were cultures specifically for acid fast bacteria without growth; the number of specimens in this category varied between patients from zero to eight. In addition, during the same admissions the 16 tuberculous patients required 102 other microbiological cultures (for ordinary bacteria or for fungi). The number of these cultures varied between the patients from zero to 19.

Because the actual collection of such data has been automated, it has been possible to establish a prospective exchange of data with another university hospital for the purpose of answering some of the questions

noted above.[8] In addition the study will establish permissible ranges for these non-numerical reports and create an automatic surveillance over the quality of our two laboratories.

EDUCATION

The computer system has for some time now been extremely useful as a means of giving individual students and investigators access to the patient care data which is stored on random access disks and magnetic tapes.[9,10] During the last fiscal year 125 category-oriented inquiries to the laboratory files were processed; these excluded those inquiries which sought information concerning specific patients. Two hundred and twenty-five inquiries from staff and students were made into the computer-based hospital diagnosis files.[11,12] In all cases it has been apparent that the meaning of a single category of information about the patients in the computer file, for instance the serum potassium, is inevitably much more valuable and meaningful for education when the system makes it possible to combine this with other categories, for instance, electrocardiograph findings or blood pressure or clinical diagnosis. Questions rarely involved more than three categories of data at once.

Doubtless, purely instructional goals could also be served by such a hospital computer system. We have only begun to introduce didactic information into the system and hence cannot report on this aspect of the problem. At present it would appear ideal to blend instructional techniques into the patient care or service-oriented hospital computer systems.[13]

SUMMARY

A system for acquiring, evaluating, and processing hospital laboratory data has refined the data and speeded their transmission to the clinical physicians. Because the data are now machine processible, the laboratory director can utilize new techniques to achieve quality control. Likewise the student can combine these data with diagnoses and other computer stored information in order to help himself learn the patterns of diseases.

REFERENCES

1. Collection, Evaluation, and Transmission of Hospital Laboratory Data. DONALD A.B. LINDBERG. 7th IMB Medical Symposium, Poughkeepsie, New York, Oct. 25, 1965

2. Self-checking Number Feature, Modulus Eleven, and its Associated Self-checking Number Generator, Modulus Eleven. International Business Machines Corp. Publication Number G24-1022-0, 1960

3. The Limitations of Laboratory Quality Control with Reference to the "Number Plus" Method. H.J. VAN PEENEN and D.A.B. LINDBERG. Amer. Jour. Clin. Path. 44:322-330, 1965

4. Patterns in Clinical Chemistry: Low Serum Sodium and Chloride in Hospitalized Patients. D.A.B. LINDBERG, H.J. VAN PEENEN and R. COUCH. Amer. J. Clin. Path. 44:315-321, 1965

5. Serum Creatinine and Blood Urea Nitrogen as Measures of Renal Function: a Computer-Assisted Study. H.J. VAN PEENEN and D.A.B. LINDBERG. Missouri Med. 62:992-995, 1965

6. The Meaning of Quality Control with Multiple Chemical Analyses. D.A.B. LINDBERG and H.J. VAN PEENEN. Technicon Symposium on Automation and Analytical Chemistry, New York, Sept. 9, 1965

7. Data Patterns in Laboratory Diagnosis: Results of Chemical Screening Procedures. H.J.VAN PEENEN, D.A.B. LINDBERG, and F.V. LUCAS. American Medical Association Symposium on Successful Computer Applications, Chicago, Illinois, June 29, 1966

8. Personal communication. E. CHEATLE, D.D. TAYLOR, and J. NEFF, Dec. 3, 1965

9. Electronic Reporting and Processing of Clinical Laboratory Data: Retrieval Techniques. D.A.B. LINDBERG. Proceedings of the Information Science Symposium. 64th Annual Meeting of American Society for Microbiology, 1964. Bacteriological Rev. 29:554-559, 1965

10. Electronic Retrieval of Clinical Data. D.A.B. LINDBERG. Amer. J. Med. Educ. 40:753-759, 1965

11. Computer Generated Hospital Diagnosis File. D.A.B. LINDBERG, G.R. REESE, and C. BUCK. Missouri Med. 61:851-853, 1964

12. A General Technique for Computer Processing of Coded Patient Diagnoses. CHARLES BUCK, GARST REESE, and D.A.B. LINDBERG. Missouri Med. 63:276-279; 285, 1966

13. Learning Equipment. L.R. KINGSLAND, D.A.B. LINDBERG, and D.D. TAYLOR. Presented at the First Annual Meeting of the Association for the Advancement of Medical Instrumentation, Cambridge, Mass., July 29, 1966.

TABLE I

TYPES OF IN-PATIENT CULTURES

January 6, 1963 through June 6, 1964

Type of Determination	Number of Specimens
Routine culture for bacteria	10,194
Culture for fungi	811
Culture for acid fast bacteria	1,865

TABLE II

RELATIVE FREQUENCY OF BACTERIAL ISOLATES FROM HOSPITALIZED PATIENTS

	Number of Isolates	Per Cent+	Patients++
Staphylococcus aureus	686	9.6	389
S. epidermidis	430	6.0	373
Diplococcus pneumoniae	212	2.9	169
Streptococcus pyogenes (beta hemolytic)	129	1.8	90
Streptococcus fecalis	235	3.2	165
Other streptococci	205	2.8	154
Hemophilus influenzae	24	0.3	18
H. aegypticus	3	0.0	3
Neisseria pharyngis group	27	0.3	27
N. meningitidis	3	0.0	2
Escherichia coli	883	12.3	614
E. intermedia	180	2.5	143
E. freundii	40	0.5	34
Aerobacter aerogenes	368	5.1	253
Pseudomonas species	333	4.6	190
Proteus mirabilis	371	5.2	241
P. vulgaris	67	0.9	54
P. morgagnii	8	0.1	7
Paracolon bacilli	132	1.8	98
Alcaligenes faecalis	1	0.0	1
Salmonella species	7	0.0	4
Clostridium perfringens	13	0.1	11
Bacteroides species	2	0.0	2
Mimeae tribe	3	0.0	3
Serratia species	5	0.0	5
Bacillus subtilis	5	0.0	5
Diphtheroids	244	3.4	213
Unidentified bacteria	2	0.0	2
No growth - routine culture	7084	69.4	2165
Mycobacterium tuberculosis hominis	52	0.7	16
Atypical acid fast bacilli	1	0.0	1
Candida albicans	241	3.3	129
Saccharomyces species	9	0.1	7
Aspergillus species	7	0.0	3
Histoplasma capsulatum	4	0.0	4

+The number of specimens from which the named organism was recovered, divided by the total number of specimens which yielded growth.

++The number of unique patients from whom the named organism was recovered.

TABLE III

PER CENT OF ISOLATES FROM FOLLOWING SITES:

No. of Isolates		Abscess	Blood	Bone marrow	Bronchial wash.	Burn site	Cerebrospinal fl	Feces	Gastric wash.	Nasopharynx	Peritoneal/pleural fluid	Skin	Sinus tract	Sputum	Urethra	Urine	Vagina	Wound swab
686	Staphylococcus aureus	2.1	5.6		0.7	2.0	0.4	5.2		15.4		1.0	0.2	12.8		11.8	0.5	23.1
129	Streptococcus pyogenes	1.5	11.5			0.7				27.0		2.3	2.3	3.1		3.0	2.3	27.1
883	E. coli	1.4	2.5			0.3				2.8	1.5			2.8		71.5	3.0	5.3
368	A. aerogenes	0.8	1.0		1.3	0.5	0.2			6.7	2.1		0.8	7.6	0.2	57.8	1.3	9.2
333	Pseudomonas species	0.3	0.9		2.7	8.7	2.1	2.4		9.3	1.3			25.8		34.2		31.1
371	Proteus mirabilis	1.0	1.0		0.2	0.8	0.5		0.2	0.8	0.2	0.5		6.4	0.5	60.3	1.8	15.6
13	Clostridium perfringens	7.6								15.3	15.3						15.3	30.7
2	Bacteroides species										100.0							
52	M. tuberculosis hominis								26.9		11.4++		3.8	53.8		1.9		1.9
4+	Histoplasma capsulatum			25.0					25.0					25.0				

+ one isolate was obtained from a lung biopsy

++ includes one isolate obtained from a pericardial biopsy

EXPERIENCE WITH THE USE OF A DIGITAL COMPUTER FOR THE STUDY AND IMPROVED MANAGEMENT OF THE CRITICALLY ILL, AND ESPECIALLY PATIENTS WITH CIRCULATORY SHOCK

By Max. H. Weil, M.D., Ph. D., Associate Professor, University of Southern California, Los Angeles, USA.

Detailed analysis of the physiological data by manual methods is inefficient, both for clinical research and for practical use at the bedside. A combination of automated sensors, with analysis of their output by an on-line digital computer, offers a means of achieving a comprehensive inventory of patient condition for immediate use by the physician and the clinical investigator. An important development, in which our group shares responsibility, is the use of the output of an appropriate combination of automated devices to develop an integrated picture of a patient's pathophysiological condition on a time-related basis. For historical justification we would record that our efforts were preceded by the work of Wilbur and Derrick (1) in the development of an automatic transducer system and digital computer used off-line for simultaneous measurement of blood pressure, pulse, respiration and temperature during anesthesia.

However, techniques by which multiple signals are handled concurrently employing time-sharing methods were previously developed and shown to be very effective in space research, handling of airline reservations and industrial process and control operations (2-5). Similar techniques, in modified form, were adapted for the clinical project of on-line clinical

---

NB Supported by the John A. Hartford Foundation, Inc., New York and Public Health Service Research Grants HE-05570 and HE-07811 from the National Heart Institute.

References at pages 284 - 287

Figures at pages 288 - 289

monitoring, which is the subject of this report.

CURRENT STATUS OF ON-LINE AUTOMATION

Automated measurement and collection of biological data in a clinical setting is feasible in a number of areas. Automated analog devices are available for measuring blood pressure by the use of indirect auscultatory techniques (6-13). These devices are not entirely satisfactory for patients in shock however since the indirectly obtained blood pressure loses accuracy under conditions of markedly reduced blood flow. Analog computers have been developed to determine cardiac output by integration of the corrected area under the indicator dilution curve (14-19). Estimates of cardiac output have also been obtained by calculation of stroke volume from components of the arterial pressure wave (20-21), and by the automated technique of relating oxygen consumption to arterio-venous oxygen differences (22). Flow through large blood vessels has been measured by computerized impedance techniques in patients during surgery (23). Respiratory ratemeters based on pressure or impedance changes during chest movements are widely used, and more recently both analog and digital techniques have been applied to the acquisition and analysis of transpulmonary pressure and airflow (24). Automatic measurement of rectal and esophageal temperatures by the use of thermistors is common during anesthesia, particularly during cardiovascular surgery. Analog devices for measuring and reporting heart rate have assumed an important role in the monitoring of patients in critical-care areas. Automated analysis of electrocardiograms has provided a new measure of objectivity in electrocardiographic diagnosis (25-30), while wireless telemetry has been especially useful for continuous electrocardiographic monitoring (31). Analog devices have been used for chemical determinations such as blood sugar, lactate, pH, PCO_2, PO_2 and, independently, computers have assumed an increasingly important role for efficient processing of data generated in clinical laboratories. The contribution that automation makes when applied to each of these technique greatly bolsters the present as well as the potential usefulness of the resulting measurements. When a combination of such automatic devices is used to obtain an inventory of the patient's condition, the acquisition, computation and presentation of the resultant information is expedetiously handled by a system which includes a digital computer as the central processor.

COMPUTER-ASSISTED RADIATION TREATMENT PLANNING

By J. van de Geijn, M.D., H. Joannes de Deo Hospital, The Hague, Netherlands.

1. This conference is devoted to the investigation of the role of the computer in clinical medicine. It is the purpose of this paper to give some information about what is being done in this respect in a special field of medicine, radiotherapy, where there is a strong interplay between medicine and physics.

Before examining the role of the computer in this particular field, it may be useful to try and give, briefly, an idea of the general problems as far as they are of interest for the purposes of this lecture.

The purpose of radiation treatment is to cause, by means of interaction between ionizing radiation and living tissue, certain biological effects: destruction of malignant growths without undue damage to surrounding normal tissues.

From this definition it will be clear at once that good knowledge of the distribution of the absorbed energy is of vital importance. It is the purpose of radiation treatment planning to try and find for every individual patient, the best irradiation procedure with the technical means available.

In the present paper, we shall limit our attention to external irradiation techniques (teletherapy).

The equipment commonly used in most small and medium sized radiotherapy departments consists of conventional (orthovoltage) röntgen

NB Literature and legends at page 300
Figures at pages 301 - 304

apparatus up to 250 KV and ^{60}Co and equivalent high energy X-ray machines.

For reasons discussed below, we further restrict ourselves to the latter category.

The whole activity of radiation treatment planning by means of calculations has sense only in so far as the calculated patterns and levels are of an acceptable accuracy. The interaction between conventional X-rays and body materials is dominated by the photo-electric effect. This effect is trongly dependent on atomic number ($\sim Z^4$) and photon energy ($\sim \frac{1}{E^3}$). Consequently there is a strong difference in energy absorption per unit weight between bone and soft tissue. Soft tissue behaves to a good approximation as water, as far as the distribution of absorbed energy is concerned.

It appears to be impossible, at present, to get sufficiently detailed information about the geometrical and physical properties of bone structures in individual patients. Therefore sufficiently general and accurate calculations for conventional X-ray treatments are not possible and there does not seem to be any sense in trying to computerize calculation methods for this range of energy.

For ^{60}Co gamma radiation and equivalent X-rays the situation is much more favourable. For this range of energy, the dominant interaction mechanism is Compton scattering. This is the interaction of a photon and a free electron. In the present case, all electrons can be considered free and therefore the probability of this effect depends mainly on the electron density, which depends very little on the sort of material. This gives the very important advantage that the whole human body, except the lungs and air cavities can be treated as water equivalent. Therefore, calculation methods derived from experimental investigations in water models (phantoms) are, in principle, meaningful for most of the human body.

2. CONVENTIONAL METHODS OF CALCULATION AND IRRADIATION

The traditional basis for the calculation of dose distributions in general is the well known isodose chart (Fig. 1). This is a

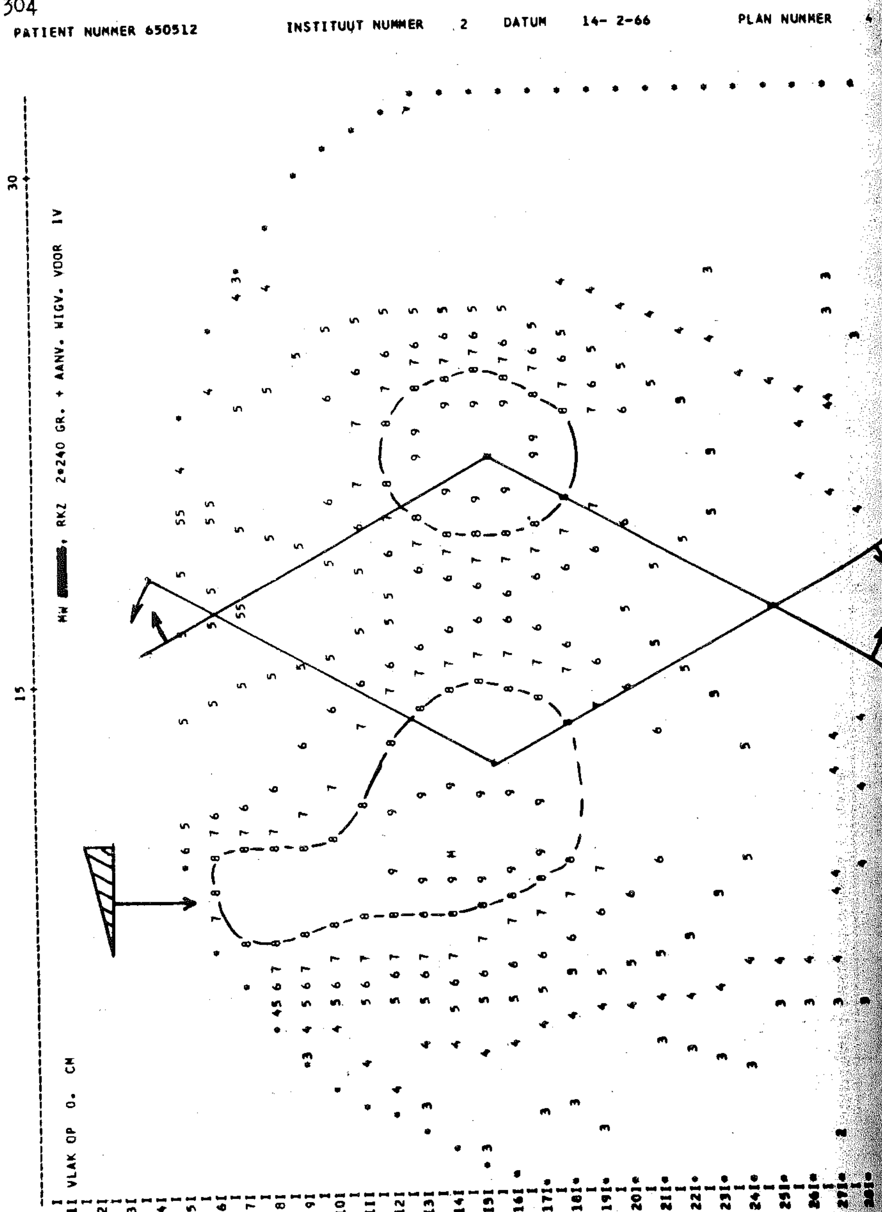


Figure 3 b

HYBRID APPROACHES TO A HYBRID WORLD

By Josiah Macy Jr, Professor, Albert Einstein College of Medicine, Yeshiva University, New York, N.Y., USA.

I. PROPERTIES OF HYBRID DATA

"Hybrid" data combine continuous analog variables with simple numerical or coded measurements. In mathematical terms, this is a combination of continua and finite sets. As a class, hybrid data includes continuous functions of time or some other variable combined with discrete numbers, or it can combine functions of time and other variables with discrete numbers. The only data excluded from this class are those which can be expressed directly as a small set of numbers or codes, and although they comprise the major part of medical records, such data can be processed by conventional digital techniques.

The one clearly distinguishing feature of hybrid data is the presence of continuous variables on which complex transformations must be performed. Thus medical data such as EEG, EKG, and X-ray films should be classified as hybrid because the raw data consist of continuous variables from which information must be derived by performing some complex transformation on the raw variables.

Some medical and biological data, such as pulse-rate, heart-rate, blood pressure, and body temperature, are effectively excluded from the hybrid class although they are basically continuous variables, because the parameters of interest are measurements taken at specific time intervals and hence can be readily expressed as discrete numbers.

II. HYBRID COMPUTING REQUIREMENTS

The problems of computing with hybrid data have involved the necessity

NB References at page 322. Figures at pages 322 - 326.

to sample and quantize them for processing in conventional digital computers. Until recently, the appropriate equipment and information about dependable techniques was not generally available and the so-called "Analog-Digital Conversion" process was, in itself, a source of many difficulties. Second, the results of such conversions are usually an all but unmanageable flood of numbers produced in an effort to approximate numerically the continuous data. For example, the numerical iterations needed to approximate continuous transformations in processing converted EEG and EKG data have been cumbersome and time-consuming even in very large digital computers. Third, results of such calculations depend, often heavily and sometimes insidiously, on the original choice of sampling rate and conversion accuracy. Although these facts suggest the use of a conventional ananlog computer for this class of data, such computers turn out to be inadequate for processes which require the use of memory, coded information, complex logical decisions or non-linear operations, as hybrid data frequently require.

Several properties of hybrid data help to establish the most effective computing techniques. The accuracy of the data is well within the range of simple analog methods, usually between 5% and 0.5%. An accuracy figure of 0.1% is achieved rarely and exceeded even more rarely. Such accuracies are low enough to permit a wide range of analog and digital methods. The range of natural bandwidths, from 100 cps to 10,000 cps, with very occasional incidences up to 20,000 cps, is also well within convenient limits for analog computation and for conventional conversion equipment.

The output format for hybrid data is another consideration in selecting appropriate computer techniques. Although a permanent record in numerical form is often wanted, this is seldom useful as interim feedback in a control situation. Interim results sent as information to the laboratory or operating room are most appropriate and most familiar if graphic, with plots or other analog displays showing results as a continuous function of time. This is particularly true when the control process requires human response and action as a result of the computation; in such cases, the response and action are facilitated by output which permits fast comprehension. For most medically and biologically trained people, information presented in analog form is grasped more quickly than the printout of accurate numbers which will serve as the eventual record of the computation. Also, if human intervention is needed to control or select action after interim calculation, turning a knob or setting a dial is the most familiar and natural way to do so. Although some of the computation is most efficiently performed by digital methods, the analog type of display and control is more "intuitive", and can be done in a straightforward manner because the required accuracy is low. Displays on a standard oscilloscope are, in many cases, sufficiently accurate.

The hybrid data, in all the cases mentioned, can be processed by solving the technical problems of adapting established standards and methods to a computer context. There is another area of clinical and diagnostic interest for which standardization, quantification, and processing techniques have yet to be defined. Interpretation of the raw data has been a human art, performed by a trained human observer in a manner that resists accurate definition or duplication by machine. Recent attempts to standardize the measurements and transformations of EEG and EKG records for computer processing seem promising. Other attempts to apply pattern recognition techniques to X-ray plates and histological slides may result in a method for processing "pictorial" material. Phonocardiograms have been picked up by microphone and recorded on analog tape in an effort to quantify some of the information usually given verbally in medical records. It is not possible to compute with descriptive data such as the subjective impressions of the examining physician. These impressions are expressed in a more-or-less conventionalized form, but the measurement has not been standardized and no procedure to relate the measurement to diagnosis or treatment has been defined.

Until the theoretical problem of establishing what procedures to use has been solved, it is naive to expect in any kind of computer system to produce useful information from the data. The technological problems previously discussed can be solved by appropriate design and programming of a hybrid computer, but there is nothing in a hybrid system that will compensate for a lack of sound theoretical and conceptual understanding of the transformations to be performed. The nature of medical and biological data make the hybrid system, with its special resources and flexibility, seem an appropriate tool for attacking the theoretical problem through extensive clinical and biological research.

III. HYBRID COMPUTATION

The hybrid methods to be discussed here consist of a mixture of digital, analog and non-standard techniques, assembled to provide maximum flexibility and to capitalize on recent advances in commercial components. Analog-to-digital conversion and digital to-analog conversion is included where needed, as well as special formats, coding schemes and programming techniques. No specific size of system is implied: small systems serving a single laboratory, large versatile systems serving many laboratories, and medium systems serving as specialized parts of a large digital computer oan utilize the proposed techniques, which impose few limits on the extent of the system and provide flexibility and simplicity for the user.

The range oovered by these techniques varies from the very effioient, inexpensive, speoial purpose equipment fully and permanently committed to a special function to the completely programmable system which is less efficient for any single job but more broadly versatile.

Flexibility costs very little in additional hardware but it increases programming complexity, for every function or parameter not wired in must be selected or de-selected by every program written.

In general, hybrid computing implies the use of the methods and equipment mentioned to operate jointly on each specifio problem. The efficient use of the system usually includes having each part perform the computations it performs most naturally. This normally means that continuous operations such as integration, differentiation, oontinuous functions of continuous variables, and frequently multiplioation and non-linear operations, are performed by the analog part of the equipment; sorting, logical decision sequences, accounting prooedures, and memory are provided by the digital portion. Control is generally a function of the digital oomputer, although control operations may be oontingent on events which take place in the analog portion.

IV. COMPONENTS FOR HYBRID SYSTEMS

Hybrid systems range in size from a few items of raok-mounted equipment with some partioular permanent assigned funotion to a oombination of

standard analog and digital computers joined by interface equipment to form a general-purpose unit of considerable size and capacity. Because of the limited aocuracy and bandwidth of medical data, modern hybrid methods usually have the capacity to process the flow of data in real time, by making the necessary calculations at a rate which is fast compared to the data rate. This not only allows various closed-loop uses of such computers, but means that much smaller requirements are imposed on digital memory than would be needed if the full set of analog-digital conversions were first stored for computation.

Several features of modern computers are especially important in the development of hybrid equipment. Recent increases in computing speed and amplifier bandwidth of analog computers has been important, as has the development of fast, solid-state control circuits. These new control circuits avoid the timing problems and response delays associated with relay controls, provide an order of magnitude inorease in computing speed, and are usually easy to work with in providing digital logic oontrol of analog elements. In addition, they make possible analog memory and the creation of fast analog transmission gates. This versatile control system with its wide range of solution speeds includes among its advantages the use of iterative solutions, forced conversion techniques, and analog time sharing. Such previously difficult problems as multiple integration, partial differential equations, and split boundary value problems can now be solved by analog methods.

An analog computer with these control circuits can easily be controlled by a small digital computer, or by special logic working on signals from biological sources. By such methods it becomes a versatile high-speed control computer for biological purposes. It can be used to perform various calculations in real time on changing biological variables, sending the results baok to the laboratory or to the digital portion of the equipment for monitoring.

Several recent small digital computers have been produced with the charaoteristios needed for hybrid work. In general, the most important of these is speed. Speeds of 4 or 5 useo for a oomplete operation, suoh as adding two words, or extraoting some part of a word, with memory cycle times of 2 useo or less, is a desirable level of speed. For data manipulation, sorting addition, and logical deoision programs, average execution times for the working part of the program should be five microseconds or less per instruction. Slower computing speeds than this make it increasingly difficult to handle the higher data rates in real time; faster speeds than this would be a distinct advantage. The more complex operations, such as multiplication or division, can be considerably slower, since in most cases in which time is critical these operations are not performed digitally.

For almost all hybrid work, a short digital word is satisfactory. For maximum speed and convenience, a binary machine with a twelve-bit word, and the binary character of the machine makes it easy to manipulate the individual bits directly. This manipulation is often essential to fast hybrid control operations. The occasional cases in which double-precision is needed in machines with a short word do not usually cost much in time or programming complexity.

A very necessary feature of a digital machine for hybrid work is fast, versatile, multiple input/output. It should be possible to carry on several input and output operations simultaneously, while continuing computation. Conventions for selecting and addressing other devices should be as simple and flexible as possible. Fortunately, many of the new small computers have been designed with the process control market in mind, and input/output operations are generally fast and flexible. The interface will undoubtedly contain elements and functions not produced as standard with the digital machine. The design of the machine will determine how difficult and expensive it is to produce and connect such devices to the digital computer.

For larger and more versatile hybrid combinations, the requirements stated above are still valid; it is useful in some cases to use a machine with a 24-bit word, but the need for speed and versatile input/output is still present. In addition, in most cases of hybrid operation, a large fraction of the programming is done in machine language, at least initially; the machine language of the chosen machine should be versatile, easy to use, and should lend itself to the easy determination of loop times and program segment times. These times are critical in many hybrid programs, and must often be determined to a higher degree of accuracy than any "average execution time" calculation can provide.

V. THE INTERFACE

The collection of components which connects the analog and digital computers to each other and to the real world, and which controls the flow of data and the conversions, is generally called the interface. This interface equipment is the basis of successful hybrid operation, and will be discussed at greater length.

One of the most necessary and important parts of an interface is the conversion equipment, analog-to-digital (A/D) and digital-to-analog (D/A). Such converters are available from several manufacturers in a wide range of speeds and accuracies. Conversion accuracies of 6, 8, or 10 bits are the most usual choices; converter speeds of up to 100 KC are useful. The relationships between conversion rate, bandwidth, frequency spectrum, aliasing error, and other variables, have been discussed in the literature. (1,2). Sample-and-hold devices and multiplexers will also form a part of the interface in most cases; they can be selected to suit the particular range of applications, and can allow considerable flexibility in number of channels and speed of sampling available. In general, it is most convenient if the multiplexer is "addressable", that is, if it can be caused to change to a specific channel under digital command. In addition to this, an independent device is often provided which is actuated by digital command, and which moves the multiplexer to successive channels sequentially after each conversion. This relieves the digital logic of the need for specifying each move of the multiplexer in cases where each channel should be sampled in turn at the same rate. The addressable feature is particularly useful when unequal sampling rates are desired, or when an event on one channel is detected, and causes the sampling of some other channel. For simultaneous conversion of several channels, it is usual to provide each channel of the multiplexer with its own sample-and-hold, so that the samples are actually gathered at the same time for all channels, even though the conversions are sequential. It is also convenient to provide an accurate master clock or source of timing pulses within the interface, and to have the multiplexer and converter timed from this clock. This makes the interface operation independent of the timing of the digital portions of the program, and allows accurate control of conversion timing.

The remainder of the interface consists of the logic components needed to

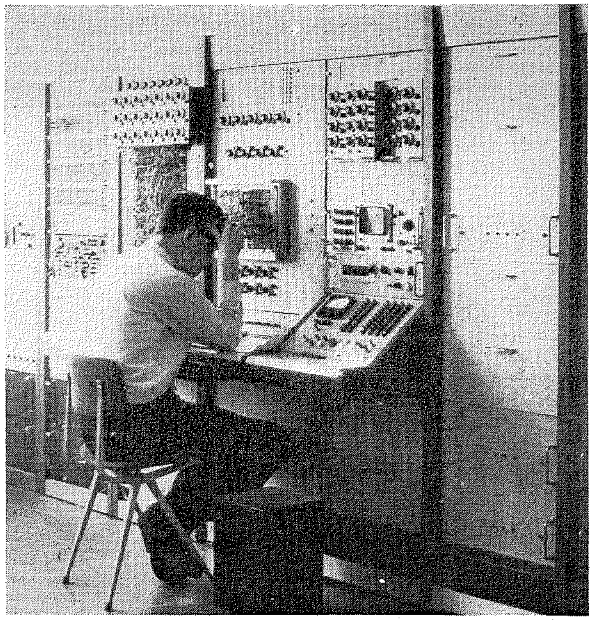


Figure 8

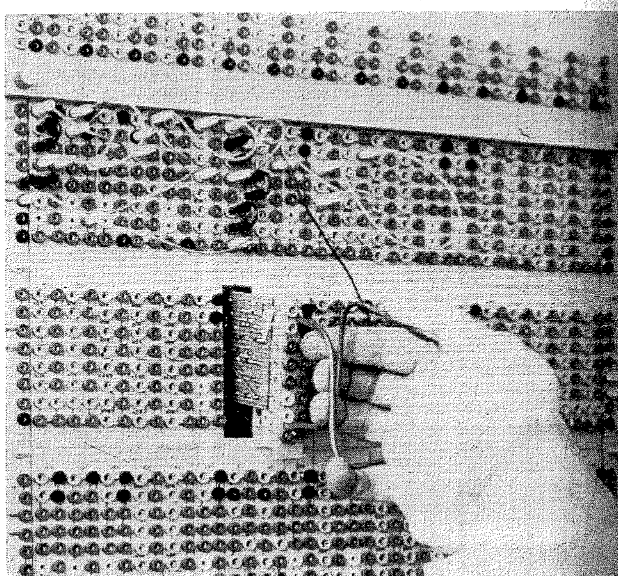


Figure 9

SIMULATION OF THE CIRCULATORY SYSTEM BY HYBRID MEANS

By D.H. Bekkering, M.D., Director, Medisch-Fysisch Institut, Utrecht, Holland.

Let me introduce myself as the head of a laboratory with a total staff of 14 qualified scientists, 11 having degrees in physics or engineering and 3 medical doctors.

This laboratory has no program of its own, but was set up to deal with problems from others, that is to say from medical institutes, hospitals, etc.

And these problems may range from telemetering electroencephalograms to detecting foetal heart beats with signal to noise ratios as bad as one to three hundred and problems such as the development of electromyographically controlled hand prostheses.

I won't go through the whole field again after the excellent speeches of Dr. Macy now and last Thursday at the conference, but I will start with some general remarks, then I will give you an idea of the work going on at our laboratory with what is becoming a hybrid computer, and then I hope there will be some further discussion about the place and possibilities of hybrid computers in the medical field.

First of all, let us be clear about hybrid computation and hybrid computers. It is possible to do hybrid computation without using a hybrid computer. Prof. Howe from Applied Dynamics once defined hybrid computation as the simultaneous use of analog and digital computing elements.

So if you have access to a digital computer and to an analog computer, you may - by interconnecting these two machines - do hybrid computations. And if you have access to a large enough digital computer - as Prof. Macy said last Thursday - then it is even possible to get rid of hybrid computation altogether. But in that case, don't talk about costs any more

10 one shot multivibrators

20 comparators

25 flip-flops (bistable multivibrators)

100 gates

This was an example of the use of a hybrid computer to simulate a complex dynamic system. - Up till now, several studies have been made on the computer on demand from outside cardiologists, for instance I remember studies concerning the right atrial pressure, about ballistocardiography, about cardiac output during catheterization, and just before I left, I was told that they are simulating now the childrens' circulation in close cooperation with and at the demand of the University Pediatric Hospital of Utrecht.

In conclusion I want to summarize the needs in the medical field for the possibility of hybrid computation. These needs can be expected first of all:

> In research situations, when studies are done of the dynamic behaviour of large, complex systems. - It is clear, I think, that in research situations you can exploit everything, and therefore also the hybrid computer.

A second possibility for the need of hybrid computation may arise in hospitals when, in the chain introduced by Prof. Flagle: observation - decision - action, there is no room for time delay.

In this connection I am thinking of the work that Prof. Weil described last week, where a digital computer was used for the management and study of the critically ill. I wonder whether a hybrid computer could do the job better.

INTENSIVE CARE OF ACUTE MYOCARDIAL INFARCTION: THE DESIGN OF A SEMIAUTOMATIC SYSTEM FOR ANALOG DATA ACQUISITION AND PROCESSING, AND SOME COMMENTS OF THE POSSIBILITIES OF DIGITAL DATA PROCESSING

By Erik Sandöe, M.D., Medical Department B, University Hospital of Copenhagen, Denmark.

My introductory remarks to the discussion of intensive care concern of the patient with acutecardiovascular disease in a department of internal medicine. I intend to focus on the problems and possibilities of analog data acquesition and processing, and on the interface between patient and analog sensors and machinery.

The most common patient with acute cardiovascular disease seen in a department of internal medicine, is the patient with acute myocardial infarction and it is usually this patient who is made the object of intensive care. For the many non-medical people being present I have to explain that myocardial infarction most often is the result of an occlusion, usually an arteriosclerotic lesion in one of the coronary arteries, i.e. in one of the vessels which supplies the heart with blood. The occlusion deprives a part of the heartmuscle from its supply of blood, and this part of the heartmuscle dies. One may say that the occlusion of the vessels leads to a wound - a sonamed infarction - in the heart. The wound may heal in the time of two or three weeks, and the patient may survive, or the wound may initiate pathological processes leading to the death of the patient. The mortality rate of acute myocardial infarction is very high. Most reports tell about a mortality rate between 30 and 40%, and most of the deaths occur in the first week of the disease (7, 12).

The following complications contribute to the high mortality rate (11):
1. serious arrhytmias, i.e. irregular heartaction, being rather so slow or so fast that the pumpfunction of the heart is abolished and the

---
NB References at page 341
Legends to figures at page 342
Figures at pages 343 - 346

Time									
heart rate									
rhythm and conduction									
arterial blood-pressure (cuff-method)									
respiratory rate									
temperature of the skin									
rectal temperature									
body weight									
pH of blood									
pCO_2 of blood									
art. oxygen sat.									
total blood-volume									
direct art. blood-press.									
central venous pressure									
cardiac output									
central blood-volume									
urinary flow per hour									
chest pain (grade 0, 1 or 2)									
clinical shock (grade 0, 1 or 2)									
pulmonary stasis (grade 0, 1 or 2)									
hepatic enlargement (grade 0, 1 or 2)									
stasis in the veins of the neck (grade 0, 1 or 2)									
Treatment:									
peroral medication: dose									
intravenous injection: dose									
intravenous infusion: time at stop total volume given									
artificial ventilation: time at stop									
cardiac massage: time at stop									
DC-defibrillation: voltage									
el-pacing: rate voltage time at stop									

Figure 4

A COMPUTER SYSTEM FOR CARDIOVASCULAR AND RESPIRATORY RESEARCH

By Mogens Jörgensen, M.D., Copenhagen County Hospital, Copenhagen, Denmark.

In March 1965 the Copenhagen County Hospital in Gentofte made a contract with IBM about the delivery of a data processing system, Model 1800, to be used mainly for cardiologican and respiratory investigations.

This system, which is to be delivered in November next, will be not only the first data processing system for this purpose in a Danish hospital, but it is also the first data processing system in Denmark to be used exclusively for medical work.

Consequently, this report can only be purely preliminary, but I hope that it will give me a chance of hearing the views of others, especially concerning the plans we have made for organizing the daily work.

It has an 1801, 1 B Processor Controller, a memory cycle time of 4 microseconds, and a memory size of eight kilowords. Obviously, a processor controller having a 2 microseconds cycle time would be more effective, but as this change can be made without much disturbance later, expenses were cut on this item. Although the system can also be delivered with 4 kilowords instead of 8, the latter was considered necessary, as 8 kilowords is the minimum core storage capacity which permits time-sharing and monitor operation. The choice of model 1800 instead of model 1802 excludes the attachment of a tape unit. We felt that this restriction was acceptable. A disc system is attached, and we feel that this provides us with a unit which to some extent will replace the ability of the tape units for storage of large amounts of data and programmes.

As you will see we have prefered a card-read-punch system to a cheaper paper tape system.

will be necessary, when the problems start presenting themselves, to refer as many off-line programs as possible to evening and night hours. During the daytime we can, for instance, by means of an anlog tape, temporarily store several input signals received at the same time and then later treat them one by one.

I feel convinced also that a certain data reduction may be done in many cases even before the signals reach the computer. This data reduction might perhaps be carried out by small analog computers placed in the individual rooms. At the outset, when the computer time will be rather ample, there is less need for these units, and we have therefore tried to obtain a digital computer as large as possible, but we are aware of the fact that an extension of this nature may become necessary.

Lastly, as you know there are fairly wide possibilities of expanding the computer system proper. Internal and external storage may be increased and the computing ability increased by reducing the cycle time from 4 to 2 microseconds.

I have tried to convey an impression of our future data processing system and the way in which I imagine it will be operated.

Of course, the size of the computer system that we could obtain was decided by political and economic regards.

I am pleased that we succeeded in obtaining a computer system with a wide flexibility and with a possibility of using high-level programming languages. On the other hand, I should have been rather uneasy if we were to start with a much larger system, as this would have made the distance between the investigator and the computer system correspondingly greater. I hope that a system of this very size is going to supply us with the best conditions for learning to utilize electronic data processing for medical research.

TIME SHARING A COMPUTER FOR BIOMEDICAL RESEARCH

By Homer R. Warner, M.D., Ph.D. and T. Allan Pryor, M.S., Latter-day Saints Hospital and University of Utah, Salt Lake City, Utah, USA.

The term time-sharing as applied to computers has come to mean a variety of things. For purposes of this presentation I will define time-sharing as a mode of computer operation in which more than one computer program occupies memory space in the computer at any given time and control of the central processor and input/output devices is switched back and forth among these programs in such a way that the users of the programs operating from remote stations appear to have immediate access to the computer for their needs.

Before launching into a discussion of the means by which time-sharing is achieved in our laboratory and some examples of programs operated in this mode, let us first consider some of the reasons why time-sharing of a computer might be a useful thing for biomedical research.

Although experiments are carefully planned and more often than not the observations which are made are largely those that were anticipated, important advances occur when observations are made which were not anticipated. If these observations become apparent only after the experiment is complete, it may be necessary to repeat the experiment in order to find the cause of an unexpected findings. If, however, the results of the data analysis are made available while the experiments is still in progress, it may be possible to determine the cause of the unexpected findings by either repeating the observation or by observing other circumstances present in the environment at the time. Thus, in the time-sharing mode, it becomes possible to feed back to the observer almost immediately the results of calculations made to derive indirectly information about the system he is studying which cannot be obtained through observation of variables which can be recorded directly. One example of

---

NB Figures at pages 360 - 361

this kind of immediate feedback used to control an experiment which involves the display of an amplitude histogram of nerve action potentials which form the basis for judging whether the recordings are being made from a single fiber or multiple fibers of the nerve.

The second advantage offered the medical researcher by time-sharing a computer is that he is able to minimize computer idle time. Because it is usually not necessary to sample events taking place in the real world at the maximum rate the computer is capable of, the computer would be spending most of its time waiting for time to elapse so it could take the next sample if only one experiment were being processed at a given time. Furthermore, the time required to get the required data on the line is often much longer than the time occupied by the sampling process once it is begun.

Another very important source of computer idle time occurs during the solution of a problem when the investigator is making a decision regarding the next set of parameter values to use in the manipulation of a model. During this time he is examining graphical representations of the previous solution and comparing these to still earlier solutions or to experimental data. It is important that he be allowed to do this in a leisurely fashion without the pressures that would result if he were completely occupying the services of a powerful computer. In the time-sharing mode his program is being swapped in and out of memory at his request and the pressure to occupy what might otherwise be idle computer time does not push him to make hasty decisions. The computer is there and ready to give him a solution when he asks for it whether this be in the next five seconds or next week.

Not only is it important to provide access to the computer for the investigator when he needs it but also where he needs it. Multiple input/output stations are essential if the computer is to be used by more than one investigator at a given time. Locating these input/output stations remote from the computer provides the user with another important advantage, the opportunity to work in his own environment. Depending on the nature of the job, this environment might be a laboratory specially equipped for certain types of experiments on humans or experimental animals, a patient room, intensive care ward or operating room in the hospital, or in the investigator's own office where he has available to him

all pertinent material relating to the model he is studying and, more important, a quiet atmosphere conducive to concentration and creative thinking. It is obvious, I think, that the usual hurried, noisy environment surrounding a computer is not the ideal situation for creative intellectual activity. As was once stated by R.W. Hamming, "The purpose of computers is insight not numbers." In a similar way it might be said that the main purpose for using a computer in a time-sharing mode is to optimize the chance that the user will gain insight into the system he is studying and maximize the number of users that can be provided this service for a given investment in computer equipment.

SALT LAKE CITY HOSPITAL DP SYSTEMS

To accomplish this in our own medical community, we have designed and implemented a time-sharing monitor program called MEDLAB which is now servicing 16 remote stations in three hospitals in Salt Lake City and has been duplicated to service 12 laboratories at the Mayo Clinic in Rochester, Minnesota.

THE HARDWARE CONFIGURATION

The central processor is a Control Data 3200. This machine has a 24-bit word and a basic cycle time of 1.25μsec. It has three 15-bit index registers and 64 24-bit high-speed registers. One of these registers is a 10KC clock and another is a clock interrupt mask. The MEDLAB monitor uses this clock to control analog-to-digital and digital-to-analog conversion rates for the user programs. The machine has floating point hardware. The core memory size is 32,768 words. There are five input/output channels. Channel 0 connects to a card reader which can read punched cards at the rate of 1200 cards per minute. Channel 1 is connected to two controllers; the tape controller controls three tape drives, each capable of reading or writing 120,000 characters per second; the other controller is connected to two IBM 1311 disc drives. The read interface consisting of the A-to-D converter and direct digital read of 12-bits is connected to channel 2. The analog-to-digital converter will convert a voltage between -1.5 volts and +1.5 volts to a number between 0 and 255 100,000 times per second. The input voltage to the analog-to-digital converter is multiplexed through 32 channels. The multiplexer and the sampling rate are controlled by the program. Likewise, the direct digital input may multiplexed under program control and is used by the monitor to read four digit octal switches on each of the remote stations.

Channel 3 is connected to the write interface which consists of both a digital-to-analog converter and a 12-bit direct digital output which is used to control banks of relays and solid state switches. The D-to-A converter also operates up to 100,000 conversions per second and is multiplexed through ten channels at the present time. Two channels of the D-to-A converter are used to write alphanumeric characters on the face of memory oscilloscopes at the remote stations and for plotting graphs on the scopes. The other analog channels may be fed to other analog recording devices or used to furnish input voltages to an analog computer for hybrid computation.

Channel 4 is connected to a printer which prints 1000 lines per second. Information to be printed is stored by each program on disc until that program has run to completion. At this point the monitor waits until the printer is not busy and then prints the entire output from that program. This arrangement avoids having the output from one program interspersed with the output from another. The disc-to-print operation is completely buffered and proceeds in parallel with input/output on other channels and computing by the central processor.

One type of <u>remote station</u>: The basic unit is a Tektronix oscilloscope equipped with a memory or long persistence tube. Information is written on the tube in the form of alphanumeric characters and graphs by controlling the X and Y axis of the scope through the D-to-A converter and blanking and unblanking the beams to limit the write-out to just that station for which it is intended. Once a message is written, it will remain on the face of the tube for several hours or until erased by closing a relay in the oscilloscope under computer control.

We have built into the oscilloscope some additional equipment. First, there is a set of eight lights which indicate at the remote station the status of the computer and the status of the program being run from this station. Second, there is a four digit octal switch which is used by the operator to call programs, to specify program options and to enter directly digital information. Finally, there is an interrupt button which, when depressed, interrupts the computer and sends a unique interrupt

code which allows the MEDLAB monitor to identify the station. Each such station has three analog channels assigned to it, but may use additional channels when available.

From <u>another remote station model,which has been developed for use in</u> remote hospitals which do not have direct lines to the computer but communicate over telephone lines, analog data is transmitted using frequency modulation over special 15KC telephone lines. Three analog signals are fed to three voltage controlled oscillators and the resulting carrier frequencies are multiplexed over a single input line. Also, the digital information is transmitted by using a very narrow band tone to represent each bit. Two telephone lines are required, one for input and one for output since these must take place simultaneously. The permissible bandwidth of the signals being fed in over the three analog channels is from 0 to 200 cycles per second which is high enough to transmit most physiological data except neurophysiological data such as nerve action potentials. Even here, however, the system is useful if a certain amount of analog preprocessing is done at the local laboratory. With this remote station, the four octal numbers are entered using a pushbutton telephone type switch and are accumulated in a register where they are displayed to the operator allowing him to check the number before he interrupts to transmit it to the computer. The subroutine called SCOPE, which writes messages at the remote station, is controlled by the clock to write at a slower speed on these stations which rely on telephone lines. Thus, the slower write-out speed is the only difference in mode of operation which is noticeable to the operator between the two types of stations.

<u>MEMORY ALLOCATION IN THE COMPUTER</u>

Approximately 5000 words in the top of memory are occupied by the MEDLAB monitor itself. Many of the routines which constitute the monitor are available for use by the individual user programs. These are written using re-entrant coding so that all intermediate information computed in the subroutine is stored in a buffered area in the user programs. In this way one user may interrupt another who is in the midst of using one of these subroutines without causing difficulty when the original user returns to that point in the subroutine where he was interrupted. Furthermore, the use of re-entrant coding permits this part of memory to be protected under hardware control since it operates in a read only mode.

Approximately 11,000 words are available for real-time user programs. These programs are loaded into memory from disc when requested by a user from a remote station. The loader relocates them in memory in whatever part of this memory is available at the time. These real-time programs usually are not large and are usually concerned primarily with controlling input and output from a remote station.

Each Fortran program is assigned a priority one, two or three. Priority one programs are programs which communicate through the A-to-D converter through a link with a real-time program, but a priority one is swapped out on to the disc at any time during its operation that its clock is not running to input data through the A-to-D converter or that computation is not actually taking place in the Fortran itself or in the real-time program with which it links. It is then swapped back into memory when the next interrupt occurs from that station.

Priority two and priority three Fortran programs do not communicate with real-time programs and differ from one another only in terms of their running times. Programs which require more than three minutes to run are assigned priority three. The operation of assembling and compiling programs is considered priority two. In the system as now running, one priority three program, one priority two program and up to four priority one Fortran programs may be active at one time.

DEBUGGING A PROGRAM FROM A REMOTE STATION

I would now like to describe two user programs which operate under the MEDLAB system and themselves are very general and in no way oriented towards any one research problem. The first of these is a program which permits the user to debug another program which has not been checked out by running in a controlled fashion and examining the contents of registers and memory locations without stopping the computer. He does this from his own office or laboratory.

After calling the DEBUG program from the remote station, the operator specifies which program is to run under it. A message containing eight

options appears on the memory scope at the user's station. By dialing a number from 0 to 7 in the low order digit of his switch and pressing the interrupt, he chooses one of these options. Two of these will be illustrated. Option 3, the read storage option, allows the user to look at the contents of any storage address in memory as well as four addresses preceding and four following the specified address. The addresses displayed are apparent addresses as they appear on the listing of the program at the time it was compiled or assembled. In fact, the program is relocated in memory at run time, but the user need not concern himself about the real address. This option is particularly useful in programs doing pattern recognition where the contents of memory can be checked against the decisions made by the program at any stage of program execution.

If, on the other hand, the operator chooses option 1, he is then asked to indicate the beginning address and the breakpoint address he desires. In this way he may begin running his program at the address specified and when the program counter reaches the breakpoint address, program execution will stop and the contents of the registers will be displayed to him. At the top are the contents of the two arithemetic registers, A and Q, the program address counter, the next instruction to be executed and the contents of the three index registers. If the breakpoint address, in his octal switch is not changed, pressing the next interrupt will cause the next step of his program to be executed and the new contents of the registers to be displayed.

In this way program execution can be carried out one step at a time or a breakpoint can be moved further down in the program and all the intervening instructions will be executed before the next breakpoint is reached. At any time the operator may dial all zeros which takes him back to a display of the options once again.

It should be emphasized that this debugging operation can be carried out from a remote station without interfering with execution of other programs which are taking place under MEDLAB. Fortran or assembly language programs can be manipulated in this fashion. Another useful feature of the DEBUG program is that on reaching a breakpoint, the clock interrupt from that station is turned off so that inputting of data will not continue after the program has been stopped. Other options permit writing into storage to change the program or to alter data in the program and options which permit reading of the high-speed registers and control of special relays.

MANIPULATING A MATHEMATICAL MODEL FROM A REMOTE STATION

Until recently the analog computer has had certain definite advantage over the digital computer in terms of the ease with which parameters of a mathematical model could be modified and the speed and convenience with which graphs of solutions could be manipulated and displayed. A general subroutine called MODEL has been written for operation under MEDLAB to facilitate these operations in a digital computer in a time-sharing mode. The use of the digital computer, given facility in performing these operations, has advantages over the analog in terms of accuracy, use of a stored program and remote access. Fig. 1 shows the MODEL subroutine as it appears to the programmer who will write his mathematical model in Fortran and use this routine to enter parameters and display solutions as graphs on the memory oscilloscope.

Any number of equal length arrays may be set up by the programmer for later display on an oscilloscope using the MODEL subroutine. Variable number one, in this case PRES, is equated to X, variable two, FLOW, is equated to $X^{(501)}$ since, in this case, each variable to be displayed is dimensioned at 500 and X is dimensioned equal to the sum of all these variables. The call list for the MODEL subroutine includes the following: K, which may be equal to 1, 2 or 3, depending upon which option is desired. When K is equal to 1 or 2, parameters may be read out on to an oscilloscope or entered in via an octal switch to the parameter array. With K = 1, parameters are read initially from cards. PARA is an array of model parameters dimensioned at 100. NVAR is the number of different variables in the data array X. NPTS is the number of data points in each variable array and INC is the increment between points to be used on the printer plot-outs when this option is called for. At the end of the program the MODEL subroutine is called again with K equal to 3. This initiates a sequence of inquiries and responses at the remote station to specify the options for the plotting of model solutions and experimental data on the memory scope.

At this point the operator may examine the current values of any of the

parameters in his parameter array, enter new values for any of these parameters or call for a solution with the values then in the array. Initial estimates for parameter values may be read in from punched cards following compilation of the program and some values in the parameter array may have been calculated as a result of previous solutions. To look at parameter 31 the operator enters this number and presses the interrupt. When he is ready to solve the equation he dials a 4 in digit one and presses the interrupt. When the solution is complete the message shown in Fig. 2 appears. The operator may ask for a plot of any or all the numbers in any given variable array. To start with the first variable he dials 1 and interrupts to get a message asking for the display code. The first two digits in the switch indicate which variable is to be plotted and the last digit causes the variable to be plotted unscaled, scaled to the lower half of the scope, the upper half of the scope or scaled to the full scope range. On pressing the interrupt the graphs will appear as shown in Fig. 3.

This plot is a graph of aortic pressure generated by a theoretical model of the circulation. This involves the solution of 18 simultaneous differential equations. The plot may be held on the lower half of the scope, new parameters entered and the new solution superimposed on the old for comparison. Likewise, the experimental data which has been stored on disc or magnetic tape may be read in by the program and plotted over the theoretical solution representing the same variable. Examples of the use of this program in our laboratory are many, ranging from the representation of the relationship between pressure in the carotid sinus and the time-course of frequency of firing of a single fiber on the carotid sinus nerve to the simulation of kinetics of granulocyte turnover in blood and bone marrow and comparison of solutions obtained against the time-course of specific activity of labeled cells in those two organs.

This subroutine gives to the user the same close interaction with his program solutions and experimental data that he has with an analog computer as well as all the flexibility for scaling, parameter and program storage, and accuracy of the digital computer under conditions conducive to concentration and interaction with real data.

MODEL PROGRAM AS SEEN BY PROGRAMMER

Equivalence (X, PRES), (X (501), FLOW), etc ----

CALL MODEL (K, X, PARA, NVAR, NPTS, INC)

where

K is option 1, 2, or 3

X is data array

PARA is array of parameters

NVAR is number of variables in data array

NPTS is number of data points for each variable

INC is increment between points on printer plotout

Figure 1

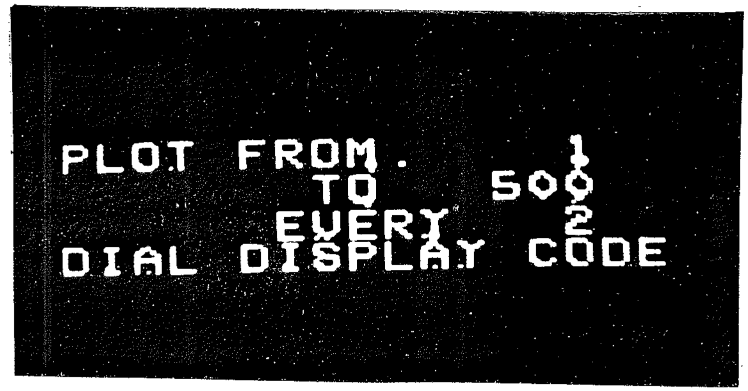


Figure 2

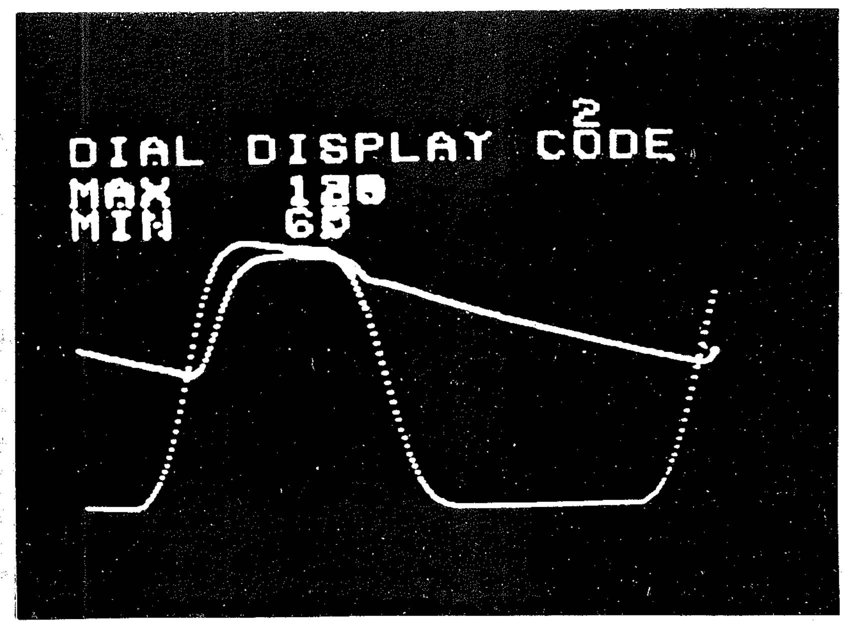


Figure 3

SOME TECHNIQUES FOR COMPUTER ASSISTED DIAGNOSIS OF HEART DISEASE

By Homer R. Warner, M.D., Ph.D. and William M. Stauffer, M.D., Latter-day Saints Hospital and University of Utah, Salt Lake City, Utah.

Two programs have been developed to aid in the diagnosis of congenital and acquired heart disease. The first accepts information collected from the history, physical examination, electrocardiographic and X-ray data and fed into the computer as answers to multiple choice questions presented to the physician on an oscilloscope at a station remote from the computer. The second is a program which accepts directly via an analog-to-digital converter the output signals from pressure transducers, oximeters and electrocardiographic leads during heart catheterization.

A DIAGNOSIS PROGRAM FOR CONGENITAL HEART DISEASES

Fig. 1 is a photograph of the face of the oscilloscope on which has been displayed the instructions to the user for inputting symptoms into the diagnosis program. As the symptoms are presented on the oscilloscope with the number opposite each symptom, the user dials the number representing the symptom if it is present in the patient in question. This will cause an asterisk to appear opposite this symptom as confirmation that it has been entered. If the symptom is not present, he does not dial that number. If for some reason he does not know whether the symptom is present or not, perhaps because the X-rays or ECG were not available, he then dials that symptom number in the last digit and a one in the first digit of the switch which will cause the symptom to be omitted from consideration when the subsequent diagnostic probability calculation is made. If an error is made in entering a symptom, the physician may dial a six in the first digit which will cause this page of symptoms to be presented again and the error corrected.

Fig. 2 shows the first page of symptoms. Symptoms which are listed with

NB Figures at pages 370 - 374

single spaces between lines represent mutually exclusive sets and only one of such a set may be entered. The patient's age is classified as less than one year, between one and 20 years or over 20 years. Cyanosis is classified into one of four categories, if present. Mild, severe, which means with clubbing, intermittent or differential, that is, cyanosis of the lower half of the body only.

Fig. 3 shows the second page of symptoms. Murmurs(i.e. abnormal heart sounds)are classified according to their point of maximum intensity(e.g. apical) and the time-course of intensity during the heart cycle. (The number of murmurs shown to yield significant information of diagnostic value is small compared to the number originally included in the program.) A physician may choose to classify a murmur as mid-systolic, or holo-systolic but if he cannot be certain, he may use the classification systolic which does not have the same diagnostic power as the other two. However, it is better to use this than to make a mistake and misclassify the murmur. The phonocardiograph has proven to be very useful as an objective means for determining murmur characteristics and we rely almost completely on this at the present time.

In this way the physical examination is continued. When it is completed, three pages are presented asking for information derived from EKG and X-ray.

On dialing a zero after display of the last page of symptoms, the computer calculates the probability that the patient represented by the symptoms just entered has any one of a series of 33 congenital heart diseases and displays the results on the oscilloscope at the remote station as well as printing them on the printer at the central facility. This is shown in Fig. 4.

At the top are written the code numbers for the symptoms entered, providing a check on the data used for the calculation. Symptoms omitted appear after the slash. In this case the computer calculated the probability that the patient had a ventricular septal defect with infundibular pulmonary stenosis at .63, with valvular pulmonary stenosis .27 and 3, 2 and one percent for three other defects.

Fig. 5 shows the equations by which these calculations were made. This equation of conditional probability is a modified form of Bayes' rule. It states that the probability that a patient has disease one given the fact that he has the set of symptoms indicated by $a_1, a_2, \ldots a_n$ where a_i assumes 1 or 0 depending upon presence or absence of symptom S_i in the patient, is equal to the apriori probability of disease one in the population under consideration times the product of the following terms over all symptoms. The term P_{S_j} given D_1 is the probability of the j^{th} symptom occuring in the disease one, is used in the calculation, since the other term (1 minus this probability) is raised to the zero power and is, thus, equal to one. On the other hand, if this symptom is absent, a_j equals 0 and the complement of the probability will be used in the calculation.

The program includes provisions for appropriate handling of mutually exclusive symptoms. Of course, the doctor is not always able to assess the presence or absence of a symptom with the certainty implied by a binary decision, that is, one or zero. For this reason a modified form of the equation, shown in Fig. 6 was developed.

In this form of the equation, the "a" values, instead of assuming only one or zero, may assume any value from 0 to 1 and represent the estimate by the physician of the probability that the symptom is present. Notice that in the case where a_j equals one or zero the equation assumes a form identical to the equation just presented. Also, if an "a" value equals .5, indicating equal probability that the symptom is present or absent, this symptom will have no influence on the diagnosis since the probability of that symptom and its complement are weighted equally. An additional dividend is achieved by this new form of the equation in that it is now possible to input any data that can be represented on a probability scale from 0 to 1. For example, a program is being used in our laboratory which calculates the probability that a given electrocardiogram belongs to each of ten electrocardiographic patterns. Using the equation shown here, the output of this ECG program can be used as input to the diagnosis program making it possible to integrate computer generated probability data concerning the electrocardiograms with physician generated probabilities concerning symptoms derived from history, physical examination and X-ray.

Fig. 7 shows still another approach to the quantitation of data for input to the diagnosis program. This approach was developed for diagnosis of hematological disorders from laboratory data which may assume any

value on a continuous scale. Consider here, for example, that X represents the volume of packed red blood cells or hematocrit and each of these three curves is the distribution of hematocrit values in patients with three different diagnoses. One is normal, two is polycythemia vera and three, chronic myelocytic leukemia. The problem is to convert a numerical value for hematocrit to a number between 0 and 1 which represents the probability that a hematocrit value in the range represented by this particular value would occur in a patient with each of these three diagnoses. This is expressed in the equation which states that the probability of a hematocrit value between D and E in a normal patient is given by integrating the distribution curve, number one, over these limits. Using this approach, it is necessary to store only the parameters of the distribution function for each symptom in each disease and not the explicit probability of each symptom category in each disease as these can be calculated from the mean and the standard deviation of the distribution function. The parameters of these functions are calculated by sending a normal or logged normal distribution curve to the collected data on many patients.

HEART CATHETERIZATIONS

The second diagnostic program is designed to accept the input of patient data directly from transducers at the time of heart catheterization, to permit the physician to edit this data, to test the data against standards of normality and to print a complete report, including a list of the abnormal findings. This program runs under the MEDLAB time-sharing monitor and can process data from at least four heart catheterization procedures simultaneously (Fig. 8).

Three types of data transducers are connected to the patient. One ECG lead chosen for its prominent R wave is used as a reference source for averaging pressure waves. A strain gage pressure transducer and a Wood cuvette oximeter are used to measure pressure, oxygen saturation and dye concentration. The output of the red cell and infrared photocells are fed independently through operational amplifiers into the A-to-D converter. All variables are scaled in the amplifiers to cover a range of -1.5 to +1.5 volts. Once the gain and bias on these amplifiers has been set they are not touched from one day to the next but calibration signals are easily put through the whole system each morning to test its performance. The total set-up time should not occupy more than five minutes and is relatively free from human error since the computer instructs the operator in the performance of each step. Transmission of the data from the amplifiers to the A-to-D converter multiplexer may be over direct wires from an adjacent laboratory or over telephone lines from another hospital.

Results of calculations made on the data are fed by the computer through a digital-to-analog converter to control the X and Y axis of a memory oscilloscope and present as alphanumeric characters and graphs. A message, once written, is retained on the oscilloscope until erased under computer control.

The physician, during the catheterization, controls the sampling of data and all other computer operations with a plastic wand which is sterile on one end so that he can handle it without contaminating himself while he is manipulating the catheter. He uses this device to change the readings on the octal switch and press the interrupt button. Each patient, is assigned a unique area on the magnetic disc where data accumulated during the procedure on that patient is stored. At any time the physician may call for a listing on the oscilloscope of the data accumulated on his patient for review. He may delete or modify this data if corrections are necessary. At the end of the procedure he may obtain any number of copies of a printed report.

At the beginning of each day a check is made on the instrument. This check is called for by dialing 00 in the code representing the catheter position. The previous day's saline readings are first displayed and the computer requests a new saline reading and tells the operator how to set the switches prior to pressing the interrupt. When saline is in the cuvette the operator presses the interrupt, the outputs of the photocells are sampled. The switches are then set as instructed by the computer and the output again sampled to give another check on the system, namely, the black level.

A subprogram is used for analyzing indicator dilution curves. In patients without shunts, this is used for measuring cardiac output in various physiological states. In patients suspected of having right to left shunts, selective injections are made into pulmonary artery, right

ventricle and right atrium and the resulting curves are displayed following each injection and in the end, superimposed. In this way the location of a right to left shunt can be determined.

Fig. 9 shows an indicator dilution curve displayed back on the oscilloscope with the extrapolation made by the computer program displayed along with it to provide the physician with some basis for evaluating the accuracy of the extrapolation process. The program finds that part of the descending limb which best fits an exponential and extrapolates using the slope of the exponential as measured from this part of the curve. This curve was obtained to make this slide from a dog being used for another experiment and, thus, the calibration values were not correct causing the cardiac output to appear very low. The other numbers on the slide represent an index of mitral insufficiency measured from the shape of the curve, the appearance time, the build-up time, the mean circulation time and the calculated central blood volume. Calculation for determination of cardiac output by the indicator dilution method is done by pulling undyed blood through the cuvette, mixing this with a known concentration of dye and pushing the dyed blood back through the cuvette.

It is possible for the physician during the catheterization to review the data he has collected up to any point in the procedure. This data is read back from the disc and displayed in the format shown here. The lefthand digit is the state of the patient, in this case a zero meaning the patient is at rest breathing room air. The three digit code following this indicates the location of the catheter tip followed by the pressure or oxygen saturation values obtained. He may bring up additional data or delete an item on the report.

Other codes allow him to add or subtract a constant from all the saturation values or to change the location code for a given item. For instance, he may have withdrawn a blood sample thinking it was right atrium and find the saturation to be very high, indicating that the catheter was across an atrial septal defect and the sample was actually withdrawn from the left atrium.

Two other subroutines are available to him during the catheterization procedure. The first of these permits him to display the time-course of pressure during one heart cycle on both sides of a given heart valve. A second program permits the operator to display superimposed dye curves obtained following injection at three different sites.

By the time the catheterization procedure is complete the physician will usually have edited his report and be ready to print out hard copy. The first part of the report is a print-out of the edited data along with the time in minutes since onset of the procedure at which each data value was collected. The printer also plots the pressure waveforms from each heart chamber which was entered, followed by a plot of all the indicator dilution curves superimposed. At the end is a summary of the abnormal findings.

OTHER PROGRAMS

There are a variety of other programs available to the physician to use both during heart catheterization or for bedside monitoring or in the operating room. These include programs which measure stroke volume beat-by-beat from the contour of a central aortic pressure wave. This method is calibrated once against the indicator dilution method and has been shown to correlate .95 or better under a variety of physiologic states against beat-by-beat measurements of stroke volume made with an electromagnetic flowmeter.

A program is available for measuring the physical properties of aorta from a video tape recording of the fluoroscopic image of the chest following injection of opaque dye into the left ventricle or aorta. On an edge track of the video tape is recorded aortic pressure. The computer measures the time of transit for one line of the video signal from one edge of the aortic shadow to the other, averages 15 lines and, thus, obtains 60 times a second, aortic diameter and aortic pressure. A mathematical model is used to calculate the properties of the system including Young's Modulus of Elasticity.

Another program measures arterial properties by determining the transfer function which relates an upstream aortic to a downstream radial of femoral pressure wave. The transfer function takes the form of a second order linear differential equation and a segment of artery can be described in terms of a resonant frequency and damping coefficient. These

properties change as a function of age and have been shown to be predictable in artificial segments of the aorta inserted into dogs when these properties are measured independently.

And, finally, programs have been developed which permit estimation of the fraction of cardiac output perfusing various segments of the circulation from determinations of distribution of transit times through those segments. These distributions are the time domain representations of transfer functions which relate upstream and downstream indicator dilution curves. Although the feasibility of this approach, using theoretical data, has been clearly shown, the method is at present limited by our inability to obtain indicator concentration curves from sites such as the inferior vena cava, which are representative of the true cross-sectional concentration of dye as a function of time. Streamlining of flow in the venous system is a prime obstacle. However, we are currently testing a method developed by Earl Wood of the Mayo Clinic for measuring concentration of a radiopaque indicator from a video tape recording of the fluoroscopic image following injection of the dye by integrating the light intensity over a window positioned at time of replay of the video image. The computer does the integration and subtracts the density measured from another window used to compensate for changes in light intensity unrelated to dye concentration.

It is encouraging that the techniques developed have been so easily implemented in other laboratories.

Figure 1

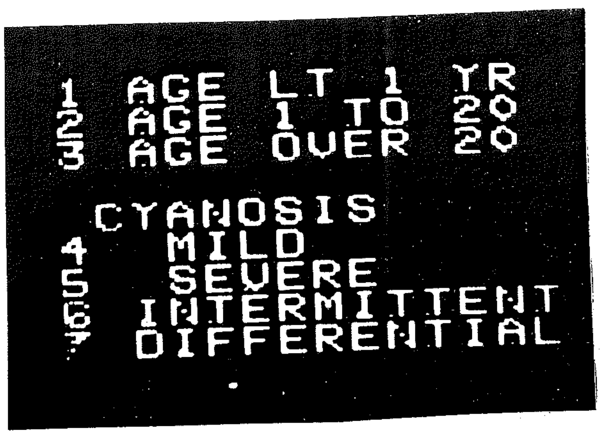


Figure 2

$$P_{(D_i/a_1,\, a_2, \cdots\cdots a_n)} = \frac{P_{D_i}\prod_{j=1}^{j=n}\left(P_{(S_j/D_i)}\right)^{a_j}\left(1-P_{(S_j/D_i)}\right)^{1-a_j}}{\sum_{i=1}^{i=K}(\text{Numerator})_i}$$

Where —

$a_i = 1$ if $i^{\underline{th}}$ symptom present

$a_i = 0$ if $i^{\underline{th}}$ symptom not present

Figure 5

$$P_{(D_i/a_1, a_2, \cdots\cdots a_n)} = \frac{P_{(D_i)}\prod_{j=n}^{j=n}\left(a_j P(S_j/D_i) + (1-a_j)(1-P(S_j/D_i))\right)}{\sum_{i=1}^{i=K}(\text{NUMERATOR})_i}$$

Where a_j is probability estimate for each symptom.

Figure 6

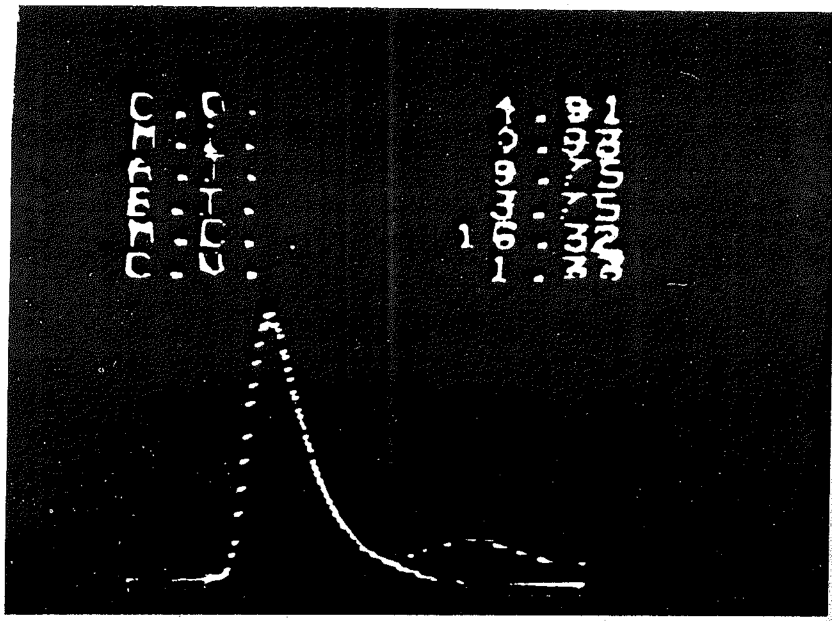


Figure 9

DIAGNOSTIC CLASSIFICATIONS OF NORMAL AND ABNORMAL ELECTROCARDIOGRAMS BY DIGITAL COMPUTER

By Hubert V. Pipberger, M.D., Georgetown University Hospital, Washington, D.C., USA.

Recent trends in electrocardiography are directed more and more toward improved quantiative methods and statements. The introduction of computer techniques has had already considerable impact on this development and promises to enhance ECG techniques in clinical use to a large degree. Since any input information prepared for digital computation needs to be prepared and processed in discrete numerical form, the need for more accurate data acquisition methods, greater precision in recording techniques and data analysis is felt more strongly than in the past.

The outcome of this development should lead to more reliable classification of ECG records than was possible in the past. The problems involved in this task are numerous and still need considerable study. Only those related more directly to record classification will be dealt with here.

Besides many instrumentation problems in data acquisition and analog-to-digital conversion, the selection of ECG leads, to be processed and analyzed, is of major importance. The most sophisticated classification techniques cannot improve input data derived from ECG leads, such as the conventional 12-lead which vary in performance from one subject to another. The introduction of corrected orthogonal lead systems into electrocardiography represented, therefore, a major step toward improved record classification. In some instances normal ECG ranges could be cut in half by application of such leads (e.g., the frontal plane QRS direction range). A more significant finding was that in most, although not in all cases, the separation between normal and abnormal ECG ranges improved consider-

NB References at page 381
Legends for tables at page 381
Tables and figures at pages 382 - 385

1) ably when more reliable leads were used (1).

The next problem in ECG record classification consists of the <u>selection of appropriate measurements</u>. The multiplicity of different parameters advocated in the past for diagnostic use is ample evidence for considerable uncertainty about optimal methods for recognition of ECG abnormalities. The many display methods developed in
2) the past all aimed at the same goal (2). Since almost all known ECG measurements, a total of more than 200, had been programmed in this laboratory for computer analysis it was possible to compare their diagnostic efficiency on the basis of large normal and abnormal re-
3) cord samples (3):

In a first comparison different types of ECG analysis were tested. A likelihood ratio test was used for this purpose. Several interesting findings evolved from this study. Series of instantaneous vectors derived from the X, Y and Z leads discriminated best diagnostic groups. They were obtained in one test at fixed time intervals of 0.01 sec. and in another from time-normalized QRS and ST-T complexes by dividing these complexes in 8 equal parts in time. Thus, the class-separation for QRS was based on 33 and 24 points respectively. For ST-T analysis 21 points were used. This type of procedure exceeded all others by a wide margin and the percentage of misclassification for almost all categories was less than 10 per cent for QRS.

A substantial deterioration in classifications by approximately 30 percent occurred when conventional ECG parameters such as Q,R and S wave measurements were used. This seemed to indicate that significant diagnostic information was lost when items from scalar leads replaced spatial parameters. The recording of ECG leads one by one without time coherence has to be considered, therefore, as a serious limitation in ECG analysis.

Evaluation of recognition rates was also performed for <u>time integrals, spatial maximal vectors and eigenvectors</u>. All of these measurements represent to a large degree summaries of ECG complexes, i.e., a whole complex is expressed in one term. The diagnostic performance of these items was very similar. It ranged 13-17 per cent

below the scalar lead measurements. When a differentiation between normal and abnormal records without further classification into specific diagnostic entities was tested, time integrals of entire cardiac cycles (<u>spatial ventricular gradients</u>) were found rather efficient. It is doubtful, however, whether such simple parameters can serve a useful purpose for specific diagnoses.

Multivariate analysis on the basis of series of instantaneous vectors has been used extensively in this laboratory for differential
3) diagnosis (3). The vector differences between an unknown patient vector and the means of various diagnostic groups are computed. A likelihood ratio test leads to a diagnostic classification and the vector differences are used as an expression of the "closeness" of the unknown vector and the various groups, which relate to the probability of the respective diagnosis.

A number of problems accompany this type of multivariate analysis. It is not known how large the record samples which form the basis of the classification have to be.

It has been stated that the sample size should be of the same order of magnitude as the number of parameters estimated, i.e., the square of the number of attributes. This statement has not been tested, however. As long as addition of new records leads to boundary fluctuations of the sample, its size is not yet adequate.

The record samples which form the basis of the classification have to be derived from patients whose diagnoses have been confirmed by autopsy or objective clinical evidence, leaving little room for doubt. Accumulation of large record samples of this type are very difficult and time-consuming. Even when records are collected from a number of hospitals as done in this laboratory, it takes many years to accumulate adequate numbers.

Use of autopsy data for correlation with electrocardiograms is more difficult than it might appear from the ECG literature. Patients who come to death with single myocardial infarcts are rather exceptions than the rule. The same is true for ventricular hypertrophies which are only rarely limited to one ventricle. The variety of possible

AUTOMATIC ANALYSIS OF THE ELECTROCARDIOGRAM AS A SERVICE TO THE COMMUNITY AND THE PRACTISING PHYSICIAN

By Cesar A. Caceres, M.D., Public Health Service, Washington, D.C., USA.

One of the most significant contributions of a medical computer system will be the aid it can provide to a community health service. Such a system can offer low cost accurate medical signal interpretation, which should improve the capability of a health department, clinic or hospital in mass screening and early disease detection. Easy recall and statistical analysis of the measurements is a further advantage.

To demonstrate feasibility of such systems, one group has worked with local community health departments. This has included taking 2000 electrocardiograms of patients in their homes a week as a mass screening effort, and is currently taking tracings in a health department clinic on a daily ongoing basis. In Hartford, Connecticut tracings are being recorded daily and transmitted to Washington for analysis.

Current work in this field includes evaluation of what leads and what other tests are of most value in health care.

Our group, the Instrumentation Field Station of the Heart Disease Control Program of the Public Health Service has had a role in this field by developing an automated system to process physiological signals used in clinical diagnostic tests. Digital computer programs have been written to recognize, analyze and classify the waveform patterns of the electrocardiogram, phonocardiogram, respirogram, and electroencephalogram. Other diagnostic signals are under current study.

The techniques are acceptable for clinical use so it is now essential to

NB References at pages 393 - 395
Legends for figures at page 396
Figures at pages 397 - 403

consider maximum utilization of these new developments not only for health screening but also by the practicing physician. By their use he saves time for patient care and increases the quality of his work.

THE AUTOMATED HEART STATION

Let's consider as an example an automated heart station. Machine-made electrocardiographic diagnoses comparable to those made by physicians have been routinely available at an experimental automated heart station now operational for over a year. Twenty-five thousand 12 lead electrocardiograms were processed last year using only one third of the system's capacity.

Electrocardiograms are recorded in the conventional manner. Each of the twelve standard electrocardiographic leads is the on-line input to the computer or the off-line input from frequency modulated analog magnetic tape. The electrocardiogram is recorded on tape on a special data acquisition device which contains a tape recorder, an electrocardiograph and a patient coding circuit. We could bypass magnetic tape and go directly from patient and electrocardiograph to computer but this is not currently economical because the computer would have to wait between patients, so magnetic tape is used as a storage medium.

(Codes are important if we are to have really automated systems. Codes are easy ways to "communicate" with the computer. The computer can respond to the code but not easily to a voice. Further, much information that might otherwise take a minute by voice can be given in a second or so.) The first items recorded on tape are a patient code number or other identifying data for the function being recorded. A calibration curve is the second group of important data recorded. It is thus possible to correct, within the computer, for any possible loss of signal linearity over the range of the recording or for technical error.

Transmission of the electrocardiograms from the data acquisition device may be by messenger delivery of the tapes or by telephone transmission of the signal in real time or tape recorded playback. At the processing system the analog signal is converted to digital samples at the rate of 500 per second and stored in the digital computer's core memory for identification, recognition and measurement.

INTRODUCTION TO THE PROBLEMS OF PHYSIOLOGICAL DATA ACQUISITION AND ADAPTION

By Antoine Rémond, M.D., Université de Paris, La Salpêtrière, Paris, France.

The most important questions when dealing with a computer is how to feed it with information. The information in the present case originates from the patients in a hospital; the interface between them and the computer is, as in other cases, most often called the acquisition of data. The problems encountered in collecting data in medicine are among the most complicated and delicate. For the benefit of the follwing discussion, my plan is to enumerate and classify some of the problems which I think are crucial, and then to elucidate those problems. I would like to take some examples in the field of electrophysiology and electroencephalography.

The relations between the physician and the patient may be of four categories: 1) they may be administrative, 2) they may be anamnestic, 3) they may be biochemical, 4) they may be biophysical. To this fourth category belongs a large part of the physical examination of the patients and particularly the various measurements and electrophysiological recordings. I would not like to minimize the importance of the three first categories, but I would like to limit myself to the fourth, the acquisition of biophysical data from the patients.

The temperature of the patient, the hue of his skin, the pressure of a cuff necessary to equilibrate the systolic pressure of the heart beat, the sounds made by the breath or by the beat of the heart - it is by the man-made evaluation and the systematic consideration of these phenomena that the symptoms of disease are recorded and that the diagnosis is precisely established by the physician. So the problem of to-day is more or less to imitate this fundamental behaviour by means of tools or instruments which extend, amplify or simplify the immediate human approach. The computer must be called for to take into consideration the much bigger masses of data which are obtained when electronic slaves replace the

doctor. At present, all the phenomena are not yet suitable for computer consideration. So one of the first necessities for the doctor is to select the phenomena, the physical entities, which are at the moment best suited for the introduction in the computer. The second problem after this selection is to use or to adapt the censors which apprehend these phenomena. They can of course be of many different kinds: some are mechanical, some are biomedical, some are electrical. Of course, most of them try to change the quality of the phenomenon into an electrical signal.

Once the selection of the best censors for the apprehension or the detection of the phenomenon is made, we have the problem of transduction, that is to say of a first adaptation of what has been acquired by the censor serving to put it in a form or to bring it to a level which will make it fit for measurement. The transduction has to be done in the most efficient way. This can of course be summarized in a unique sentence: one has to maintain or to increase the signal-to-noise ratio. To do this, one has to deal with different aspects. The first one may be to protect the signals or the source of signals from external interference. The second one is to try to avoid reactions from the source of signals (the patient) when the measurement is being done. One must not change the behaviour of source of signals when one tries to measure it. This is in fact a very important aspect which until to-day has probably not been dealt with correctly but many efforts are made for instance to keep the patient away from wires. Everybody knows that when a patient is in the middle of a forest of wires conducting various waves, the number of artefacts that appear is big; moreover, the patient becomes restless, and after a short while the quality of the data acquisition decreases. Telemetry has been used in the recent years to deal with this type of reaction problems and will probably be used more and more.

Now we are going to treat three different aspects: The spatial resolution of the data acquisition, the temporal resolution of same, and the precision with which they are obtained. Let us first speak of the spatial resolution. Here I would first like to say that one has always to think of where to place the sensor. This is by no means indifferent when one listens to the heart; if one is not immediately at the point of highest intensity of these sounds, one can of course have difficulty in finding the essential differences between those noises. This is of course quite

APPENDIX

MATHEMATICAL AND STATISTICAL PROCEDURES USED IN COMPUTER-AIDED DIAGNOSIS

1. BAYES' THEOREM AND CONDITIONAL PROBABILITY

A. Congenital Heart Disease

1. WARNER, H.R., TORONTO, A.F. and VEASY, L.G. A mathematical approach to medical diagnosis. Application to congenital heart disease. J.A.M.A. 177:177-183, 1961

2. WARNER, H.R., TORONTO, A.F. and VESAY, L.G. Experience with Bayes' Theorem for computer diagnosis of congenital heart disease. Ann. N.Y. Acad.Sci. 115:558-567, 1964

3. AMOSOV, N.M. and SHKABARA, Ye, A. Experience in formulating a diagnosis by means of diagnostic machine. (In Russ.) Eksp.Khir. Anest. 4:15-22, 1961. (Engl.transl.available JPRS: 11815, 3 Jan 1962)

4. BYKHOVSKIY, M.L., VISHNEVSKIY, A.A., and KHARNAS, S.Shr. Problems in the development of a diagnostic process by means of mathematical machines. (In Russ.) Eksp.Khir.Anest. 4:3-15, 1961 (Engl. transl.avaiable JPRS: 11814, 4 Jan 1962)

5. GUSTAFSON, J.E. The computer - for use in private practice. Proc. 5th IBM Medical Symp. 99-112, 1963

6. GUSTAFSON, J.E. and THROCKMORTON, T.D. The computer as a consultant. Med. Times. 93:115-120, 1965

7. TAKAHASHI, K. Logic of diagnosis and its processing by computer - with respect to congenital heart diseases and brain tumors. To be published in Proc.International Conference on Automated Data Processing in Hospitals, Elsinore, Denmark, April 1966

B. Acquired Heart Disease

1. BRUCE, R.A. and COX, A.R. Medical diagnosis by computer. Northwest Med. 60:1075-1082, 1961

2. BRUCE, R.A. Computer diagnosis of heart disease. Proc. 5th IBM Medical Symp. 77-98, 1963

C. Electrocardiography

1. KUMURA, E., MIBUKURA, Y. and MIURA, S. Statistical diagnosis of electrocardiogram by Theorem of Bayes. Jap.Heart J. 4:469-488, 1963

D. Cushing's Syndrome

1. NUGENT, C.A., WARNER, H.R., DUNN, J.T. and TYLER, F.H. Probability theory in diagnosis of Cushing's Syndrome. J.Clin.Endocr. 23:621-627, 1964

E. Thyroid Function

1. OVERALL, J.E. and WILLIAMS, C.M. Conditional probability program for diagnosis of thyroid function. J.A.M.A. 183:307-313, 1963

2. SILINK, K. A new mathematical approach to endocrinological diagnosis. In Progress in Biocybernetics. Vol.1. Edited by N. Weiner and J.P. Schade. New York:Elsevier, 1964. pp.20-35

F. Gastrointestinal Tract

1. RINALDO, J.A., SCHEINAK, P., and RUPE, C.E. Symptom diagnosis. A mathematical analysis of epigastric pain. Ann.Intern.Med. 59:145-154, 1963

2. WILSON, W.J., TEMPLETON, A.W., TURNER, A.H., LODWICK, G.S. The computer analysis and diagnosis of gastric ulcers. Radiology 85:1064-1073, 1965

G. Bone Tumors

1. LODWICK, G.S., HAUN, C.L., SMITH, W.E., KELLER, R.F. and ROBERTSON, E.R. Computer diagnosis of primary bone tumors. Radiology 80:273-275, 1963

2. LODWICK, G.S. Computer analysis of tumor roentgenograms. Proc. 5th IBM Med.Symp. 213-224, 1963

3. LODWICK, G.S., TURNER, A.H., LUSTED, LB., and TEMPLETON, A.W. Computer-aided analysis of radiographic images. J.Chron.Dis. 19(4): 485, April 1966

H. Neuropathology and Psychiatry

1. BIRNBAUM, A. and MAXWELL, A.E. Classification procedures based on Bayes' Formula. Applied Statis. 9:152-169, 1961

2. MISYUK, N.S., LEPESHINSKIY, N.A., LISKOVETS, A.O., and MASTYKIN A.S. Diagnosis of brain tumors with the help of the Universal "Ural-1" Computer. (In Russ.) Zhur.Nevr.i Psikh. (im Korsakava) 64:453-458, 1964 (Engl.transl.available JPRS: 24759, 26 May 1964

3. OVERALL, J.E. and GORHAM, J.E. A pattern probability model for the classification of psychiatric patients. Behav.Sci. 8:108-116, 1963

4. TAKAHASHI, K. Logic of diagnosis and its processing by computer - with respect to congenital heart disease and brain tumors. To be published in Proc.International Conference on Automated Data Processing in Hospitals, Elsinore, Denmark, April 1966

II. LIKELIHOOD RATIO

A. Hematology

1. LIPKIN, M. The likelihood concept and data processing in the diagnostic process. In The Diagnostic Process. Proc.of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacquez. Ann Arbor: Malloy Lithographing, Inc., 1964. pp.255-273

2. LIPKIN, M. The likelihood concept in differential diagnosis. Perspec. Biol.Med. 7:485-497, 1964

B. Screening and Diagnosis

1. COLLEN, M.F., RUBIN, L., NEYMAN, J., DANTZIG, G.B., BAER, R.M. and SIEGELAUB, A.B. Automated screening and diagnosis. Amer.J. Public Health 54:741-750, 1964

2. BAER, R.M. A note on computer diagnosis. IEEE Trans.Biomed. Engin.ME-11:8-12, 1964

III. DISCRIMINANT FUNCTION

A. Diabetes

1. BOGDANEK, T. The application of mathematical discrimination in the differential diagnosis of two types of diabetes. In Progress in Biocybernetics. Vol.1. Edited by N. Weiner and J. P. Schade. New York: Elsevier, 1964. pp. 45-54

B. Thyroid Function

1. OVERALL, J.E. and WILLIAMS, C.M. Models for medical diagnosis. Part One: Theoretical. Med.Dok. 5:51-56, 1961. Part two: Experimental. Med.Dok. 5:78 80, 1961

IV. FACTOR ANALYSIS

A. Thyroid Function

1. OVERALL, J.E. and WILLIAMS, C.M. Comparison of alternative computer models for thyroid diagnosis. Proc. San Diego Symp. Biomed.Eng. 1963. 3:141-145, 1963

B. Psychiatry

1. OVERALL, J.E. A configural analysis of psychiatric diagnostic stereotypes. Behav.Sci. 8:211-219, 1963

C. Use of Cornell Medical Index

1. VAN WOERKAM, A.J. and BRODMAN, K. Statistics for a diagnostic model. Biometrics 17:229-318, 1961

D. Urology

1. ZINSSER, H.H. Patient derived autodefinition of a disease: Pyelonephritis. In The Diagnostic Process. Proc.of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacquez. Ann Arbor: Malloy Lithographing, Inc., 1964. pp. 103-128

V. MATCHING PROCEDURES

A. Dermatology

1. NORINS, A.L. Computers in dermatology. Arch.Dermatol. 90:506-511, 1964

B. Hematology

1. LIPKIN, M., ENGLE, R., DAVIS, B.J., ZWORYKIN, V.K., EBALD, R. and SENDROW, M. Digital computer as an aid to differential diagnosis. Arch.Int.Med. 108:56-72, 1961

C. Neurology

1. SHTABTSOV, V.I. and OSTROUMOV, Yu.A. Experience with the diagnostics of nervous diseases using an electronic digital "Ural-2" (ETsVM) Computer. (In Russ.) Zhur.Nevr.i Psikh. 64: 515-519, 1964

D. Personality Assessment

1. ROME, H.P. Automation techniques in personality assessment. J.A.M.A. 182:1069-1072, 1962

2. PEARSON, J.S., SWENSON, W.M., ROME, H.P., MATAYA, P., BRANNICK T.L. Further experience with the automated Minnesota Multiphasic Personality Inventory. Proc.Mayo Clin. 39:823-829,1964

E. Surgery

1. PETERSON, O.L. and BARSAMIAN, E. Diagnostic performance. In The Diagnostic Process. Proc. of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacquez Ann Arbor: Malloy Lithographing, Inc., 1964. pp.347-362.

REFERENCES

1. The Diagnostic Process. Proc. of a Conference Held at the University of Michigan May 9-11, 1963. Edited by John A. Jacquez, 391 pp. Ann Arbor: Malloy Lithographing, Inc., 1964. Library of Congress Catalog Card: 64-21210

2. ENGLE, R.L. and DAVIS, B.J., "Medical diagnosis: Present, past and future. I. Present concepts of the meaning and limitations of medical diagnosis," Arch. Int. Med. 112 (1963), pp. 512-519

3. -"- , "Medical diagnosis: Present, past and future. II. Philosophical foundations and historical development of our concepts of health disease and diagnosis," Arch. Int. Med. 112 (1963), pp. 520-529

III."Diagnosis in the future, including a critique on the use of electronic computers as diagnostic aids to the physician," Arch. Int. Med. 112 (1963), pp. 530-543

4. SPROSTY, P.J. "The use of questions in the diagnostic problem solving process," In The Diagnostic Process. Proc. of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacques. Ann Arbor: Malloy Lithographing Inc., 1964. pp.281-308

5. RIMOLDI, H.J.A. "Testing and analysis of diagnostic skills," In The Diagnostic Process. Proc. of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacquez. Ann Arbor: Malloy Lithographing, Inc., 1964. pp. 315-343

6. ADAMS, W.R. "Studies of teaching of the diagnostic process," in The Diagnostic Process. Proc. of a Conference Held at the University of Michigan, May 9-11, 1963. Edited by John A. Jacquez. Ann Arbor: Malloy Lithographing, Inc., 1964. pp. 367-385

7. KLEINMUNTZ, B. "Profile analysis revisited: A heuristic approach," J. Counsel. Psychol. 10 (1963). pp. 315-321

8. REITMAN, W.R. "Information processing models in psychology," Science 144 (1964), pp. 1192-1198

9. NEWELL, A, and SIMON, H.A. "Computer simulation of human thinking," Science 134 (1961), pp. 2011-2017

10. SIMON, H.A. Administrative behavior. (2nd ed.) Mac Millen, New York, p. xxix. 1957

11. TAYLOR, D.W. "Cognitive processes in solving algebra word problems." In Problem Solving. B. Kleinmuntz, Ed., Wiley & Sons, New York, 1966, p. 120

12. KLEINMUNTZ, B. Computer Simulation of Diagnostic Judgments. Presented at the 36th annual meeting of Midwestern Psychological Assoc., St. Louis, Missouri, 1964. Personal communication.

13. SWETS, J.A. and FEURZEIG, W. "Computer-aided instruction," Science 150 (1965), pp. 572-576

14. LUSTED, L.B. "Logic of the diagnostic process", Methods of Information in Medicine, Vol. IV(2), (Jun 1965), pp. 63-68

15. SCHEFF, T.J. "Decision rules, types of error and their consequences in medical diagnosis," Behav.Sci. 8(1963), pp. 97-107

16. NISSEN-MEYER, S. "Evaluation of screening tests in medical diagnosis," Biometrics 20(1964), pp. 730-755

17. COLLEN, M.F., RUBIN, L., NEYMAN, J., DANTZIG, G.B., BAER, R.M. and SIEGELAUB, A.B. "Automated screening and diagnosis," Amer.J.Publ. Health 54(1964), pp. 740-750

18. FRIEDEN, J., SHAPIRO, J.H. and FEINSTEIN, A.R. "Radiologic evaluation of heart disease," Arch.Int.Med. 111(1963), pp.44-50

19. HORVATH, W.J. "The effect of physician bias in medical diagnosis," Behav.Sci. 9(1964), pp. 334-340

20. LUSTED, L.B. "Logical analysis in roentgen diagnosis," Radiology 74 (1960), pp. 178-193

21. BIRNBAUM, A. and MAXWELL, A.E. "Classification procedures based on Bayes' Formula," Appl.Statistics 9(1961), pp. 152-169

22. TORONTO, A.F. VEASY, L.G. and WARNER, H.R. "Evaluation of a computer program for diagnosis of congenital heart disease," Progr.Cardiov. Dis 5(1963), pp. 362-377

23. GROSZ, H.J. and GROSSMAN, K.G. "The sources of observer variation and bias in clinical judgments: 1. The items of psychiatric history," J.Nerv.Ment.Dis. 138-2 (1964), pp. 105-113

24. KING, D.J. and MANEGOLD, R.F. "Consistency of cardiologist' prognostic judgments in cases of myocardial infarction," Amer.J. Cardiol. 15 (1965), pp. 27-32

25. PICKERING, Sir GEORGE W. In Conclusion: The Physician. Controlled Clinical Trials - A Symposium. Oxford: Blackwell, 1960, pp,163-167

26. LEDLEY, R.S. and LUSTED, L.B. "Reasoning foundations of medical diagnosis," Science 130(1959), pp. 9-21

27. LEDLEY, R.S. and LUSTED, L.B. "Medical diagnosis and modern decision making," Proc.ofSymposia in Applied Math. 14 (1962), pp. 117-158

28. CORNFIELD, J. "Bayes' theorem," Proc. 6th IBM Med.Symp. (1964), pp. 163-196

29. CORNFIELD, J. "Discriminant functions," Proc. 6th IBM Med.Symp. (1964), pp. 567-589

30. EDWARDS, W., LINDMAN, H. and SAVAGE, L.J. "Bayesian statistical inference for psychological research," Psych.Rev. 70(1963), pp. 193-242

31. MOSTELLER, R. and WALLACE, D.L. "Inference in an authorship problem," J.Am.StatisticalAss. 58 (1963), pp. 275-309

32. MOSTELLER, F. and WALLACE, D.L. Inference and disputed authorship: The federalists. Addison-Wesley Pub. Co., Reading, Mass., 1964

33. WARNER, H.R., TORONTO, A.F. and VEASY, L.G. "Experience with Bayes' theorem for computer diagnosis of congenital heart disease," Ann. N.Y.Acad. Sci. 115 (1964), pp. 558-567

34. BRUCE, R.A. "Computer diagnosis of heart disease," Proc.5th IBM Med. Symp. (1963), pp. 77-98

35. LODWICK, G.S., HAUN, C.L., SMITH, W.E., KELLER, R.F. and ROBERTSON, E.R. "Computer diagnosis of primary bone tumors," Radiology 80 (1963), pp. 273-275

36. OVERALL, J.E. and WILLIAMS, C.M. "Conditional probability program for diagnosis of thyroid function," J.A.M.A. 183 (1963), pp.307-313

Note: A booklet entitled "Computer diagnosis in thyroid disease," has been prepared by Dr. Williams. This excellent booklet contains an interesting and detailed description of items used in the study, i.e. case histories, data collection sheets, data card layout, data matrices, computer programs, program flow charts. The physician should be able to start his own project with the use of this booklet. A copy may be obtained by writing to Dr. Clyde M. Williams, Department of Radiology, University of Florida College of Medicine, Gainesville, Florida.

37. NUGENT, C.A., WARNER, H.R., DUNN, J.T. and TYLER, F.H. "Probability theory in the diagnosis of Cushing's syndrome." J.Clin.Endocr. 24 (1964), pp. 621-627

38. WILSON, W.J., TEMPLETON, A.W., TURNER, A.H., and LODWICK, G.S. "The computer analysis and diagnosis of gastric ulcers," Radiology 85 (1965), pp. 1064-1073

39. NEYMAN, J. First Course in Probability and Statistics. New York: Henry Holt, 1950, Chapter 5

40. OVERALL, J.E. and WILLIAMS, C.M. "Comparison of alternative computer models for thyroid diagnosis," Proc. Dan Diego Symp. Biomed. Eng. (1963), 3 (1963), pp. 141-145

41. OVERALL, J.E. and WILLIAMS, C.M. "Models for medical diagnosis: Pt.1: Theoretical," Med.Dok. 5 (1961), pp. 78-80.

(From Kleinmuntz, B. Personal communication.)

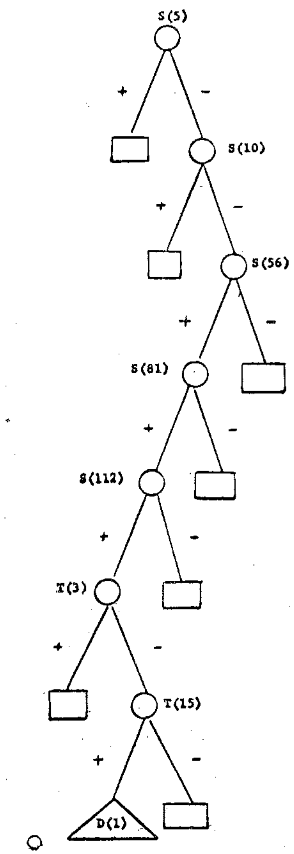


Figure 1. The process of diagnostic decision-making may be visualized as a tree structure which represents the search strategy of a diagnostician. The circles represent nodes at which the physician may give a differential diagnosis or he may elect to proceed by asking for more information. The plus or minus branches represent the presence or absence of the preceding symptom, e.g. S(56).

(From Lusted, L.B., Radiology 74 (1960), pp. 178-193.)

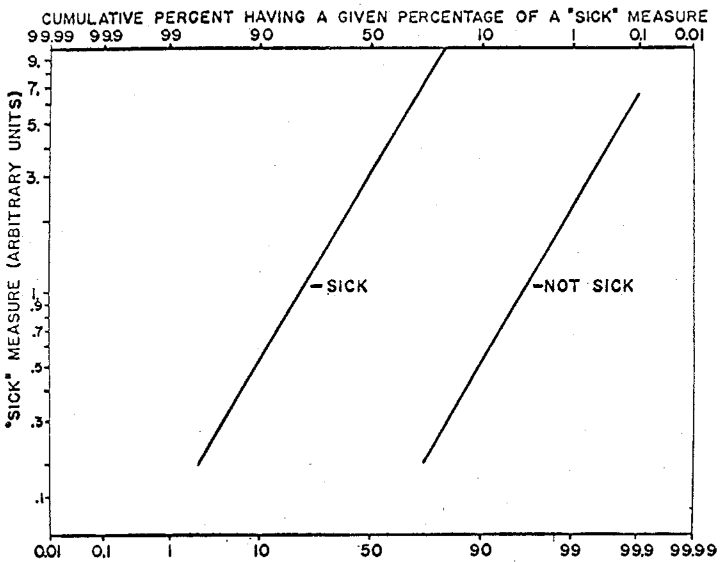


Figure 2. A population has been divided into two groups, "ill" and "not-ill" on the basis of arbitrary criteria for "ill". The plot is made in logarithmic probability paper.

(From Lusted, L.B., Radiology 74 (1960), pp. 178-193.)

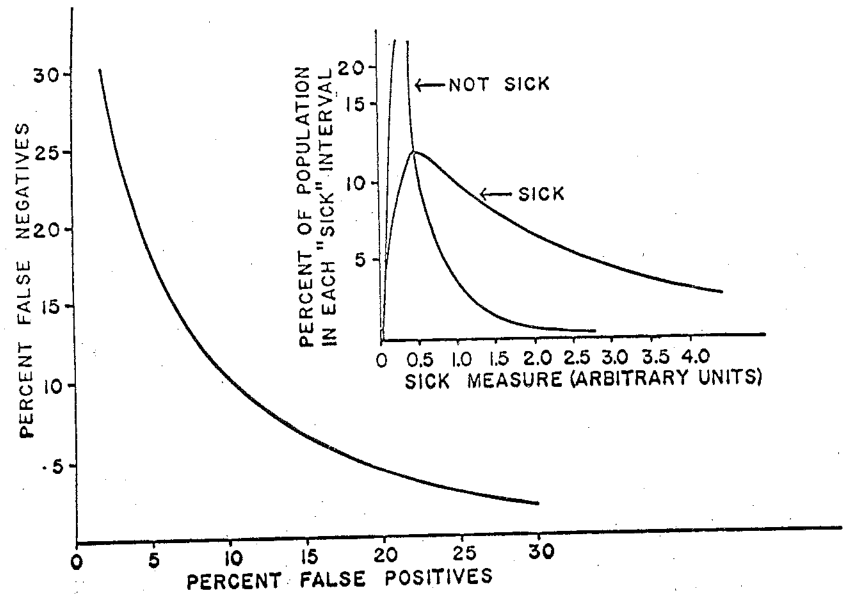


Figure 3. Upper right graph: The population distribution from Figure 2 is replotted to show that the curves are longnormal. Lower left graph: The reciprocal false positive-false negative relationship for the population is very near a hyperbola and is a good fit for the experimental data curve in Figure 4.

(From Lustsd, L.B., Radiology 74 (1960), pp. 178-193.)

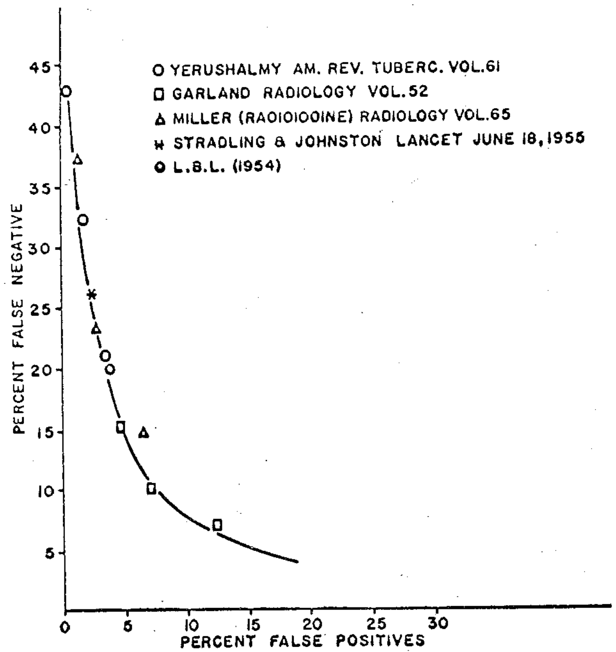


Figure 4. Graph showing the reciprocal relationship between the percentage of false positive and of false negative readings obtained from studies on the accuracy of chest film interpretation. Miller's study on the interpretation of the radiation effect on thyroid gland and other studies on the accuracy of diagnosis of Papanicolaou smears have shown a similar reciprocal relationship.

CYBERNETICS IN MEDICINE

By Professor Aldo Masturzo, The University of Naples, Naples, Italy, President of the International Society of Cybernetic Medicine.

This paper will discuss the following topics:

1) general principles of Cybernetic Medicine
2) cybernetics applied to space biology
3) cybernetics applied to rheumatology
4) diagnosis with the aid of computers.

Biology, like other sciences, does not tell about reality in itself, but provides us with models representative of it, which mimic natural phenomena and are liable to previsions, deductive reasoning and experimental methods.

Mathematical models are by far the most effective for a systematic study of reality: which is witnessed by the achievements of physics in these last years.

But whils in physics traditional mathematical methods proved to be satisfactory to a logical arrangement of observed data, in biology such methods were inadequate and not sufficient.

In scientific research, mathematics avails itself of systems of classic mechanics, differential equations, which were successfully employed in astronomy and physics.

Such systems, however, in biology do not give the same results because vital phenomena are subject to aleatory factors, which make it quite impossible to determine the history of a living organism at every moment of its evolution.

NB Figures at pages 542 - 547.

THE INFORMATION RETRIEVAL PROBLEM IN MEDICINE

By G. Wagner, Professor, M.D., Deutches Krebsforschungszentrum, Heidelberg. Germany.

Within the scope of our meeting, I have the pleasure to speak about the problem of information and information retrieval in medicine. Although the word "Information" applies to different kinds of systems (references, facts and data), the term "Information Retrieval" customarily is restricted to the retrieval of printed information. But, as I have been asked to use the term in its broadest sense, I will also stress some problems of retrieval of facts and data. However, as these fields together are so broad and complex and furthermore only partly explored, my paper must necessarily remain very fragmentary.

Let me start with some general remarks upon the Information problem in science.

There is no doubt that the rapid development of all sections of science may be called one of the most outstanding characteristics of our age. The extent of this expansion may be illustrated by the statement of the American science historian de Solla Price (1963), that out of all scientists that have ever crowded our planet, about 85% are living in our time. But the rapid devebpment of science, exercising an essential influence on our present form of society, our standard of life, and our world-view, also raised problems that have never existed before or, at least, did not become evident in acomparably aggravating fashion, e.g., the increasing difficulty of keeping oneself informed on scientific progress.

Until some few decades ago, it was not too difficult for a scientist to keep himself up to date in his speciality and also to get sufficient

NB References at pages 574 - 575
Figures at pages 576 - 579

information about the progress made in the adjoining disciplines. Nowadays - at least in the sector of natural science and technology - there is hardly any scientist able to overlook the entire literature of his special working field, still less, to obtain a general survey of the adjoining scientific branches. Principally two factors complicate the scientist's survey of the literature that is relevant or at least potentially important to him: on one hand, the increasing number of publications, on the other hand, the growing difficulties of communication between the different scientific specialities.

During the last years, the growth of literary production in the scientific sector has often been labeled as a "publication explosion" or "literature flood". But considering the situation from an impartial viewpoint, the exception must be made that these slogans have pretty often been abused by sensation mongering, and that things have been magnified deliberately (Coblans, Wagner (1955)). The purpose of those exaggerations is - as Shaw (1962) mocked - " to scare us into accepting radical solutions to the documentation problem, for the alternative pictured is that of science dead of autointoxication, from immersion in its own effluvia". However, though a certain reservation as to bombastic overstatements seems to be imperative until better founded data will be presented, one thing is a matter of fact: never before mankind has faced such an inundation by printed paper as it does at present. According to American estimations (Kent (1960)), every sixty seconds - Sundays and holidays included - about 2000 pages of newpapers, periodicals or books are being produced. The total of scientific periodicals actually published on earth are valued at no less than 30 000; the estimated number of the articles published in then annually fluctuates from 1.5 to 6 millions.

In the field of chemistry, the number of scientific publications is said to have nearly doubled about every eight years during the last decades (Pietsch (1962)). Wheras in 1961 some 150 000 papers appeared in this scientific sector, an amount of about 300 000 publications are expected to come out in 1970 (Kent (1962b)).

The determination of the growth rate in biomedical literature is much more difficult. In all probability, the expansion of medical literature is not so strong as the increase of publications in the chemical field.

COMPUTERS IN MEDICAL TEACHING

By J.C. Pagès and F. Grémy, Professor at the School of Medicine, University of Paris, France.

All countries, even developed, have to face hard problems in the educational field. Among these problems, there are the following:

1) the tremendous growth of new information and the difficulty of every one to keep up with development in his own field
2) the growth of student population
3) the difficulty of having sufficient and competent teachers, this resulting in the fact that education is becoming more and more impersonal in large universities.

These educational problem can be partly overcome by the help of computer technology, in the following field, information storage and retrieval, computerized problem solving, and machine programmed learning, which is our topic to day.

PROGRAMMED INSTRUCTION

It is impossible to speak of computers as teaching machines without keeping in mind the exact notion of programmed instruction.

Programmed instruction is a technique of instructing

- without the presence of a human instructor
- where the material is presented to the student in small increments
- which requires from the trainee frequent responses and informs him immediately whether his response is right or wrong. As the questions are very easy, the answer is supposed to be very often

NB References at page 559
Figures at pages 560 - 565

a right one. This results in a very strong positive feed back between the programmed instruction and the student, which is called reinforcement.

It can be a linear programmed instruction (B.F. Skinner): after the student has compared the computer answer with his own, he goes on to the next step, which is always the same, whatever his answer has been.

More sophisticated is the branching programmed instruction : here the next step depends on the previous answer, which determines which information will be presented to him next : new material, if the answer was correct, explanatory or review material if it was not. One can conceive that the next step depends not only of the just preceding answer, but of all the previous answers of the student, and even takes into account his capabilities as revealed by some psychologic texts such as Q. I. or M. M. P. I[1]. So the teaching strategy could be very closely adapted to the personality of the learner.

What are the advantages of programmed instruction over more classical ways of teaching?

- as there is no group instruction, there are no absences and interruptions to disturb the pupils. A student who works alone can start and stop whenever he wants
- the work required from the student is an active and personal one : lazy students cannot wait for the answers of other students : shy students are not inhibited by the presence of others
- the rates of learning are different : the faster is not held back; the slower is not left behind; the levels of achievements are more uniform.

Very often it is reproached to this kind of teaching to be impersonal because it separates the learner from the instructor. But it is possible to point out that programmed instruction is more personal in two ways : first, it matches the student capability : second it frees time for valuable contacts between student and teacher : they get rid of factual notion acquisition and they are able to spend time for interesting and

(1) Minnesota Multiphasic Personality Inventory.

advanced work, which no machine can do, for the time being, where imagination, synthesis and moral capabilities are at work.

Now what do we call a Teaching Machine?

In the wide sense, it is any device which contains a programmed instruction and presents it to the students (slides, filmstrip, magnetic tape, or the combination of these).

In a narrower sense, a teaching machine is a general purpose computer with all the relevant peripheral devices. Among all teaching devices, data processing machines have several advantages :

- it is easier to store in a computer a branching process than in any other device, and no other possibility exists for the very complex teaching strategies we mentioned previously.
- the computer can monitor very closely how quickly the student learns, how many mistakes he makes and which ones, and so on. So it gives a possibility of evaluating either the student or the program itself or both of them.

The major drawback of computers as teaching machines is the high price, but this disadvantage can be partly overcome with time sharing machines using multiprogramming. Up to thirty students now, and more than one hundred in the future, can converse simultaneously with the machine as a private interlocutor and unaware of the presence of others. Another difficulty is the reluctancy of many teachers, who fear to be supplanted in their work. This way of thinking will be perhaps common in Europe, and specially in latin countries where a kind of humanistic and literary tradition is very strong, even among the scientists. "I do not believe in your gadget" told one of our eldest and most cultured colleagues in Paris.

Their fear is vain, because preparation of programmed instruction is a very difficult time consuming activity, which requires from the teachers experience and high pedagogic qualities. This difficulty may be seen as another drawback of this method, because of the lack of trained people, in psychology as well as in programming.

THE PLATO[1] PROJECT (Illinois University)

We have now some general idea of programmed instruction and teaching machines. It is time to speak of some effective programs. We think it is interesting first to give a quick description of the PLATO project [2], even if its point of interest is not medical teaching, but computer language teaching. We do so because it is not only a project but it has already actually worked since the year 1960. Since its very beginning, it was one of the best and more complete achievements in teaching by machines.

The device (see fig. 1)

First they used an old and slow machine : ILLIAC, the University of Illinois general purpose digital computer with a rather small memory and then has been much improved using now a CDC 1604 with 32K memory. Only the first and simple version will be described here as being sufficient for our purpose, that is just to give an example.

Each of the students could communicate with the computer :

a) through a keyset with the usual alphanumeric characters and additional keys for special symbols

b) through a television set which is used both as an electronic book and as an electronic blackboard.

The electronic book consists of a set of slides containing the material to be taught. And a slide selector, controlled by the computer selects the slide to be displayed on the screen.

The computer can also write graphs, figures, letters on the screen like a teacher on a blackboard.

(1) Programmed logic for Automatic Teaching Operations.

(2) in the same university, another computer assisted instruction project is developed : the SOCRATES (of course) PROJECT (L.M. STOLUROW). It is essentially devoted to fundamental research in psychology, using an IBM 1710.

containing the material of the lesson, the answers to the problems and also some home-work. This record could provide a very good knowledge of each of the students : how fast he works, how long he ponders before he makes a decision, how often he calls for help ... and finally what kind of intelligence he has. It provides a evaluation on the student.

4) at the end of the lesson, when all the students have made their work, the computer is requested to print synthetic records of the whole lesson (see Fig. 4). This is a great advantage above all other teaching devices.

For the teacher has to write just one part of the dialogue. The other part, the students' one, must be guessed, and the teacher must provide an appropriate reply to each answer he can imagine. For instance, it is very difficult to write the help sequence guessing at the actual difficulties the student will meet. So the great interest of PLATO project is to be able to collect data about these difficulties, so the help sequences could be modified. The compter appears to be a very powerful teaching tool, much more powerful than "programmed teaching books" which keep no record of the answers and have no such flexibility and potential improvement facilities. The teacher has now a great feedback to judge the effectiveness of his lessons.

Teaching machines in the medical field

Now, what has been made in medical area with teaching machines? Several programs have been written. We shall describe three of them, which have several common features.

They are not exactly programmed instruction because they do not deliver new knowledge. They are rather training games which forces the students to use the knowledge they have already acquired. And with this knowledge the machine trains them to make medical diagnosis : the machine plays the part of both the patient and the medical instructor. As well as training exercises, these programs could be used as programmed examinations.

The second common feature : as far as we know, none of them have been actually used, because the medical material to be included into the program has not been written yet, except for some demonstration examples.

First example:

The first program to be presented here has been conceived and written by one of us. This program(+) enables a single computer to simulate many patients at the same time, permitting diagnosis by a number of students - the students asking for questions, then proposing their diagnosis, the computer answering the questions and discussing the diagnosis.

A session devoted to this training would gather fifteen or twenty students under the responsibility of an instructor (a man up to now...) Ten to twenty diseases are chosen to be simulated by the computer. Each student has to distinguish between all these diagnoses by an adequate inquiry. To make possible this dialogue, he is provided with two lists.

1) A list of the examinations he can ask for. This list contains clinical examinations as well as lab-tests. For instance, among the clinical questions, one can find : "auscultation of the heart", and among the laboratory ones "white blood count and differential" or "chest X-ray", etc...

2) A list of diseases among which is the diagnosis to be found.

For each case, with the help of these two lists, the student first "examines his patient", then offers his diagnosis.

The first list enables each student to ask questions to the machine in a way very similar to medical practice in a hospital : the visiting physician asking the resident about clinical or laboratory information related to a recently admitted patient.

The questions and diagnostic propositions are processed inside the machine in ciphered codes. This coding can be done by the student or by the machine operator, or by a mark-sensing system. It could even be possible to use the actual hospital forms. But the best solution, disregarding cost, would be two-ways terminals, one for each student : the question being typed either in code or in plain language by the student himself. Within a few seconds, the answer reaches the student, always in plain language, either printed or displayed on a cathode-ray terminal.

(+) This program is written for an IBM 1401 with disks in autocoder, with the help of A. Stephant.

This answer includes the name of the student, the number of the case, the text of the question and the answer itself : this answer is a list of the symptoms which would be usually collected through this examination on a typical patient suffering from this disease.

Diagnostio propositions

When the student thinks he is informed enough and believes he can set a diagnosis - and justify it -, he proposes this diagnosis to the computer.

Within a few seconds the machine prints "right" or "wrong", but cal also discuss the wrong propositions. Thus, the computer keeps in its memory each student's questions, all the answers and all the consequences of these answers : that is to say that the computer updates by itself the differential diagnosis with the help of the clinical files. Therefore, it can print one of the four following types of answers:

a) RIGHT DIAGNOSIS - STUDY THE FOLLOWING PATIENT.

b) RIGHT DIAGNOSIS - BUT THE SYMPTOMS YOU HAVE COLLECTED UP TO NOW ARE ALSO COMPATIBLE WITH THE FOLLOWING DISEASES : (here, the computer lists the diseases which should have been eliminated by a proper inquiry and which have not been).
PROCEED WITH THE DIFFERENTIAL DIAGNOSIS.

c) WRONG DIAGNOSIS, THOUGH IN ACCORDANCE WITH THE SYMPTOMS YOU HAVE COLLECTED.

d) WRONG DIAGNOSIS.
   1) THIS DIAGNOSIS DOES NOT EXPLAIN THE FOLLOWING SYMPTOMS YOU HAVE ALREADY COLLECTED :....
   2) YOU ARE IN CONTRADICTION WITH THE LACK OF THE FOLLOWING SYMPTOMS WHICH WOULD HAVE BEEN FOUND OUT BY YOUR QUESTIONS IF YOUR DIAGNOSIS HAD BEEN RIGHT :....

Normally the student must proceed with his inquiries and propositions until he obtains the first answer. But other "game-rules", programmed or not, can be decided by the instruotor.

What advantages can be expected from this method?

The advantages of such a simulation-teaching-method have to be

established by a long range pedagogic experimentation. But one may expect the following ones :

- A good motivation based on a real-life pattern.
- Each student leaves a session with a printed list of all the questions, answers, and comments. He can, therefore, think the case over again checking with his books. By this method, a student can work rather like a physician improving himself than like a pupil.
- It is easy to record on a digital tape all the sessions. It could be thus possible to analyze by a special program each student's progress and to warn him of some idiosyncraic attitudes (too much self-confidence, for instance or, on the contrary, too much "anxiety".)
- By letting well-trained physicians play this game and discuss it, one can expect a progressive improvement of files.

Second Example

The second program we shall describe is a program written by A.D. Kirsch of Datatrol Corporation.

The point of interest about this program is that it combines game and learning theories.

The program requests the machine to generate synthetic patients from probability tables. There are four of these : symptom-disease probability table ; physical finding-disease p.t. ; medical history-disease p.t. and labtest-disease p.t. It must be pointed out that only marginal probabilities and not conditional probabilities are taken into account. A fifth table is devoted to test whether the lab. exams are required in the optimal sequence. None of these tables have been written yet, because it is a cumbersome work.

The instruotor is supposed to put in a card containing the age, the race, the sex and the disease, that he wants the student to work on.

Then using a Monte Carlo method, the machine works out symptoms and physical findings. A random number generator will generate a number RN to be compared with the probability P, of the first symptom ; if $RN < P_1$ the symptom is accepted and delivered to the player. If not, it is discorded and the computer goes on immediately to the second symptom, and so on... Then the machine does the same thing with the physical findings.

Now the player is requested to make a diagnosis. If he is not ready, he may decline the invitation and ask for medical history. Through the same process as before, the machine looks up through the third table and generates a medical history.

The player is then again asked if he is ready to make a proposal.

If not, he can ask for a lab-exam, one at a time, and the computer will use the fourth table and look up the probability of the test having a positive result. If the probability is too low (for instance.I), the choice of this particular lab-exam is considered as not reasonable, and the game is over.

If the choice is reasonable, the computer uses the fifth table, in order to judge if the sequence is correct or not, that is, if that test is a good one to ask for at the present time of the diagnosis process. If it is a good one, the machine comes back to the fourth table and generates a random number to give the result - positive or negative - of the test selected.

It is of interest to compare this program with the previous one.

In both programs, the disease is selected by the instructor. But it is not the same with the patient : in the first program, the patient is entirely determined, because it is assumed that he is a typical one, with <u>all</u> the symptoms and physical findings, and <u>without any</u> other symptoms. By the Datatrol program, on the contrary, there is much uncertainty for the student but also for the instructor : the patient synthetized by the machine may be a typical one or an unusual one, like in the true medical game when patients are not always like hand-book descriptions. The consequence of this if the following : all the students in the first program play the same game, because they have the same patient. In Kirsch's program the patients are different even if they have the same disease : the consequence of this is a possibility of fruitful discussions between

students and instructor after the lesson.

Another comparaison can be made about the conversational features of these programs :

Kirsch's player has very few decisions to take : he has not to choose for chest exam or for abdominal exam ; he has no question to ask for medical history. In the other program, the student asks for each exam separately. The judgment given by the program is rather poor in the Datatrol program : the player has either won or lost. It is more elaborated in the other program, where the differential diagnosis is thoroughly considered. The diagnosis strategy is neglected by both of the programs, except for the lab tests in the second one.

The last but not the least common feature of those programs, is the following : both of them are general programs : one has just to change the tables, and it is possible to use these programs for every topic of medical field, or even for other fields, such as the diagnosis of the breakdown of a motor car, an oscilloscope or a computer.

<u>Third Example</u>

This third example is quite different from the former ones because each disease, each medical program requires to use a <u>different program</u>.

W. Feuerzeig developed a teaching system which made possible a true <u>student-oomputer dialogue</u>, and which he called a socratic-system, because it simulates rather well the Socrates' maieutio way of teaching.

The program is so designed that the student is very <u>free</u> in his strategy for solving the problem : he is able to ask for every information he wants and to propose his diagnosis whenever he likes. The only restriction is the vocabulary : he is given a list whioh specifies the words he is allowed to use.

The responses of the computer are very elaborated : it answers by long sentenoes, in full language. More interesting is the fact that each answer depends not only on what the student has just said, but also on every answer that preceded it.

Here is now a medical example : player was given a medical history of a patient who arrived at the hospital with a high fever, and the diagnosis to be made was a pneumococcal pneumonia. A list of the allowed questions and declarations was provided to the student. One sees again that, even if correct, a diagnosis is refused if the logical process of diagnosis is not completed. Only a little part of the dialogue is presented on the figure 5.

One may now emphasize some very important aspects of this socratic teaching system, more elaborated than any other, as they have been outlined by W. Feuerzeig himself:

a) the student computer interactions are very flexible : the student has to answer the questions, but he can make any assertion he wants. The program must be able to maintain a continuity of context, taking into consideration all previous statements of the student. The analyst who writes the program must foresee all the ways the student can take to the diagnosis. So the program is always specific of the case under study, and it must be written by a member of the medical staff.

b) the language used for these interactions must be a natural language. It is easy for output : the statements of the computer as stored in the memory. But the machine must also understand the input natural language : this problem has been solved by the use of a restrained list of sentences allowed. The larger is this dictionary, the more sophisticated is the logic structure of the program.

These two remarks pose the problem of the programming languages used in programmed instruction. When the programs are general, it is conceivable to write them in specific machine languages as it was made in the first example, where 1401 autocoder was used. But one must remember that the present hardware will be obsolete when the experimentation in programmed instruction reaches interesting results.

When specific programs are involved, and when medical people are asked to provide not only the input data (files, matrices and so on), but also the logic structure of the program, it becomes impossible to require from them to write the program even in a symbolic language such as FORTRAN, where the input/output statements are tedious : and it would be

cumbersome to ask for the mediation of professional programmers. So the need for a specific language adapted to teaching applications, facilitating input and output of alphanumeric character, appears to be very great. Such languages are now available. For instance the COURSE WRITER, developed by IBM, has a basic repertory of twelve operation codes, most of which represent a series of instructions to the computer.

Here are some final remarks as a conclusion:

- these teaching methods intend to help the student in his hospital practice, not to replace it of course.
- one must remember that the difficulty to set a diagnosis lies less in the mental deduction from the symptoms than on the search and estimation of these symptoms and physical findings on the patient.
- the programmed instruction by computer, which has been presented here as a mean of solving some current educational problem, may appear in the future constitute improved ways of teaching, complementary to the traditional ones, and be kept, even if the cause of their utilization has disappeared.

REFERENCES

1. "Conversational Teaching Machine" by Wallace FEURZEIG, in Datamation June 1964, p. 39

2. "The Use of an Automatic Computer System in Teaching" by P.G. BRAUNFELD and L.D. FOSDICK in report R. 160, September 1962 from Coordinated Service Laboratory, University of Illinois, Urbana, Ill.

3. "Utilisation d'un ordinateur en pédagogie médicale" by J.C. PAGES in "La Presse Médicale". 12 Décembre 1964.

4. " A computerized Medical Training Game as a Teaching Aid" by Arthur D. KIRSCH in Data Acquisition and Processing in Biology and Medicine, Vol. II, Pergamon Press (page 139).

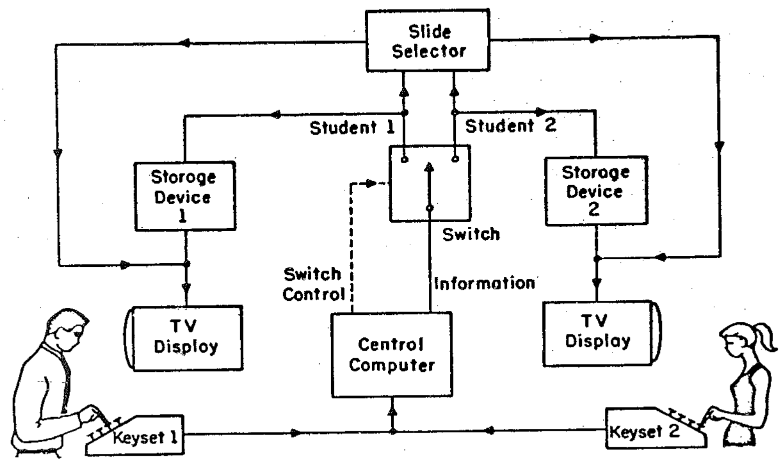


Figure 1: General Organization of PLATO II Equipment

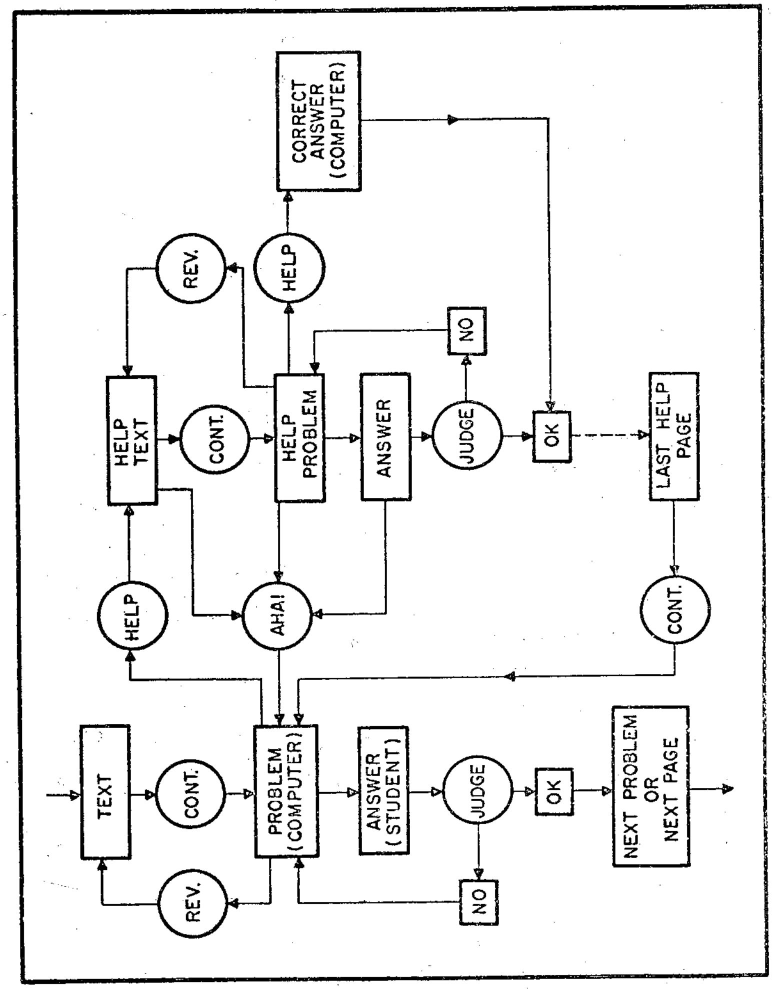


Figure 2: PLATO II Programmed Teaching Logic

The logic (see Fig. 2)

The lesson proceeds in the following way : each student - quite independently of the others - is shown a sequence of slides, which is called the main sequence : this sequence contains the material to be taught (text slides), and asks the question (question slides) which the student is to answer correctly.

After each text slide, the student pushes a button labelled "continue" and the following slide is provided. If it is a "question slide", the student must answer every question correctly before he is allowed to go on to the next slide. He types his answer and pushes the "judge" button : the computer writes O.K. or NO.If NO, the trainer tries another answer, a.s.o/. When he is in difficulty, he may ask for past material (button "reverse") or obtain additional aid by pushing the "help" button, which begins a help sequence. Each problem of the main sequence is associated with an appropriate help sequence, which serves to lead the student to a better understanding of this particular problem.

At the end of the help sequence - or before - at any point he wants - he may come back by pushing the "Aha" button : he is then routed to the main sequence and he answers again the primitive question. If he is still unable to make the right answer, he may "give up". The size of the memory precluded, in the first version, the program from having secondary help sequences : so asking for "help" in a help sequence was equivalent to giving up. But it is obvious that secondary help sequences, can be very easily stored in a more recent and powerful machine.

The principal features of programmed instruction by computer appear very clearly in the PLATO project.

1) it is in the full meaning of the words a "teaching device" because it provides new knowledge
2) it tests how this knowledge has been understood, by the use of a very strong feed back between machine and student. The answers are judged by the computer and not by the student.
3) The computer keeps a detailed record of the progress of the student (history of events" see Fig. 3) and prints it after the end of the lesson. This record is given to the student with a booklet

PLATO II: HISTORY OF EVENTS

STUDENT 2, RUN , LESSON , CHAPTER .

TIME IN MINUTES	PAGE NUMBERS (MAIN)	(HELP)	ANSWER	OK/NO
0	1			
.3			JONES	OK
.4	2			
.8	3			
1.1	2			
1.2	3			
1.7			ACCUMULATOR	OK
2.0			DISTRIBUTER	NO
2.6			DISTRIBUTOR	OK
2.8			NO	OK
2.8	4			
2.9			10	OK
3.0			SIGN	OK
3.1			20	OK
3.1			SIGN	OK
3.2	5			
3.3			B	NO
3.8		108		
4.0		109		
4.1			2.0	OK
4.3			+1.0	OK
4.3	5			
4.4			A	OK
4.5	6			
4.8			OVERFLOW	OK
4.8	7			
5.2			T	OK
5.4			F	OK
5.5			T	OK
5.6	8			
5.9			DA	OK
5.9		117		
6.1	8			
6.3	NO MORE HELP			
6.4	HELP-HELP GIVEN		OP	
6.6			DA	OK
6.7			IA	OK
6.9	9			

Figure 3: Sample Printed Student Record

Student/ Lesson	A	B	C	D	E	F	G	H	J	Row Ave.
1	25.7	11.3		23.2		20.2		29.8	13.8	20.7
2	23.7	40.6	38.3	33.4	22.5	24.8	36.0	46.7	15.0	31.5
3	21.5	30.7	32.9	39.2	33.2	35.0	43.7	54.2	20.4	34.5
Col.Ave.	23.6	27.5	35.6	31.9	27.9	26.7	39.9	43.6	16.7	

Time (minutes) spent on lessons; missing entries indicate mechanical difficulties which resulted in loss of data.

Student/ Lesson	A	B	C	D	E	F	G	H	J	Row Ave.
1	1	5		16		3		5	6	6.0
2	8	6	19	8	4	8	8	9	6	8.4
3	2	16	18	18	7	13	24	18	15	14.6
Col.Ave.	3.7	9.0	18.5	14.0	5.5	3.0	10.7	10.7	9.0	

Number of wrong answers.

Lesson/ Student	A	B	C	D	E	F	G	H	J	Row Ave.
1	2	0		4		1		3	1	1.9
2	5	1	7	4	2	1	6	6	3	3.9
3	2	1	7	6	2	3	6	11	4	4.6
Col.Ave.	3.0	0.6	7.0	4.7	2.0	1.7	6.0	6.7	2.7	

Number of helps requested.

Figure 4: Examples of synthetic records

Figure 5. Reproduced from Computer-aided teaching in Medical Diagnosis by W. Feuerzeig, P.Munter, I. Swets, and M. Breen in I. of Medical Education 39, August 1964, 746.54.

PNEUMOCOCCAL PNEUMONIA

Are you absolutely sure?
    YES.
You can't be until you have irrefutable evidence. What would help you nail down the diagnosis?
    EVIDENCE FROM SPUTUM SMEAR AND CULTURE.
Good. Anything else?
    NO.
I think you might also want evidence from x-ray findings. Go ahead and we'll see what happens.
    X-RAY?
Not ready yet. Try the blood lab.
    DIFFERENTIAL?
Marked shift to left. Even though all the lab reports aren't in, you have enough information to make a presumtive diagnosis. If you care to make one go right ahead.
    ACUTE PYELONEPHRITIS.
It's getting hard to hold onto this diagnosis with neither history nor signs of dysuria. Moreover the pain is getting away from the flank region into the lower belly. Still thinking of a renal problem? Reconsider.
    TB PNEUMONIA.
Any other thoughts as to diagnosis? The sudden high rise in WBC and differential is against it and you don't have tubercle bacilli in the sputum.
    URINE EXAM?
The urine test shows presence of albumen, few hyaline casts, few WBC, occasional RBC.
    X-RAY?
The x-ray shows: area of consolidation, right lower lobe, consistent with pneumonic process. This is a good time to review the total findings and make another tentative diagnosis - if you like.
    EVIDENCE OF FLUID?
No evidence of fluid.
    SPUTUM SMEAR?
Gram positive rods.
    STREP PNEUMONIA.
Fine, but how can you distinguish this from any other bacterial pneumonia? Wait until you have an irrefutable case.
    VIRAL PNEUMONIA.
But what about shaking chills, rusty sputum, elevated WBC, etc. - all the evidence of an acute febrile bacterial process? You are in error, doctor.
    SPUTUM CULTURE?
Presence of pneumococci.
    STREP PNEUMONIA.
Incredible! You have disregarded the positive findings that should clearly have established the correct diagnosis.
    PNEUMOCOCCAL PNEUMONIA.
Very good. It's obvious now, though there were some confusing signs and symptoms in the picture. You've seen your way through to the correct diagnosis.

THE INFORMATION RETRIEVAL PROBLEM IN MEDICINE

By G. Wagner, Professor, M.D., Deutches Krebsforschungszentrum, Heidelberg. Germany.

Within the scope of our meeting, I have the pleasure to speak about the problem of information and information retrieval in medicine. Although the word "Information" applies to different kinds of systems (references, facts and data), the term "Information Retrieval" customarily is restricted to the retrieval of printed information. But, as I have been asked to use the term in its broadest sense, I will also stress some problems of retrieval of facts and data. However, as these fields together are so broad and complex and furthermore only partly explored, my paper must necessarily remain very fragmentary.

Let me start with some general remarks upon the Information problem in science.

There is no doubt that the rapid development of all sections of science may be called one of the most outstanding characteristics of our age. The extent of this expansion may be illustrated by the statement of the American science historian de Solla Price (1963), that out of all scientists that have ever crowded our planet, about 85% are living in our time. But the rapid devebpment of science, exercising an essential influence on our present form of society, our standard of life, and our world-view, also raised problems that have never existed before or, at least, did not become evident in acomparably aggravating fashion, e.g., the increasing difficulty of keeping oneself informed on scientific progress.

Until some few decades ago, it was not too difficult for a scientist to keep himself up to date in his speciality and also to get sufficient

NB References at pages 574 - 575
Figures at pages 576 - 579

information about the progress made in the adjoining disciplines. Nowadays - at least in the sector of natural science and technology - there is hardly any scientist able to overlook the entire literature of his special working field, still less, to obtain a general survey of the adjoining scientific branches. Principally two factors complicate the scientist's survey of the literature that is relevant or at least potentially important to him: on one hand, the increasing number of publications, on the other hand, the growing difficulties of communication between the different scientific specialities.

During the last years, the growth of literary production in the scientific sector has often been labeled as a "publication explosion" or "literature flood". But considering the situation from an impartial viewpoint, the exception must be made that these slogans have pretty often been abused by sensation mongering, and that things have been magnified deliberately (Coblans, Wagner (1955)). The purpose of those exaggerations is - as Shaw (1962) mocked - " to scare us into accepting radical solutions to the documentation problem, for the alternative pictured is that of science dead of autointoxication, from immersion in its own effluvia". However, though a certain reservation as to bombastic overstatements seems to be imperative until better founded data will be presented, one thing is a matter of fact: never before mankind has faced such an inundation by printed paper as it does at present. According to American estimations (Kent (1960)), every sixty seconds - Sundays and holidays included - about 2000 pages of newpapers, periodicals or books are being produoed. The total of scientific periodicals actually published on earth are valued at no less than 30 000; the estimated number of the articles published in then annually fluctuates from 1.5 to 6 millions.

In the field of chemistry, the number of scientific publications is said to have nearly doubled about every eight years during the last decades (Pietsoh (1962)). Wheras in 1961 some 150 000 papers appeared in this scientific sector, an amount of about 300 000 publications are expected to come out in 1970 (Kent (1962b)).

The determination of the growth rate in biomedical literature is much more difficult. In all probability, the expansion of medical literature is not so strong as the increase of publications in the chemical field.

INPUT/OUTPUT HARDWARE

By J.F. Davis, M.D., Director, International Institute for Medical Electronics and Biological Engineering, Paris, France.

In this ensemble of relationships which we have called here the "interface" which surrounds the use of computers in the hospital, there are many problems of communication. The particular one I shall talk about is the interface between man and machine. We also spoke about interfaces between man and man and between machine and machine. But I have been given the job of talking about the input and output devices, where man and computer come face to face. Even though our carefully selected speakers have gallantly avoided too great an emphasis on the input/output hardware which they have used to solve their respective communication problems, preferring to leave the subject to me, their secrets have been exposed none the less, so it is hardly necessary for me to run down the detailed list of available machinery - you have already seen most of it in slides already presented. But perhaps a few generalizations are in order.

You are aware of the classical I/O devices - I mean the nonspecialized hardware that is applicable to any field. Perhaps is is hardly suitable to call "classical" those techniques which have only been in use a few years - for example, the C.R.T. display which, along with its keyboard or light-pen, is such a natural in good human engineering that its proliferation in a very few years has been phenomenal and now at least a dozen makers are offering different versions. It will surely take its place as a "classic" amongst D.P. terminals.

Mr. Marchman referred to the production-planning aspect of hospital work as being similar in many respects to production planning in the business world (we know that a part of every nurse's and every doctor's duty and the major part of an administrator's work falls in this category). It is obvious that terminals developed for the entry, exit and transmission of data in any management information system will serve very well for those same functions in a hospital system or a system serving a group of hospitals. Therefore it is not surprising to find standard keyboard devices, punched card and punched tape devices working very well in hospital installations for admissions, transfers, requisitions, doctor's orders, daily reports, census-taking, etc.

NB Table at page 586

These are essentially information-bookkeeping and message-switching functions which do not create much demand for the "conversational" type of communication with the computer nor for extremely rapid enquiry and response. Beyond this point we run into difficulty with these classical I/O devices.

But in the clinical management end of our hospital system we also deal with information-bookkeeping and with message-switching (not essentially different in conceptual terms from ordinary business information but, as we shall see, greatly different in dimensionality). Here is where we are led up the garden path - we are tempted to use the same I/O hardware for work in connection with our clinical duties - but now we want it, not only in time-sharing configuration (because there are many of us making demands at the same time) but, of course, on-line as well - and with "zero" response time.

At least it is fortunate, that what looks like zero time-delay to a human may be many thousands of operation cycles to a modern processor. But even at that, our brains find our fingers very clumsy, so that many input procedures seem awfully slow to us. Such neat devices as the common dial or the newer pushbutton (touch-tone) telephone seem obvious, at first, for the registration of simple messages, but even the most patient of users (like Dr. Weil) soon complain about their slowness. And the much more complete keyboard devices such as teletypewriters and full panels of pushbuttons, each serving different purposes, are also limited severely in the amount of "communicating" you can do in a given space of time. As an aside here, I would like to refer to Professor Tybjaerg-Hansen's opening talk where he referred to the choice between using plain language and coded messages at the input. Where the typewriter mode may give a better impression of talking to the computer, the push-button array, by virtue of its densely packed coding may get more information "down the

TERMINAL	INPUT RATE	OUTPUT RATE	MONTHLY RENTAL or EQUIV. ($)
Dial (Telephone type)	1-3 ch./s.	--	- 10
Pushbutton Array	2-5 ch./s.	--	20 - 50
Typewriter	5-10 ch./s.	0-15 ch./s.	100 - 200
Paper Tape terminal	100-1000 ch./s.	50-100 ch./s.	150 - 400
Punched Card terminal	100-1000 cd/m.	100-250 cd/m.	150 - 500
C.R.T. Display	---	1 kch.- 4 kch./s.	80 - 250
Line Printer	---	300-1200 lines/m.	500 - 1000
Magnetic Tape Drive	20-100 kch./s.	20-100 kch./s.	2000 - 3000
Digital Plotters	---	100-300 incr./s.	200 - 500
Hard Copy Generator	---	3-60 pages/m.	2000 -
Document Reader, mark	30 doc./m.	--	200 - 400
Document Reader, char.	400 doc./m.	--	500 - 2000
Analog-Digital conv.	1K-50K conv./s.	--	200 - 1000
Computer	100 Kch./s.	100 Kch./s.	500 - 5000

Table I

SUBJECT INDEX

input source so that from the output the computer will be able to evaluate the gain of each channel, the time constant, the damping, and so on.

A last aspect of data acquisition is the question of being able to intervene, to make some kind of dialog between us and the source of data (the patient) during the acquisition. By that I mean not only to be able to make stimulation, but to judge more or less instantaneously from the evolution of the physical phenomenon in order to act in relation to this evolution. This dialog between the observer and the patient necessitates some kind of more or less instantaneous pattern recognition.

The control of the various specifications that we could then enumerate necessitates the use of a very particular organisation. It is necessary to have personnel of a very special kind, having what has been called in French at least "géni médical". This personnel was not evidently necessary when the doctor was dealing alone with his patient; it was not even necessary at the research laboratory when we could deal with a unique instrument, an electrocardiograph or an electroencephalograph, because then the observer (the scientist) was quite able to deal with the various little problems presented by that instrument. But when there is a series of instruments and of connections which are all important for the final success of the computation, then the observer has to have collaborators who will enable him to attend to all those specifications.

SPECTRAL ANALYSIS TECHNIQUES AND PATTERN RECOGNITION METHODS FOR ELECTROENCEPHALOGRAPHIC DATA

By W. Ross Adey, Professor, University of California, Los Angeles, USA.

INTRODUCTION

From the melting pot of relatively modest experiences in data processing that, in most aspects, began little more than five years ago, the neurophysiologist has progressed to substantial use of large computing systems, and has taken account of his need for rapid processing of large amounts of data by comprehensive computational techniques that go to the very fringes of the mathematical arts. Granted that this rapid growth in analytic capabilities has not arisen de novo, and that it has its origins in earlier methods of frequency analysis (Grey Walter, 1950) and simple averaging and correlation analysis (Dawson, 1950; Brazier and Barlow, 1956), the exponential growth in the **armamentarium** of the neurophysiologist's analytic capabilities represents a series of essentially new developments. They rest upon a trinity that will be a recurring theme in this paper: data acquisition systems using analog or digital magnetic tape recording techniques, with appropriate coding for stimuli and epoch marking; the use of statistically valid analytic techniques, that take account of uncertainties inherent in limited epochs of physiological data; and automated display techniques that achieve required degrees of compression of the primary records to provide an overview of long and complex epochs of data, while retaining fine resolution of subtle shifts in pattern within the epoch (Adey, 1965a,b).

NB Supported by the United States Air Force Office of Scientific Research (Contracts AF-AFOSR-246-63 and 61-81), and the National Aeronautics and Space Administration (NsG502, NsG505, NAS9-1970), and the National Institutes of Health (NB-02501)

References at pages 423 - 425
Legends to figures at pages 426 - 428
Figures at pages 429 - 438

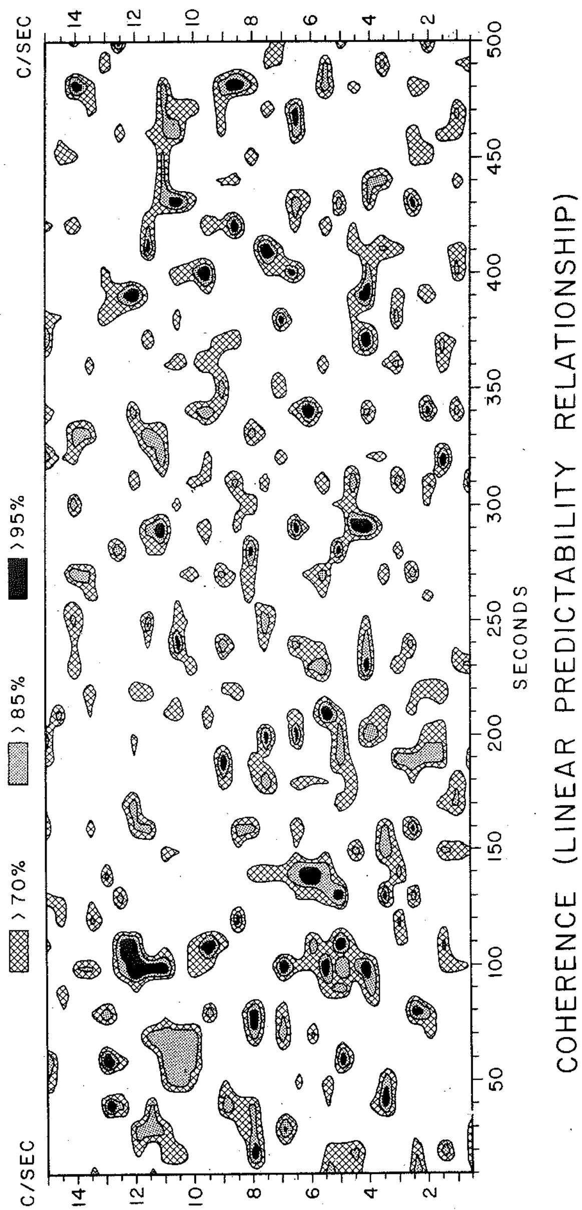


Figure 10

ON DATA PROCESSING OF THE IMPULSE RESPONSE FROM A SINGLE NERVE CELL IN THE BRAIN

By Christian Guld, Lecturer, Cand. Polyt., Institute of Neurophysiology, University of Copenhagen, Denmark.

The electrical activity of the brain can be measured in two essentially different ways.

The summed electric activity from many cells in the brain can be picked up with electrodes which are large in relation to the size of a single nerve cell. This electroencephalogram is usually recorded from the surface of the brain, but can be measured by electrodes placed deep in the brain structures as well. The EEG waves have amplitudes of up to 300 µV and are normally characterized by low frequencies of about 10 c/s.

The other way to study the processes in the brain is to measure the activity of single nerve cells by means of microelectrodes, which have a leading-off area which is small as compared to the size of the single cell. Repetitive firing of the action potential of the cell can be recorded in this way, the single potential having an amplitude of 1-10 mV or more and a duration of about 1 msec. The potentials appear at irregular time intervals ranging from a few msec to several seconds. It is usual to characterize this as a pulse-time modulated signal and I shall present some physiological experimental situations and some problems in data processing of the information in such records.

Microelectrodes have tips with diameters of the order of 1 µ. It is possible to place the tip of these thin electrodes near to or in a cell without severely affecting the dendritic structure around the cell.

The great difficulty in recording from single cells with a microelectrode is to keep the electrode in place near the cell. Due to movements of the

---

NB References at page 447. Figure at page 447.

AUTOMATIC EEG TAPE-COMPUTER SYSTEM FOR ANALOG-DIGITAL DATA PROCESSING OF FREQUENCY-SPECTRUM INFORMATION

By Edmund Kaiser, Kaiser Laboratory, Copenhagen, Denmark, and Ingemar Petersén, Laboratory of Clinical Neurophysiology, Sahlgrenska Sjukhuset, Gothenburg, Sweden.

EEG, has during the past 30 years, attained increasing importance as a tool in clinical diagnosis. The qualitative performance in most routine EEG work has, however, practically remained unchanged during the last 25 years, except that the EEG machines have been provided with more and more channels and a higher technical standard. The increased number of EEG-examinations with increasing number of channels used causes, together with the enormous variability of EEG (Related to wakefulness and sleep, age, bloodchemistry, sex, handedness, position of electrodes, clinical diagnosis etc.) increasing difficulties of correct description and interpretation of an individual EEG. Great difficulties are also encountered in dealing with differences between different groups of individuals.

Obviously there is a demand for a more precise description of EEG-phenomena and their relation to clinical diagnoses. In fact several valuable methods have been developed to obtain a mathematical and/or statistical description of the EEG.

METHODS OF MATHEMATICAL OR STATISTICAL DESCRIPTION

1. The simplest method ought to be a statistical examination of the time intervals between consecutive base-line crossings or pairs of base-line crossings, as in determining the "alpha index" (Davis and Davis, 1936) and of the "delta index" (Hoagland, Rubin and Cameron, 1936; modified by

---

NB Summary at page 458
References at pages 460 - 462
Legends to figures at page 463
Figures at pages 464 - 468

Frey, 1946). This method was later automated by Stein, Goodwin and Garvin, 1949; Prast and Noel, 1949; Yound, 1954.

2. A quantitation of EEG is also possible via automatic amplitude measurement as described by Drochocki, 1939; Drochooki and Drohocka, 1939. (Via the Drohocki method it is possible to obtain the cumulative distribution function and via the derivative the distribution density.) King (1951) describes an amplitude-distribution analyser for the frequency range from DC to 1000 Hz. The determination of amplitude-density distribution is a valuable tool in comparing EEG phenomena with mathematical models (Saunders, 1963).

3. A combination of the methods under 1 and 2 has been suggested by Lonsdale (1952) showing the width of successive waves at several preselected amplitude levels to give the percentual time for which the function was equal to or greater than the selected amplitude levels. A cumulative distribution curve characterizes in this way the wave pattern.

Another example of a combination of the above mentioned methods is a "Yes-no" pattern analyser, described by Kaiser and Sem-Jacobsen (1962). This method is based on a number of detector outputs, stating whether the amplitude and/or duration is higher than several preset threshold values and whether the amplitude and/or duration exceeds the preceeding three second average by more than 50%. The combination of answers are presented as characters by means of a digital tape puncher. The last mentioned way of pattern detection could be used in a more simple way as an event detector.

4. The above mentioned methods only deal with the statistical treatment of duration and/or amplitude of EEG. In order to quantitate the apparent periodicity in the EEG-activity a mathematical analysis of Fourier type would yield a frequency-amplitude spectrum describing the wave pattern. The first analog approach was taken by Grass and Gibbs in 1938. An automatic high Q multifilter low-frequency analyser was described by Walter (1943 a and b; Baldock and Walter, 1946). A purely optical method based on the principle of wave diffraction giving a continuous frequency power spectrum was described by Krakau (1951, 1953).

The vast amount of spectral information presents, in itself, a "data

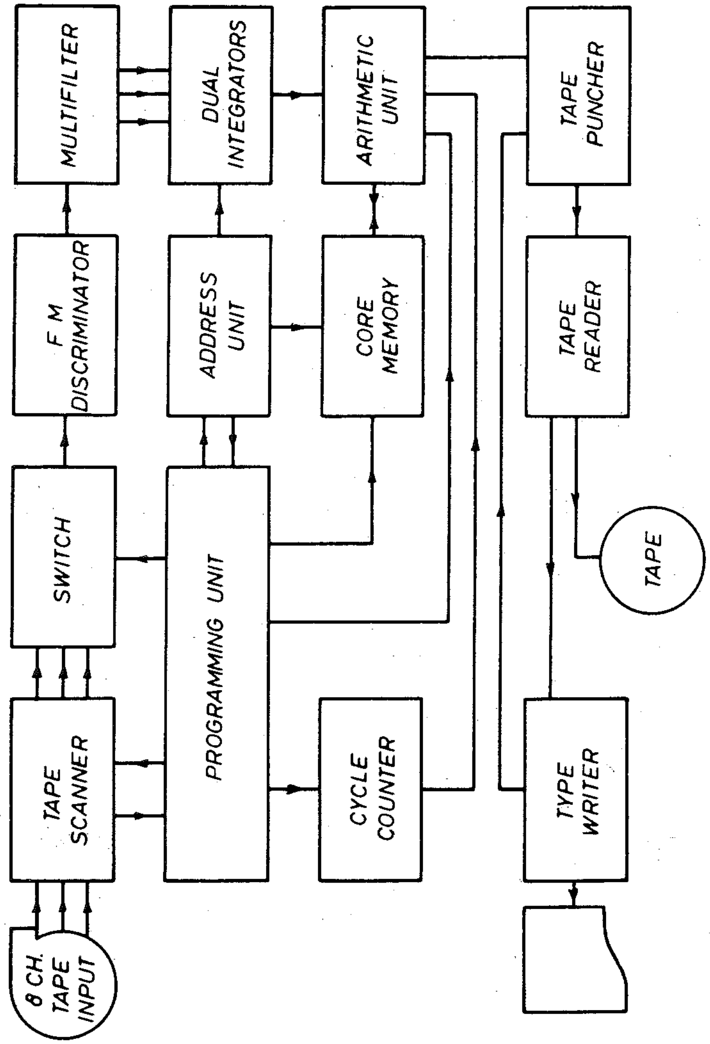


Figure 5

CYBERNETIC METHODS IN MEDICINE: SOME RECENT EFFORTS AND AN OUTLOOK

By Joseph Wartak, Computation Centre, Polish Academy of Sciences, Warsew, Poland.

ABSTRACT

After a brief outline of the fundamentals of cybernetics the author indicates several fields of medicine, neurophysiology, biochemistry and diagnostics which have already been aided by the application of the analytic methods of cybernetics.

The author, then, describes the operation of making a diagnostic decision in terms of information theory. He postulates the use of probabilistic models for decoding and processing of medical signals and symptoms as an aid to diagnosis. Next the author defines computer programming based on these postulates. Finally, he indicates the importance of using computer techniques for making much more rational diagnoses and shows some perspectives for the near future when the machine-patient relationship is established.

Cybernetic methods are being increasingly applied to problems of biology and medicine. By making use of them, we can come up with a quantitative approach to activities on the psychological, physiological, cellular or biochemical levels which not so long ago were described only in qualitative terms.

In short we can define Cybernetics as the science which deals with the processes of receiving, processing, storing and transmitting messages between living and not living systems mainly for the purposes of communication and control.

In the latter case there exists a feedback mechanism which is based on appropriate processing of information to diminish deviation from the programmed state (for example, the state of equilibrium).

The theory of information is used in cybernetics to deal with the problems involved in coding and decoding of messages, measuring the quantity of information contained in messages and transmitting information through the channels.

A simplified picture of an information system can be represented schematically as shown in Figure 1. For a telegraph the meaning of the boxes is obvious. For the process of hearing, the originator might be a piano player. The coder would be the piano. The transmission line represents the air. The decoder is the ear of the listener, and the receiver is his central nervous system. Similar analogies can be made for the synthesis of proteins, for vision, etc.

Mathematically, a originator is defined by stating the repertory of messages it can generate and, for each message, the probability that it will occur. These messages are not directly communicated but fed into an effector organ, called a coder, which modifies or codes the message to make it compatible with the process of communication. The medium into which the encoded message is fed is here called a channel. Also fed into the channel, and mixed with the message, are perturbations generated outside the system; this is called noise. Mathematically the channel is specified by the repertory of messages it can receive, by the repertory of messages it can pass on, and by a matrix of conditional probabilities stating the likelihood of any given output message resulting from any given input message. The mixture of message and noise is transmitted to a decoder which feeds into the receiver.

Some authors expect that the theory of information will play a role in biology similar to that of thermodynamics in physics and chemistry. The theory of information has already been applied to the problems of the functioning of the nervous system, protein synthesis, antigenio specificity, polymorphism in hematology, the action of ionizing radiation, aging, genetics and other problems.

This report attempts to describe a diagnostic decision-making process as a communication between the patient (originator) and the physician (receiver), a process to which the theorems of information theory can be applied.

Under the influence of various pathogenic factors - e.g. bacteria, viruses, injuries, etc. - a healthy organism can undergo pathological transformation so as to manifest a series of symptoms such as fever, changes in blood, in urine, etc. which may be regarded as signals s_1, s_2, ... s_n.

The term "signal" should not grate the physicians ear, considering the fact that it very often more adequately expresses the physical reality. For instance, the electrical phenomena of the working heart muscle (both healthy and diseased) possess all the features of an electrical signal.

Also it would be better to call the yellow colouring of the skin in jaundice a signal in comparison with the yellow signal in traffic lights.

We may consider these signals as elements of a certain set $S = \{s_1, s_2 \ldots\ldots s_n\}$ which we assume to be finite.

The signals never occur individually but in groups (blocks) which we shall call signal-complexes and shall denote by S^c.

Thus the disease itself may be treated as a message coded in a certain signal-complex S^c and transmitted with a certain frequency by a patient's organism.

Medical knowledge, on the basis of many years of experience, also establishes a certain manner of coding this information, that is, a certain representation φ of the set of diseases (messages $D = \{d_1, d_2, \ldots\ldots d_m\}$ into the set of signal-complexes S_j^c, $(j = 1, 2, \ldots\ldots m)$.

If a disease (message) d_i occurs in a patient then it is coded into appropriate signal-complex S_j^c which will appear with a probability

$$P\left(S_j^c / \varphi(d_i)\right)$$

We can group messages transmitted by a patient's organism into classes, which we identify with diseases. We shall assume that each of these classes is defined by several representatives which we know to be members of an established class.

A physician who receives signal-complex S_j^c has to find out which class (disease) has been transmitted, that is, he has to determine the conditional probability that the message d_i has been transmitted on the condition that he has received the signal-complex S_j^o. He assumes then that a message has been sent, for which the probability $P(d_i/S_j^c)$ is the greatest.

From the definition of conditional probability it follows that

$$P(S_j^c \cdot d_i) = P(S_j^c) \cdot P(d_i/S_j^c) = P(d_i) \cdot P(S_j^c/d_i)$$

hence

$$P(d_i/S_j^c) = \frac{P(d_i) \cdot P(S_j^c/d_i)}{P(S_j^c)}$$

Since any signal-complex S_j^c cannot occur unless together with one of the mutually exclusive diseases $d_1, d_2, \ldots\ d_m$ we get

$$S_j^c = S_j^o d_1 + S_j^o d_2 + \ldots\ldots + S_j^c d_m$$

and then

$$P(S_j^c) = P(S_j^c d_1) + P(S_j^c d_2) + \ldots\ldots + P(S_j^c d_m)$$

hence, according to the theorem of the probability of the product of two dependent events,

$$P(S_j^c) = P(d_1) \cdot P(S_j^c/d_1) + P(d_2) \cdot P(S_j^c/d_2) + \ldots\ldots + P(d_m)P(S_j^c/d_m)$$

or, shorter

$$P(S_j^o) = \sum_{\mu=1}^{m} P(d_\mu) \cdot P(S_j^c/d_\mu)$$

thus we obtain

$$P(d_i/S_j^o) = \frac{P(d_1) \cdot P(S_j^c/d_i)}{\sum_{\mu=1}^{m} P(d_\mu) \cdot P(S_j^c/d_\mu)}$$

where $P(d_1)$ denotes absolute probability of occurence of the ith disease;

$P(S_j^c/d_i)$ denotes the conditional probability that a signal-complex S_j^c will occur in a patient having the ith disease; and

$P(d_i/S_j^c)$ denotes the conditional probability of the occurence of the ith disease if a signal-complex S_j^o has occurred in the patient.

As can easily be seen, the last equation is Bayes' rule for the probability of causes. On the basis of the event that has occurred, that is, on the basis of the signal-complex S_j^o received, this equation allows us to define and choose the most probable hypothesis which will explain this event.

The most probable hypothesis (disease) will be that for which the probability $P(d_i/S_j^c)$ is the greatest.

It may happen that there will be two or more diseases which have the same conditional probability $P(d_i/S_j^c)$ for a given complex of signals received. The physician will then choose one of these diseases which he regards as the one actually presented; his choice will be entirely subjective.

A received signal-complex S_j^c can be interpreted as a point or as a vector in the N-dimensional abstract space. Each of the N-dimensions represents a different signal. The points corresponding to the representatives of established classes make sets which we can call "clouds", that means that distances between the points representing the same class are smaller than distances between points representing the various classes.

In such a hypothetical N-dimensional space, we can calculate for each signal-complex S_j^c the probability of its occurring, on the assumption that the ith disease is present, i.e. $P(S_j^c/d_i)$.

This can be obtained by using the well-known multidimensional linear regression method. First we compute coefficients of covariance inside the data matrix S_{ji}^c. In this way we obtain the covariance matrix A_i $(i=1, 2, \ldots\ m)$. For each covariance matrix A_i must be found the value of the unknown λ which satisfies the equation:

$$\left| A_i - \lambda I \right| = 0$$

Here the matrix I is a unit matrix, of the same rank as A_i but with unity in the leading diagonal, and zero elsewhere.

To each of the latent roots, being a value of the unknown λ , there is a corresponding latent vector of the matrix, and each of these vectors is orthogonal to all of the others.

In such a way, the latent vectors make a new uncorrelated system of coordinates; relative to which we oan compute the coordinates of the signal-complex S^c_{jk} received from any unknown patient K.

Using this technique we can obtain for a signal-complex S^c_j received from patient K the distribution density defined for eaoh area containing the points of a given set ("cloud"). In other words we can obtain the value of $P(S^c_j/d_i)$ that represents the probability with which, in a patient K having the disease d_i, signal-complex S^c_j will occur.

Assuming a priori that the distribution of the set of points is normal (which can be checked on a computer) we can show that the minimal areas having with given sizes the highest probability of containing the points of the set considered, are the corresponding hyperelipsoids.

Thus a whole diagnostic program has the following steps, which we can represent schematically as shown in Figure 1 (see page 478).

This program allows us to obtain a meaningful index of probability that an unknown patient has a speoified disease.

It seems that the formalization of the process of establishing medical diagnosis is now possible on the basis of both medical knowledge, which assigns certain blocks of symptoms (signals) to particular diseases, and of the theorems of information theory. What presents the most difficulty here is the lack of appropriate medical statistics, which would enable us to determine the probability of incidence of particular diseases and their symptoms. There seems to be no doubt, however, that the observation of great numbers of diseases as well as appropriate statistical elaboration of this material will make it possible to arrive at exact numerioal

data of the following kind: the probability of the incidence of disease d_1 is 0.01, the probability of the signal-complex S^c_j occuring in disease d_1 is 0.5, etc.

The importance of such a formalized approach is due to the fact that the process of establishing a diagnosis may be considered as a strictly defined and objective one. It can occur, too, outside the mind of the physician. It may then be presented in the form of an algorithm which is solved automatically by e.g. electronic computers. By algorithmization of the process of establishing diagnosis we bring to medical diagnosis the fields of exact sciences such as statistics and the calculus of probability, thus increasing its reliability.

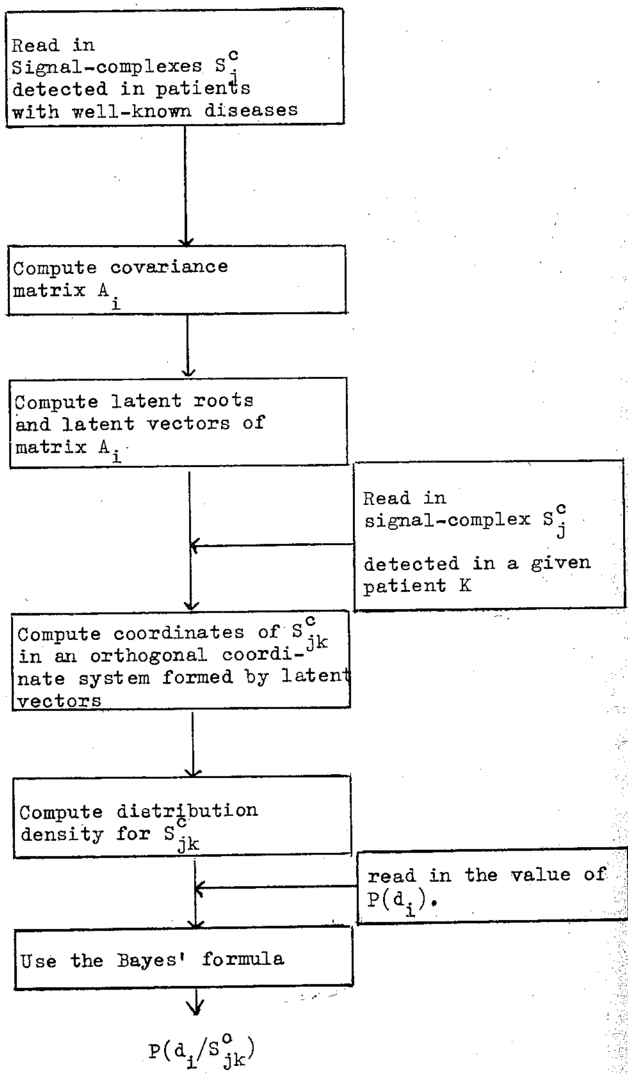
The process of reproducing information contained in signals and symptoms occurs, so far, in the mind of the physician, which makes it liable to powerful "noises". "Noise" may denote here psychological factors (e.g. optimistic or pessimistic attitude, fear, bias, egoism, conceit and self-assertion, making efforts to arrive at "a particularly interesting diagnosis") or ignorance, basic or due to forgetting, which disturb the reproduction of information transmitted by the diseased organism. Often, even experienced physicians come to different diagnostic decisions on the basis of identical symptoms, which they find in the same patient; in other words, they reproduce different messages (different diseases). The "noise" might apparently be eliminated or, at least, reduced to a low tolerable level if diagnostic procedure could be transferred to a system of automatically operated electronic computers.

In the not too distant future every man shall be supplied with a certain number of transducers made of monocrystals which can record many quantities (in a physical sense) which describe the functions of an organism. From time to time, perhaps very often (for example, every week or every day), the examinee shall "connect himself" to a detector unit situated in a convenient place (for example in the office, factory, or school). The results of such measurement shall be sent immediately to the central computer system located, perhaps, very far away, at which place they shall be processed. If the computer detects a deviation indicating the beginning of a disease it shall immediately call another device which shall order the patient to contact the physician. At the time the patient visits the physician the computer system shall supply the whole information of the patient and will advice which treatment plan should be used in a given case.

Let me finish by stating that I am no longer alone in the firm conviction that in the very near future a great revolution will occur in medicine, which depends on almost complete depersonification.

Figure 1

Diagram of the computational process involved in medical diagnosis.



LOGIC OF DIAGNOSIS AND ITS PROCESSING BY COMPUTER - WITH RESPECT TO CONGENITAL HEART DISEASES AND BRAIN TUMORS

By Kosei Takahashi, M.D., University of Tokyo, Japan.

I. DIAGNOSIS AS A SCIENTIFIC INFERENCE

Systematic medical knowledge and techniques are of course the indispensable basis of medical diagnosis -- yet the storage of an enormous number of memories in our brains does not perforce mean that we will be excellent in diagnositic practice. Diagnostic thought is a sort of selfdeveloping inference, comprising a design or plan for examination, plus decision-making.

If enough information is available, deterministic logic may be the best way of scientific inference. The trouble is that our knowledge on pathological changes, when they exist deep inside the living body, is not necessarily decisive but appears rather uncertain. Moreover, we must labor under a severe restriction: we cannot perform destructive or disassembling tests, which are usual in tracking down information in the industrial field. Moreover, we must pay attention to the dissociation between the deductive description in the medical books and the inductive thought of diagnostic inference: there is an obvious inversion of thinking process.

Medical diagnosis, as a result, employs the process of probabilistic inference plus decision-making on integration of information which are uncertain and insufficient. This is evident in the statistics on the frequency of symptoms in the various diseases, though decisive information can be provided, in part and in approximation.

If the logic of diagnostics can be clarified, it will be able to be processed by computer, how complicate it might be. The orderly, systematized taxonomy of diseases in the medical books is not always as useful as it

NB Figures and tables at page 490 - 498.

might be in diagnostic practice, since it is built up of all sorts of data, even autopsy findings.

SOURCES OF DATA:

1. Data on congenital heart disease (N = 187) were supplied by the Department of Surgery, Tokyo University.
2. Data on cerebellar tumors (N = 149) came from the Department of Neurosurgery, Tokyo University.
3. Data on obstructive jaundice (N = 67) obtained from literatures.

In all cases, the diagnosis was confirmed by operation.

2. INITIAL STAGE OF DIAGNOSTIC THOUGHT

a). Developing a Symptom Vector

After a chief complaint is given, at the initial stage of diagnosis, questions by the doctor toward the patient may be raised from the view point of patho-physiology. That means to ask symptoms or to try to find signs which are known to have large correlations to the complaint. Therefore no vivid hypothesis on diseases may not be in doctors' brain at this stage.

b). Making a Set of Possible Diseases

After the symptom vector is developed to a definite extent, several diseases come to light as hypotheses to be tested. Generally speaking, one disease is described by a number of symptom vectors and a symptom vector corresponds to a number of diseases. Such a disease - symptom vector correspondence will be simplified for the convenience of memory and stored in doctors brain. Using this, doctors make a set of possible diseases for a given symptom vector. This means the restriction of disease space. Mathematical principle of this procedure may be given by Boolean Algebra, a priori probability problem with regard to making a set of hypothetical diseases will be discussed later.

3. INFORMATION CHARACTERISTICS OF DATA

The acquisition of too much information is not only a real burden to patients, but also seems to confuse doctors in their intuitive judgement.

Although confirmation of a pathological process by several related observations is helpful for full development of intuitive understanding, the actual amount of data will not be important from the viewpoint of information theory. The quality of these facts, however, must be studied in order to get rid of the less vital information. Of course, the relative importance of the data must be considered in regard to a set of diseases to be discriminated from each other, using a process that lists the possible diseases.

a). Selection of Contributory Information

After a set of possible diseases has been prepared, several tests may be used as indicators to find the informative symptoms or findings.

1. Significance Test on Difference of Incidence.
2. $G_i = -\sum_a^m p_{ia} \log p_{ia} + m\,\bar{p}_i \log \bar{p}_i$
3. $I_i = \sum_\alpha^m P(S_{ij}, D_\alpha) \log \{P(S_{ij}, D_\alpha)/P(S_{ij})P(D_\alpha)\}$

In the research on congenital heart disease, we threw out about half of a hundred kinds of information just by inspection, and about half as much again was eliminated by the methods noted. General examination revealed that most of the symptoms were of lesser importance, and only three were retained from clinical examinations. This elimination step is indispensable to save time and trouble during the next stage of hunting down independent information.

b). Uncovering Independent Groups of Information

The next step is to distinguish the independent information among the data selected. This step is necessary in order to eliminate the theoretical difficulties that are linked to the processing of dependent information, especially in the case of discrete variates.

We have factor analysis for uncovering independent factors or groups in a large number of tests. Ordinarily factor analysis is applied to continuous multivariates. However, in view of the deduction process in the basic formula of factor analysis, no assumptions appear to be necessary on the statistical characteristics of variates provided that all variates are standardized.

We tried factor analysis on congenital heart disease cases. These comprised 187 cases described by binomial and/or multinomial multivariates, and ten diseases described by frequencies after the transformation to logarithm, log p_{ia}, or $\theta = \sin \sqrt{p_{ia}}$. From another viewpoint, a direct analysis of data on binomial and/or multinomial multivariates was tried. (Table 1)

In the former case, the total sum of eigenvalues up to the sixth accounted about 95% of total variances. After varimax rotation, test variables were bunched into five groups. The large magnitude of their factor leadings enables us to detect them as follows:

I. Anoxaemia Polyglobulia

II. Overloading on left Auricle

III. Pulmonary Hypertension

IV. Location and Characteristics of Murmurs

V. Overloading on right Ventricle

With cerebellar tumors, most symptoms or clinical findings appear to be independent with the exception of two dependent groups: general symptoms and cranial nerves.

4. DIAGNOSIS BY ALGORITHMIC INFERENCE

In view of the statistical tables of symptoms, it is quite obvious that algorithmic diagnosis is an approximation of probabilistic diagnosis. For the human brain, however, it is convenient to remember and to carry on the thinking process at the same time. No method of establishing the most efficient approximation appears to have been investigated as yet. The requirements would be the minimum amount of information to be used and the maximum discrimination of the groups.

We startedfrom a basis of seven types of tests, with each test representative of one of the independent groups. By trial and error we devised a taxonomic system of congenital heart disease, which on the average can classify cases so accurately that the figure for accuracy is 88%. (Fig. 1 Table 2)

5. DIAGNOSIS BY PROBABILISTIC INFERENCE

In probabilistic inference, the joint probability of variates corresponding to a given disease profile is computed for a series of diseases listed. These values are used as a measure of the likelihood of the disease for a given disease profile. Such a statistical entity which is used in estimating the population for a given sample is called the likelihood, to differentiate if from the probability.

C.R. Rao says:"In statistical literature the term likelihood of the observations is often wrongly used to mean the probability density of the observations. The probability density for a given set of observations may be considered as a function of the parameters which is otherwise termed as the likelihood of the parameters." (Advanced Statistical Methods in Biometric Research, Wiley, 1952, p. 150)

One difficulty inherent in diagnostic inference carried out on the basis of likelihood is that the computation of joint probability is not simple, because of the dependency of biological characteristics. One method is to eliminate all the dependent variables and start out from a set of independent variables, so as to use the ready-made formula for computing probability on the assumption of independency of variates on each other. Otherwise we must investigate the robustness of the formula for neglecting dependency of variates on each other.

a). Discrete Multivariates

In k-dimensional binomial multivariates, the possible disease profiles may comprise $2^k - 1$. If we could acquire a large number of cases, the relative frequencies of these disease profiles would give the probability density needed. In actual fact, this is impossible. So we must have an effective method of estimating the joint probability in dependent multivariates.

Lazarsfeld has developed a formula in which the joint probability is given as a product of A-part and B-part. The A-part is the probability itself which is computed on the assumption of independency of variates on each other. The B-part is a polynomial equation on standardized variates. (Table 3)

From our experience in the analysis of variance, we can reasonably expect that the higher-order terms will become smaller, yet such a monotonically-decreasing tendency is not actually proved, mathematically.

In the case of the cerebellar tumors, we computed the A-part and the approximate value of the B-part truncating the terms of a higher order than fourth degree. The logarithm of the B-part ranges from 0 to 3. As that of A-part ranges from 6 to 13, the computation of joint probability on the assumption of independency of variates can be employed only as a gross approximation, from the viewpoint of the theorists. It will be noted that, in our cases, the monotinically-decreasing tendency in higher-order terms not verified. (Table 4)

Most of the published papers on computer diagnosis on the basis of binomial or multinomial variates have been based on such an A-part of probability, up to the present. Yet in spite of this, the percentages of correct diagnosis are excellent. This suggests the robustness of the formulae of joint probability as used in diagnostic inference, in contrast to the negation effect of dependency -- which must be investigated as a separate problem.

b). <u>Continuous Multivariates</u>

In the case of normal multivariates, dependency of variates is not a problem, if the inversion of the covariance matrix is not troublesome.

The original form of the logarithm of likelihood contains three kinds of terms: constant terms, variance terms and distance terms. If we stand on the assumption of homogeneity of variances throughout the diseases, the variance term and the quadratic part of the distance term may be eliminated, and a linear function of the variates will be obtained. This is <u>Rao's discriminant score</u>. The assumption of independency of variates eliminates the bothersome computation of matrix inversion. If both independency and homogeneity are assumed, a very simplified linear function will be obtained.

Fisher's linear discriminant function gives a linear function on the assumption of equality of covariance matrix by least-square method. <u>Perkal's discriminant function</u> (used in Poland) is given on the assumption of homogeneity and independency. The constant term derived from the distance term is eliminated in both discriminant functions. The degree

of approximation in these discriminant functions should be investigated as a problem of robustness of the formula.

The multiple discrimination method is an expansion of Fisher's <u>discrimination function</u> into several groups. It will be convenient to use for an over-all view of the constellation of diseases, and for estimating the contribution of variates.

C. Cochran shows: if we replace a normal variate by a qualitative variate for the sake of simplicity, a moderate amount of loss of information may take place, according to the number of states r: if $r = 6$, the relative power is 0.942, and if $r = 2$, it is 0,636.

c). <u>Multiple Discrimination Method</u>

When there is a number of diseases to be discriminated, the method of discrimination will be determined by the structure of the symptom vector. If the difference between diseases is derived merely from the quantitative difference of the level of the mechanism behind the symptom vector, one discriminant function may be enough for the discrimination.

However, if the difference between diseases corresponds to the difference of several mechanisms behind diseases, several discriminant functions are needed. In such a case, the discrimination will be performed step by step. In multiple discrimination method, it is performed step by step on principal component axes. Although this method determines the order of discrimination, this pre-determined orthogonality of axes disturbs the efficiency of discrimination. A free multiple discrimination is thought to be more general than a principal multiple discrimination. (Fig. 2)

6. <u>DIAGNOSIS BY RELIANCE SCORING</u>

In case of discrete multivariate, we are able to approach to the approximate value of joint probability, neglecting the dependency of variates to one another and neglecting the assumption on distribution. It was revealed that the latter is not good enough.

Apart from the probability concept, a new method is proposed. This is rather similar to doctors' thinking process which confirms one pathophysiological mechanism with a lot of same kind of information. This is a positive application of redundancy in symptom vectors. In separating PDA from PDA+PH, ASD from ASD+PS and VSD from VSD+PH, a different set of information was selected. When a symptom or finding is positive, one point is accounted. The distribution of total points was examined. In each pairs of diseases, such a simple scoring method revealed very excellent in discrimination. PDA was discriminated completely from PDA+PH. ASD vs ASD+PS with 12.1 vs 8.3% of misdiagnosis and VSD vs VSD+PH with 4.3 vs 18.2% of misdiagnosis. (Table 6)

Before this operation, however, the following should be done: the difference between the incidence of one disease and the other should have the same sign. Otherwise, the signs have to be changed.

7. A PRIORI PROBABILITY IN MEDICAL DIAGNOSIS

From the numerical viewpoint, the Bayesian probability is based on the product of a priori probability of the disease and the likelihood of the symptom vector of the disease.

From the theoretical viewpoint, however, the Bayesian probability is a principle of random sampling from a set of all possible diseases. On the contrary, the likelihood is a principle of testing hypothesis by statistical inference. As we are of the opinion that medical diagnosis is principally a process of testing hypothesis, we accept the likelihood here.

There are several differences in using a priori probability. Sometimes we cannot compute it because of the difficulty in determining the population. Even if it could be computed, it could be changed from time to time and place to place. This may contradict to the absoluteness which is required in medical diagnosis. We are requested not to miss any disease because of its low incidence, if a patient shows a typical symptom vector. Such an undesirable case might occur, if we accept the Bayesian probability as a principle of medical diagnosis.

8. DATA PROCESSING BY HUMAN BRAIN

In our congenital heart disease study, our preliminary data consisted of 23 classes of information, including PCG and ECG, but excluding X-ray and catheterization. The cases comprised ASD, VSD and Fallot. With the

cooperation of 13 top medical students of the third-year class at Tokyo University, 123 cards were diagnosed (punched cards), and compared with the computer's diagnosis. The students' diagnoses were correct in 70% of ASD, 80% of VSD and 82% of Fallot, or and average of 75% of total cases.

The computer's diagnoses were correct in 92% of ASD, 94% of VSD and 98% of Fallot, or an average of 95% of the total, even though the likelihood was computed on the assumption of independency of variates. This percentage is near that of the preoperative diagnoses established on all types of information, including X-ray and catheterization. Experienced physicians and surgeons co-operated in this diagnostic experiment on punched cards. Their percentage of correct diagnosis was about 80%.

In respect to the cerebellar tumors, our data consisted of 29 classes of clinical neurological information, excluding X-ray, ventriculography, arteriography and other complicated tests. Sixteen neuro-surgeons were asked to co-operate in the card diagnoses. They had had from one to ten years' experience. Here the accuracy of diagnoses amounted to 33.9%, or if the misdiagnosis in laterality of cerebellar hemispheres is allowed, it was 44.3%. It will be worthy to note the fact that no obvious connection with years of experience was discernible. Under the same conditions, the computer was able to diagnose correctly by 81 to 84%, even with the rough computation of likelihood, on the assumption of independency of variates.
(Fig. 3)

The preoperative diagnoses vs computer's were correct in 46.7% vs 46.7% of the right hemisphere, 0% vs 91.0% of the vermis, 50.0% vs 76.2% of the left hemisphere, 81,3% vs 87.5% of the right angle, 86.7% vs 80.0% of the left angle, 83.3% vs 66.7% of the pons, 77.8% vs 81.9% of the pineal body, 23.8% vs 81.9% of the fourth ventricle and 20.8% vs 72.0% of the paraaqueductal area.

Most of the former was low in their deepness of the discrimination detecting more rough localisation of tumors as cerebellum, infratentorial tumor or merely brain tumor.

9. DECISION IN MEDICAL DIAGNOSIS

a). Interval Estimation

The likelihood method discussed above is called point estimation, which accepts the hypothesis showing the highest degree of likelihood. If we apply to this an interval method of estimation, several hypotheses may be turned aside, so that the percentage of missing true diseases decreases. Physicians sometimes appear to follow such a thinking process. In regard to the cerebellar tumors, we tried to set a one-sided interval of 0.5 from the maximum value of the logarithm of likelihood. This means that we sequester diseases with one-seventh the value of likelihood in relation to the maximum.

The percentage of correct diagnosis was improved to 92.1%, from 84.3% by such an interval estimation method of diagnosis, while 34% of the cases were reserved for further discrimination of several diseases. This method may be called the collateral-load accompanied-by-interval-estimation method: it is equivalent to the loss related to the cost of observation, which in decision theory is denoted by r_2. In respect to medical diagnosis this r_2 is negligible compared with the loss related to misdiagnosis, denoted by r_1.

b). Risk Function

There are various kinds of loss in medicine, statistically speaking. One criterion may be the death, permanent disturbances, or retrievable social and economic losses of the individual himself. The other criterion relates to the contagion (infectious disease) or disturbance of the lives of other persons (as with mental disease)-- that is, a loss in public health. These losses are not qualitatively equal, and cannot be pooled by any equivalent rate of exchange. We would make specific studies of some of these, as representing medical criteria. As a rule, the testing of hypotheses in diagnostics is not performed on all possible diseases, but is applied to the most serious losses.

c). Types of Decision Making

Two types of decision-making have been generally adopted, and their usefulness has been verified in many fields. One is called Bayesian, and the other method derives from Wald. Both of them account the risk function

into consideration. The formal difference is that the Bayes' principle uses a priori probability of the hypothesis, but Wald's principle does not take this into account.

There is a difficult problem of epistemology here: whether to accept the a priori probability or not. In the case of algorithmic inference we do not use a priori probability, but instead we proceed to a conclusion according to the branching of logic, no matter how rare its incidence may be. With the method of probabilistic inference on the basis of the likelihood, the structure of logic is somewhat embellished, and the validity of the conclusion is expressed by the magnitude of the likelihood. Nevertheless, the fundamental structure in epistemology is nearly the same. Wald's principle belongs with this epistemological viewpoint.

Fron another standpoint, the Bayesian principle chooses the decision which can minimize the Bayesian risk function through all possible hypotheses. Its consideration is global, so to speak, and the loss must be numerically evaluated and averaged. Contrary to this thinking, Wald's principle involves choosing the decision that minimizes the maximum loss that could occur in the worst possible case. It is called the minimax principle. After the disease with the maximum loss becomes known, this principle is applied locally, neglecting other hypotheses.

d). Strategy in Medical Practice

By detailed analysis of the process of decision-making in medical practice, we suppose that Wald's minimax principle is primarily needed here. This may be gathered from the uniqueness of diseases under treatment, and in the quality of loss connected with misdiagnosis.

Sometimes we can observe patients under our control, when the disease profile does not show a decisive sign or symptom. This principle is valid only when the possible delay does not cause much loss. Using the same thinking process, doctors may try provisional treatment, and change their decision according to the response of their patients.

In the offices of general practitioners, a priori probability may be widely used, on the basis of the fact that most diseases are cured spontaneously, and that rare or severe diseases are screened out under the doctor's control by their courses. If the emergency cases are eliminated from a particular situation, a delay of correct diagnosis for several days will not induce so large a loss. However, even in the ambulance, if the disease profile is highly specific for a certain disease, we do not balance the likelihood by a small _a priori_ probability.

Sometimes we cannot express _a priori_ probability numerically, but it seems to be enough to know that a disease is possible at the time. It will be considered when a set of hypothetical diseases is made.

Diseases which were excluded from a initial set of hypothetical diseases because of their low incidences, will be considered during the course of observation, when the patient shows some queer symptoms which contradict to the present diagnosis.

10. FUTURE PROBLEMS IN COMPUTER DIAGNOSIS

Medical practice has many aspects which should be studied in the light of the theory of scientific data processing. Men of medicine must deepen their insights on the thinking processes involved in the practice of their profession. They must improve their thinking, and make it more effective, through co-operation with the mathematicians.

At times accurate computation processes are very complicated, and at times they do not appear to have a practical application to medical problems of diagnosis. Nevertheless, we have been impressed throughout the course of this research with the truth that the theory of inference may be fairly robust, in application to changing assumptions to a small extent, or in approximation, by making it possible to dispense with complicated computation terms. From a practical viewpoint, the robustness of the theories in this field should be investigated more thoroughly.

The second major problem in this type of research is to obtain a large number of cases which can be minutely examined according to a definite schema, and confirmed in diagnosis. We have already coded a thousand cases of brain tumor, yet when we calssify them according to type and location, they are insufficient in number. The most urgent matter for reconstruction of medical diagnostics into a system of scientific inference is the establishment of an organization to promote a plan to collect the required data according to a definite schema, on a world-wide scale.

There is a lot of misunderstanding on computer diagnosis: it is a profanation of human being to have the computer think. It is a destruction of human relation between a doctor and a patient. It is nothing but pulling down the doctor's diagnostic ability and so on. Computer is doctors' tool, which should make doctors' diagnosis stable and accurate. Scientific inference in medical diagnosis assisted by computer does not mean to exclude doctors as a psychologist to the patient. We hope that wisdom of human being may raise up the efficiency of medical diagnosis by using computers, overcoming many difficulties. It also depends on our wisdom whether we have computers work as God or as Gohlem.

F.1 **Algorithmic Diagnosis of Congenital Heart Diseases**

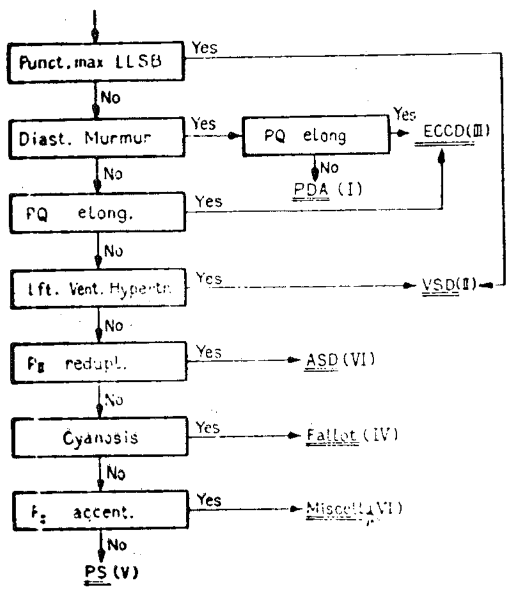
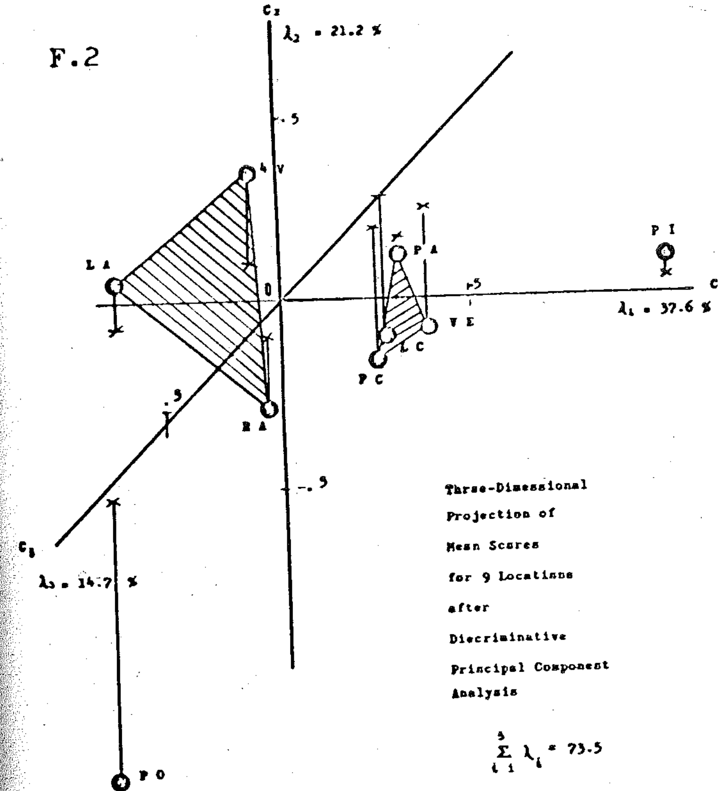


Figure 1

F.2



Three-Dimensional Projection of Mean Scores for 9 Locations after Discriminative Principal Component Analysis

$$\sum_{i=1}^{3} \lambda_i = 73.5$$

Figure 2

Distribution of Scores

Score	PDA	PDA +PH	ASD	ASD +PS	VSD	VSD +PH
1	6			5	1	
2			1	4	1	
3	12		3	2	3	
4			4	0	5	
5	9		3	1	3	1
6			6		5	1
7	1		5		3	2
8			8		1	
9		3	3			
10					1	4
11		3				4
12						4
13						3
14						2
15						1

F.3

Experiment on Card Diagnosis by Doctors

Material : Tumor Near Cerebellum (9 Locations) Information : Clinical Neurological Data (29 Symptons)

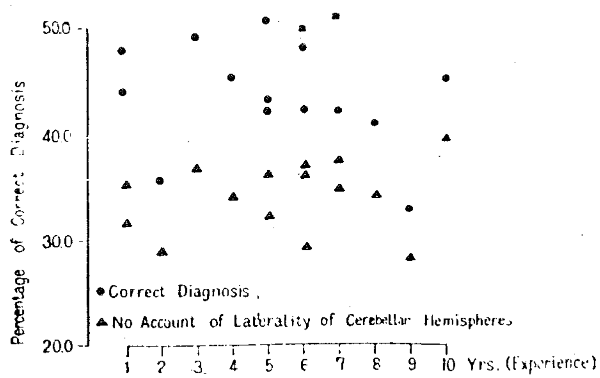


Figure 3

T-1

FACTOR LOADs obtained by FACTOR ANALYSIS

- on log P_{ia} -

VARIATES \ FACTORS	F_I	F_{II}	F_{III}	F_{IV}	F_V	F_{VI}
Clubbing digit	-.51	-.52				-.4[illegible]
Squatting	-.[illegible]		Hypoxaemia-Polyglobulia			
Cyanosis	-.92					
Red Blood Corpuscles		-.97				
Axis deviation to lft			-.69			
PQ interval elong.			-.90	Overloading of r. Auricle		
Mitral (notched) P			-.65			
QRS Breadth short.			.83			
Sinus Tachycardia				.90		
Chest Deformity		Pulm. Hypertension		.92		
P_{II} accentuated				.78		
Respir. Infect. freq.				.68		
PUNCT max in LLS3				.41		.46
PCIR-area continuous or diastolic		Pattern of Murmurs				
LLS-area ditto						
APEX-area diastolic				.51		
P_{II} reduplicated			-.74			
$ST_{I,II,III}$ elevated			-.51			
Axis Deviat. to rt	-.57					-.57
R' or qR in V_1	-.47	Overloading of r. Ventricle			-.53	-.47
R/S >1 in V_1					-.49	-.77
$ST_{v5,v6}$ elevated	-.67			-.46		-.55
RS or RS in $V_{5,6}$						-.90
QRS Breadth enlarg.			-.41	-.66		-.41
Fatigue						-.87
$S_{V1} + R_{V5} \geq 4mV$		.45				.64

Table 1

T-6 APPROXIMATION OF JOINT PROBABILITY IN NORMAL MULTIVARIATES --With Various Assumptions--

ASSUMPTION	CONSTANT	VARIANCE	DISTACE
Normality of Distr. $\log L_a$	$\frac{k}{2}\log 2\pi$	$\frac{1}{2}\log V_a$	$\frac{1}{2}\sum\sum\lambda_{ija}(X_i - m_{ia})(X_j - m_{ja})$
Normality of Distr. Homogeneity of Var. (throughout all diseases)	$\frac{k}{2}\log 2\pi$	$-\frac{1}{2}\log V$	$\frac{1}{2}\sum\sum\lambda_{ij}X_iX_j - \sum_i\{\sum_j\lambda_{ij}m_{ja}\}X_i + \frac{1}{2}\sum\sum\lambda_{ij}m_{ia}m_{ja}$
Normality of Distr. Independ. of Variates	$\frac{k}{2}\log 2\pi$	$\frac{1}{2}\sum\log\sigma_{ia}^2$	$\frac{1}{2}\sum\frac{X_i^2}{\sigma_{ia}^2} - \sum\frac{m_{ia}}{\sigma_{ia}^2}X_i + \frac{1}{2}\sum\frac{m_{ia}^2}{\sigma_{ia}^2}$
Normality of Distr. Homogeneity of Var. (throughout all diseases) Independ. of Variates	$\frac{k}{2}\log 2\pi$	$+\frac{1}{2}\log\sigma_i^2$	$\frac{1}{2}\sum\frac{X_i^2}{\sigma_i^2} - \sum\frac{m_{ia}}{\sigma_i^2}X_i + \frac{1}{2}\sum\frac{m_{ia}^2}{\sigma_i^2}$

Table 6

CYBERNETICS IN MEDICINE. SOME STUDIES ON THE USE OF LOGICAL ANALYSIS IN MEDICAL DIAGNOSIS

By Lee B. Lusted, M.D., University of Oregon Medical School and Oregon Regional Primate Research Center.

1. INTRODUCTION

As medical data has increased in volume the physician has had difficulty in using the data effectively. One explanation seems to be that the large volume of data makes it a formidable task for the physician to remember the interrelationships among signs, symptoms and laboratory tests for a wide variety of diseases. Thus it seems that studies need to be undertaken which will attempt to analyse and synthesize medical data in order to permit a reorganization of the data in a more meaningful, more manageable and more useful form. The logical analysis procedures which I wish to discuss are presented with this objective in mind. Some of the procedures were developed quite a few years ago and have recently been reinvestigated with renewed interest because of the availability of computers. In this paper I have attempted to summarize and to place in new perspective some studies which I hope will stimulate further research on techniques that will make medical data more manageable for the physician in his efforts to maintain health and treat disease in his patients.

Research projects on logical analysis in medical diagnosis seem to be grouped in two main areas neither of which has a very distinct boundary. One area of research might be called studies on the diagnostic process

---

NB Acknowledgement:
I appreciate discussions with Professor B. de Finetti and Professor Lyle D. Calvin concerning some of the material presented in this arricle.

Summary at page 511
Appendix at page 512 - 516
References at page 517 - 520
Figures at pages 521 - 524

and a second area has been referred to frequently as computer-aided medical diagnosis. An interesting group of papers which discuss the diagnostic process will be found in the Proceedings of a conference on The Diagnostic Process (1) and in a series of articles by Engle (2) (3).

2. STUDIES ON THE DIAGNOSTIC PROCESS

The clinician in his role as a decision maker has become a subject for study by physicians who are interested in improving medical school teaching methods and by psychologists concerned with problem solving and judgement (4) (5). The interest in the decision maker is his role as a medical teacher was expressed recently by Adams (6) at a Conference on the Diagnostic Process when he noted that "the questions the teachers asked the students may represent the kinds of questions that he asks himself when he solves problems alone. This, in fact, may be a possible way into the problem solving process of the teacher. We will have to come to grips with the question of what goes on in the teacher's mind. One of the big questions of the symposium seems to be "What can we learn about the processes that actually take place in the mind of the person solving problems of medical diagnosis through the study of computer processes or by other means?"

A particular approach which uses information processing computer languages to simulate the clinical decision-making processes has been selected for further discussion. This information processing approach is one way of looking at psychological activity and the use of information processing computer languages encourages the investigator to specify precisely and explicitly the systems of cognitive structures he induces from the behavior he observes (8). Furthermore, this type of study deals with processes and functions and it is concerned with the fine structure of behavior.

The concept of simulating human behavior with a computer program may be unfamiliar to some readers. Newell and Simon (9) have provided considerable evidence that an electronic computer can carry out complex patterns of processes that parallel closely the processes observable in humans who are thinking. These investigators collect data from subjects who are asked to "think aloud" during problem solving sessions. The subjects' comments are tape recorded and they then try to write computer programs

that will simulate the behaviors observed in their data. They view a program capable of simulating behavior as a theory of the system of psychological processes and structures underlying the behavior. Such theories are held to have a status comparable to those framed in words or in methematical symbols and to be subject to the same theory of adequacy (8) The significance of adopting this approach to the study of clinical decision-making is that it forces us to specify with complete rigor the processes which are involved.

To illustrate how the "think aloud" and information processing computer programs are used in the study of medical diagnosis, I have selected an example from the work of Kleinmuntz (12) on clinical decision-making in neurology. The technique used is a diagnostic game played somwhat in the form of "Twenty Questions" with one player, called the experimenter, thinking of a disease and a second player, called the subject, trying to diagnose the disease the first experimenter has in mind. The experimenter must be an expert neurologist if the game is to be meaningful because in the various roles he may assume he must be able to recognize the appropri ateness and validity of the questions which the subject asks him in trying to make a diagnosis. The subject's questions and the experimenter's answers are tape recorded for further analysis and for use in developing a computer program. An end product of the diagnostic game may be picture as a tree structure which represents the search strategy which the subject used in arriving at the diagnosis.

For a particular diagnostic game a decision tree structure might result as shown in figure 1. The experimenter asserts he has a disease in mind (shown as D1) and the subject asks whether a certain symptom S (5) is present. He receives a negative reply. Then the subject asks whether symptom S (10) is present and again receives a negative reply. Based on information he has so far he asks whether symptom S (56) is present and affirmative replies lead him to ask about S (81) and S (12). He then ask for evidence in terms of laboratory tests T (3) and T (15). The experimenter replies that T (3) results are negative and T (15) yields positi results. On the basis of the cumulative information the subject suggest diagnosis D (1), the correct diagnosis, and the game is concluded. If D (1) had been incorrect the game would have continued.

The circles in the tree represent decision nodes and the number of node which designate decision levels plus the number of tree branches may be considered as a rather crude model of the fine structure of the decision process.

The number of branches in a particular decision tree semms to be a function of the experience and ingenuity of the diagnostician. For the same diagnostic problem one diagnostician may use a decision tree with four branches wheras another diagnostician may use forty. By varying the amount of initial information supplied to the subject and, by controlling the kinds and amount of information available during the game it should be possible to make a systematic study of diagnostic decision trees and thereby to determine optimum decision tree structure and to evaluate the effects of various signs, symptoms, etc. on diagnostic accuracy. The use of one type of diagnostic decision tree has been described for a programmed computer system to be used in teaching medical diagnosis. The decision tree developed by Swets and Feurzeig (13) illustrates the manner in which a student can be led through a differential diagnosis, to a diagnosis which in the example was pneumococcal pneumonia.

3. OBSERVER VARIATION

The investigator who has started research on the use of logical analysis in medical diagnosis must sooner or later consider how his data are influenced by observer variation and the accuracy of diagnostic procedures. This statement applies whether the original information on incidence of signs and symptoms comes from "personal" probability estimates, textbooks or medical records. Several hundred articles (14) have been published on the subject of observer variation. Although many investigators already are familiar with some of the work in this field a short discussion of the topic will be presented to act as a bridge between the subjects of studies on the diagnostic process and computer-aided medical diagnosis.

Suppose that a physician has a group of 100 patients from whom he has collected an extensive amount of information on signs, symptoms, laboratory data, etc. and after a study of this information he reaches a judgement on each patient concerning whether the patient is ill or not-ill. We allow the physician only one of two possible choices for each patient. Now suppose we present this same group of patients to an additional nine

physicians of equal competence. There will be some difference of opinion among the ten physicians concerning the number of ill and not-ill patients in the group. We suppose also that a proven diagnosis is available on each patent.

The judgment of the ten physicians can be divided into the four categories true positive (patient truly ill), true negative (patient truly not-ill), false positive (not-ill patient judged to be ill) and false negative (ill patient judged to be not-ill). The patients who are found to be in the false negative (Type 1 error) and false positive (Type II error) categories represent variation in physician judgment and this group, therefore, require further study to determine the possible sources of variation (15).

It is important for the physician to understand that the demarcation line between the ill and not-ill population will vary depending on the criteria (signs and symptoms) he uses to make his decision. No matter what signs, symptoms or tests he uses to define an ill patient he will find these same signs and symptoms occurring in some not-ill patients. Part of the uncertainty in diagnosis results from this overlap of the populations of truly ill patients and truly not-ill patients (16). Collen (17) has shown how this uncertainty is treated in the screening procedure for asthma by making use of the likelihood ratio.

When signs and symptoms are chosen to define an ill patient the physician has established two things. First, he has set a demarcation line between ill and not-ill populations. Second, he has established a ratio of false negative to false positive cases. This is illustrated in figures 2 and 3 in which a population has been divided into two sub-populations (ill and not-ill) on the basis of a set of criteria chosen (in this case arbitrarily) to define "ill".

False negatives and false positives are errors which should be recognized as a "cost" to a system and it is important to find a combination of false positives and false negatives which represent a minimum total cost. The "cost" may actually represent dollars in the case of false positives where follow-up studies are required. The "cost" of the false negative case to the individual and to society is perhaps harder to evaluate and requires further study.

Numerous experimental studies (19) on the accuracy of cytological diag- nosis of Papanicolaou smears and on the interpretation of chest roent- genograms have shown that a reciprocal relationship exists between the number of false negatives and the number of false positives. This inverse relationship which is demonstrated in figure 4 for the reading of chest x-rays (20) shows that as the number of false negatives decreases the number of false positives increases. As the criteria (signs and symptoms) are made more stringent for diagnosing the ill patients, the number of false negatives will decrease but at the expense of increasing the false positives.

From considerations of statistical decision theory it can be shown that a family of error probability curves may be developed which are similar in shape to the curve shown in figure 4. Within this family each curve represents the results obtained with a particular set of diagnostic cri- teria. One curve will represent a lower, left boundary and along this curve there is no simple and categorical basis for choosing the "better" of two points of operation (21) (equivalent to choosing a "better set of diagnostic criteria) since each of the two points is better with respect to one of its error probabilities (that is, better with respect to either the false negatives or the false positives).

In a study of congenital heart disease diagnosis Toronto, Veasy and War- ner (22) were able to demonstrate that participating clinicians made two types of errors: namely, errors in symptom recognition and errors in logic (coming to a wrong conclusion from valid available data), and that the two types of errors could be investigated separately. The physi- cian working on computer aided diagnosis, needs to have some appreciation of how observer bias effects his original probability estimates and his subsequent observations on signs and symptoms. As Grosz and Grossman (23) have pointed out there is a much greater degree of observer variation in clinical judgment when data are subjective and ambiguous than when the data are objective and factual.

To conclude the discussion on observer variation two points can be made which emphasize the contribution that observer variation studies can make to analysis and synthesis of medical data. First, the studies can help to indicate which signs, symptoms and laboratory tests are really most useful in terms or repeatability and validity (16) for making a diagnosis

The real benefit, however, can come from studies which show the effect of different diagnostic criteria on the number of false negatives and false positive diagnoses. The work cited earlier of Frieden, Shapiro and Feinstein (18) on the radiologic evaluation of heart size in rheumatic heart disease is an example of such a study. Signs and symptoms that seem to have great diagnostic importance may not be as important as we think and we need not continue to carry the symptoms which are low in diagnostic value. The second thing that observer variation studies can do is to help us to formulate odds concerning how sure we are about a sign or symptom with respect to a diagnosis. How these odds may be used in Bayesian inference will be discussed in the next section on computer- aided diagnosis.

4. COMPUTER-AIDED DIAGNOSIS

Medical diagnosis as a logical process by which the physician follows a chain of events from cause to effect and, thereby, reaches a decision has stimulated the interest of investigators for a long time. For many areas of medicine and perhaps well over half, the chain of events is short and causal relationships are obvious enough so that mathematical models of diagnosis and computer techniques will never be needed. However computer diagnosis studies already have demonstrated the statistical significance of various clinical findings for several areas of diagnosis. For these areas diagnosis has become less cumbersome because the most important diagnostic signs have been separated from signs of less impor- tance. At present the entire diagnostic process cannot be described in terms of a computer program but some parts of the process can be expres- sed in terms of mathematical models. Some of these models will be discus- sed in the material which follows and a selected list of articles (Appendix A) will give the reader an impression about the mathematical models which have been applied to a group of medical specialities.

5. BAYES' THEOREM AND CONDITIONAL PROBABILITY

The eminent English physician, Sir. George W. Pickering (25) wrote re- cently, "Diagnosis is a matter of probability, as those of us who follow the fate of our patients to post-mortem room know only too well. Prognos- is is a matter of probability, and in judging treatment we have to base our judgment on knowledge of probabilities." Only in the past five years have studies in a number of medical specialties indicated the usefulness of conditional probability models to diagnosis.

Conditional probabilities are the probabilistic expression of learning from experience and a particular form of conditional probability known as Bayes' theorem is a useful mathematical device that combines evidence from data with prior information. Bayes' theorem (26, 27) has been used extensively in experimental studies of medical diagnosis with favorable results but its application to diagnosis has been disputed on the grounds that conditions under which the theorem is meaningful are rarely met and that sources of the probabilities often are vague. Bayesian statistics is based on a definition of probability as a particular measure of opinions of ideally consistent people. Statistical inference is a modification of these opinions in the light of evidence, and Bayes' theorem specifies how such modifications should be made. On the other hand Bayes' theorem is universally accepted as mathematically correct. Some of the problems which arise in the application of Bayes' theorem will be discussed later. An excellent general discussion of the use of Bayes' theorem in medical diagnosis has been by Cornfield (28, 29). Edwards, et al (30) have published an interesting report on the use of Bayesian statistical inference in experimental psychology. These articles and the work by Mosteller and Wallace (31, 32) on a problem of disputed authorship of twelve Federalist papers using Bayesian analysis will be helpful to the investigator who is interested in the use of Bayes' theorem in diagnosis.

Bayes' theorem applied to diagnosis may be written:

$$(1) \qquad P(D \mid S) = \frac{P(D)\ P(S \mid D)}{\sum_j P(D_j)\ P(S \mid D_j)}$$

This expresses the conditional probability of disease D, given symptom complex S as a function of the unconditional probability of disease D and the conditional probability of symptom complex S, given disease D.

An expandable form of Bayes' theorem has been used by Warner et al (33) for the diagnosis of congenital heart disease.

Warner and colleagues summarize their work with Bayes' theorem as follows: "Experience with this approach in the field of congenital heart disease

indicates that these diseases can be diagnosed with an accuracy equal to that of an experienced specialist in this feld.' Bruce (34), Lodwick (35) and Overall (36) have used similar computational forms of Bayes' theorem in the diagnosis of acquired heart disease, bone tumors and thyroid disease. All of these models assume that the symptoms are independent of one another within a given disease and that the diseases are mutually exclusive. Independence of symptoms can be tested by chi square analysis if sufficient data are available on the coincidence of symptoms in each disease.

The condition of symptom indepen-dence has not always been met and yet the Bayesian analysis has worked quite well. This seems to bear out Cornfield's statement (28) - "there is no assumption --- that the components of the multi-dimensional event are independent. We stress this obvious point only because there seems to be a widespread misconception in the medical literature that the application of Bayes' theorem to the diagnosis problem requires the assumption of independence of the various symptoms."

The assumption that diseases are mutually exclusive has been handled by Warner by noting that if a patient is found to have more than one disease the new combination of diseases is considered as a new single disease entity. This point needs further study.

However, the most controversial issue concerns the assignment of the unconditional or prior probabilities and on this point a distinction between frequency and non-frequency interpretation of probability must be carefully considered. Probability may be considered a degree of belief or as relative frequency and the two interpretations are both used in the application of Bayes' theorem to medical diagnosis. As Mosteller (32) has pointed out the degree-of-belief interpretation is more widely applicable, but specifying the probabilities is often difficult in the absence of conditions of symmetry or long-run relative frequencies. When the two interpretations are simultaneously applicable, the same numerical values would normally be assigned in each interpretation.

The prior probability P(D) may be constructed from a variety of sources such as medical textbook, patient medical records or a subjective probability (30) from the judgment of an experienced clinician. Since the sources and prior distributions of such data are not well documented the correctness of P(D) may be questioned. Of course the best available data or estimates should be used to determine conditional probabilities and prior probabilities and when this has been done the investigator may derive some satisfaction from a methodological conclusion of Mosteller and Wallace (31): "Prior distributions are not of major importance. While choice of underlying constants (choice of prior distributions) matters, it doesn't matter very much, once one is in the neighborhood of a distribution suggested by a fair body of data. We conclude from this that the emphasis on the difficulty, even impossiblility of choosing prior distributions as a criticism of the use of Bayes' theorem is not well placed."

Edwards (30) has noted that subjects participating in decision making studies very often are unwilling to change their diffuse initial opinions P(D) and P(S|D) into sharp posterior ones P(D|S), even after exposure to overwhelming evidence. Whether this reluctance to extract from data as much certainty as they permit occurs in medical diagnosis needs to be investigated. With the data collection organized it is possible to expand the data by using a computer program which adds each new case as it is diagnosed. Such a computer program is now used by Williams (36) for the diagnosis of thyroid disease.

The experience with Bayes' theorem in medical diagnosis during the past five years can be summarized by noting that for certain diseases new insight has been gained into the relative contribution a symtom or symptom complex will make to a diagnosis. For instance in the clinical and radiologic diagnosis of being gastric ulcer versus malignant gastric ulcer it was shown by Wilson et al (38) that seventeen statistically valid signs and symptoms would make the diagnosis with a high degree of certainty. These seventeen variables were selected from a group of 70 signs and symptoms listed in medical textbooks as significant variables in the diagnosis of gastric ulcer.

Cornfield (29) has shown recently that diagnostic procedures based upon either likelihood ratios or posterior probabilities are closely related rather than as sharply opposed as has been suggested in the past. Because of the importance of likelihood ratios a discussion is given in the following material.

6. LIKELIHOOD RATIO

A consideration of Bayes' theorem will be used to develop the likelihood principle and the likelihood ratio. Suppose that two possible observations on a symptom s_1 and s_1' - not necessarily from the same experiment - can have the same bearing on your opinion about a diagnosis d_j. Another way to state this is that $P(s_1|d_j)$ can equal $P(s_1'|d_j)$ for each S_1. The question may be asked, when are s_1 and s_1' equivalent or of the same importance with respect to your opinion about the diagnosis d_j.

The conclusion concerning the question above is given in terms of the likelihood principle which says that s_1 and s_1' are of the same import if the following equation is valid.

$$P(s_1 | d_j) = KP(s_1' | d_j) \tag{4}$$

For the purpose of making a diagnosis this equation says that the sequence of numbers $P(s_1 | d_j)$ is, according to the likelihood principle, equivalent to any other sequence obtained from it by multiplication by a positive constant (K). We will see shortly how Collen (17) has used this principle as an aid in diagnosis.

Equation 4 may be rewritten in the form

$$K = \frac{P(s_1 | d_j)}{P(s_1' | d_j)} \tag{5}$$

This equation, known as a likelihood ratio, in an application of the likelihood principle and it has been used as an aid to diagnosis in a multiphasic screening program by Collen (17) as follows.

A given population is to be screened for the presence of certain stated diseases. The only data needed for the likelihood ratio method are the determinations for each disease (d_j, $d_2 \ldots d_j$) to be screened for the proportion of individuals with a disease d with certain sets of symptoms (s) and the proportion of individuals not having the disease d_j but with the same set of symptoms (s).

The statistical procedure used to implement the likelihood ratio method of screening was suggested by Neyman (39) in 1950 and is based upon selecting sets of symptoms whose relative frequency in diseased and non-diseased patients will define a region having a maximum likelihood of detecting the disease, if present, and at the same time, a predetermined likelihood of indicating that a disease is present when actually it is absent. The ratio of the frequency of a set of selected symptoms (S) in the diseased patients (D) to the frequency of this same set of symptoms (S) in the non-diseased patients (N) when written in the form of the likelihood ratio defined in thepreceding equation, and using the notation of Collen is

$$(6) \qquad \theta = \frac{P(S|D)}{P(S|N)}$$

This method offers a simplification in the theory of testing because it does not require that the prevalence of a disease in the population be known and it is possible, therefore, to use this method at the outset of a diagnosis study when a large amount of data is not yet available. The method may also be used in the presence of changing incidence of disease, in various geographical areas where disease prevalence may differ and under circumstances where it is not necessary to prove symptom independence., From the mathematical point of view the question of mutually exclusive diseases does not need to be considered when using the likelihood ratio but from a medical point of view the question cannot be ignored. Likelihood ratio studies may be used to determine the sensitivity and specificity of a diagnostic test and to show observer variation, a subject which was discussed earlier in this paper.

6. OTHER STATISTICAL METHODS

Overall has made a comparison of alternative statistical models and clinical diagnosis in the diagnosis of thyroid function. The computer-based statistical models used were a Bayesian conditional probability approach (36), factor analysis (40) and discriminant function analysis (41). With the Bayes' model the investigators were able to classify patients in agreement with clinical diagnosis more than 97 per cent of the time and the factor analysis results agreed with clinical opinion in over 95 per cent of the cases. An additional group of studies best categorized as

matching procedures which are related to multivariate analysis are listed in the appendix.

7. SUMMARY

In this article I have attempted to draw together some examples of research activity which represent various approaches to a subject I have called logical analysis in medical diagnosis. Studies on computer-aided diagnosis and on the diagnostic process complement each other and share a common interest in problems of observer bias. The statistical information gleaned from research on computer-aided diagnosis can supply data to be used in the diagnostic process studies and these studies in turn may supply new methods for testing and utilizing the statistically based diagnostic algorithms.