

Health Data Banks

Report on a Joint IFIP/WHO Working Group

Prague

23-27 August 1976



Not for Sale
Distributed by the
REGIONAL OFFICE FOR EUROPE
World Health Organization
COPENHAGEN
1977

ICP/DHS 002

Note

This report has been prepared by the Regional Office for Europe of the World Health Organization for governments of Member States in the Region and for those who participated in the Joint IFIP/WHO Working Group on Health Data Banks. A limited number of copies are available to persons officially or professionally concerned in this field of study from the WHO Regional Office for Europe, Scherfigsvej 8, 2100 Copenhagen Ø, Denmark.⁴

The views expressed are those of the participants in the Working Group and do not necessarily represent the decisions or the stated policy of the World Health Organization.

The designations employed and the presentation of the material in this report do not imply the expression of any opinion whatsoever on the part of the Secretariat of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Where the designation "country or area" appears in the heading of tables, it covers countries, territories, cities or areas.

This report is also available in French and Russian.

CONTENTS

	Page
1. Introduction	5
2. The role of health data banks — general considerations	6
2.1 Introduction	6
2.2 The uses of health data	7
2.3 Problems in data bank establishment and operation	8
2.4 Conclusion	9
3. Types of health data bank	9
3.1 Hospital administrative data banks	9
3.2 Community-oriented data banks in health care	10
3.3 Data banks on absence through ill-health	11
3.4 Administrative data banks on ambulatory patients	12
3.5 Data banks in general practice	13
3.6 Child health status data banks in the United Kingdom	14
3.7 Disease registers	15
3.8 Health manpower data banks	16
3.9 Other health data banks	17
4. Privacy and confidentiality	21
5. Communication with users	22
6. Organization and techniques of input into data banks	23
7. Data bank concepts and considerations on hardware	24

	Page
8. Training	25
9. Exchange of information	25
10. Terminology	26
11. Conclusions and recommendations	26
Annex List of participants	30

1. INTRODUCTION

A Joint IFIP/WHO Working Group on Health Data Banks met in Prague from 23 to 27 August 1976 under the auspices of the World Health Organization Regional Office for Europe, and in collaboration with the Government of Czechoslovakia. The members of the Working Group were welcomed by Professor M. Kusak, Deputy Minister of Health of the Czech Socialist Republic, who stressed the need for health administrators to have information readily available for assistance in their decision-making, noting that if this was to achieve its maximum effectiveness, there would be a need to integrate health statistics with other social statistics.

Opening the meeting on behalf of the Director of the WHO Regional Office for Europe, Dr Leo A. Kaprio, Dr D.K. Sokolov, Chief, Strengthening of Health Services at the Regional Office, said he attached great importance to the growth of scientifically based planning of health services. Mentioning various meetings organized by WHO to further this development, he referred in particular to the Conference on Health Information Systems held in Copenhagen in 1973 which had called for the further development of multi-entry, multi-purpose data banks as well as systems for the collection, aggregation, storage and retrieval of problem-oriented data according to users' needs.⁴

The Working Group was one of the follow-up activities recommended by that Conference. It was a multidisciplinary group of clinicians, epidemiologists, public health doctors, computer scientists, statisticians and administrators. The objectives of the Working Group were to discuss the planning and development of health data banks and the linkage of information from various health programmes; to examine the relationship between the organizing data banks and their users, and means of identifying users' needs; and to consider the problems of privacy and confidentiality. On behalf of the International Federation for Information Processing (IFIP), Dr J. Roukens also welcomed the members of the Working Group and stressed the need for health decisions to be based on factual evidence. He hoped that the Working Group's deliberations would be of assistance in meeting this objective. Professor J. Červenka was elected Chairman of the Conference, Dr J. Roukens, Co-Chairman, and Dr M. Heasman, Rapporteur. Dr Z. Brzezinski acted as Secretary and Dr J. van Egmond as Co-Secretary. A full list of participants is attached as an Annex.

The papers presented to the Working Group are being published separately by IFIP as part of the proceedings of the meeting.

⁴ WHO Regional Office for Europe. *Health information systems: report on a conference. Copenhagen, 1973 (EURO 4914)*

2. THE ROLE OF HEALTH DATA BANKS – GENERAL CONSIDERATIONS

2.1 Introduction

In 1973, the Conference on Health Information Systems,^a held in Copenhagen, defined a data bank as "a system for accepting or storing data concerning defined groups of persons, events or institutions in such a way that the information may be aggregated according to the users' needs." In this system, the most important element is the concept of aggregation, or re-aggregation to suit different needs, some of which may be unknown when the data bank is established.

It should be made clear that not all health data banks are concerned with people and patients. In addition to institutions mentioned above they may also, for example, be concerned with environmental factors, drugs, etc.

Health data banks can be of value in all aspects of health service work from the planning and management of health services, through epidemiological and other research to care of the individual member of the population whether or not that individual is a patient. A health data bank is generally considered to be only part of a health information system. The information system consists not only of data banks but also of interpretive and advisory services and also has a capacity to undertake *ad hoc* research.

In general, the purpose of a data bank is to provide a basis for decision-making. Decisions may be taken by health personnel, administrators, research workers or other responsible persons inside or outside the health services.

Possible classifications of the statistical uses of data banks are many and varied. They are of value in both administration and research, and it is very often difficult to distinguish clearly between the two uses. Some of the broad statistical uses of health data banks are:

- (a) simplifying complex problems by providing objective data for use in decision-making;
- (b) predicting probable consequences of different actions and measuring the effectiveness of intervention;
- (c) discovering ways of improving performance in health services and assessing the skills necessary;
- (d) comparing results in different groups or against specified objectives;
- (e) assessing the resources required to maximize benefit at minimum cost.

^a WHO Regional Office for Europe, *Health information systems: report on a conference*. Copenhagen, 1973 (EURO 4914) p. 34

Again, data can sometimes be used for:

- (1) describing distribution of normal values (depending on the source of the data);
- (2) finding cause-and-effect relationships;
- (3) determining factors of disease incidence and prevalence;
- (4) assessment of risks;
- (5) monitoring performance, such as improvement in vaccination rates, mortality, morbidity, etc.

Some statistical applications of data banks can be described as "data-driven", e.g., the data are used for monitoring a particular trend; in other instances the application can be described as "theory-driven", i.e., the data are used in order to establish and test a particular hypothesis. Another function of a data bank includes its use as a reference to records which can then be accessed for further and more detailed study.

There was general agreement that data in medical data banks are often of great value in the initial stages of epidemiological study. Such uses cannot be foreseen in detail at the initial design stages of the system. Whilst a good knowledge of the validity of the data is essential, its use in this way is one of the most valuable, even though much of it is not susceptible to testing by normal statistical methods.

2.2 The uses of health data

The uses of routinely collected health data can be classified as primary, secondary and tertiary, although such a subdivision is not entirely mutually exclusive.

The primary uses of health data are those for which the data is immediately collected. Examples of this are registration of births and deaths, an act undertaken primarily for legal purposes; information collected to enable a patient to enter hospital and be treated; documentation surrounding immunization procedures, etc. In administration, payroll can be cited as an example of primary data, as can stock records, or records of institutional registration.

Secondary uses involve re-aggregations of data, still in the local setting, to facilitate, for example, patient, institutional, or health service administration. Examples are the use of details of patient registration to provide bed-state data, or the use of birth registration to arrange immunization schedules, the use of stock records to set up and maintain stock control and ordering procedures and so on.

The tertiary uses of the data are, in the main, statistical ones, for example, the use of vital registration information to produce birth, fertility and mortality statistics. Hospital inpatient statistics can be derived from patient

admission and discharge records; payroll and personnel records can be used as a basis of manpower statistics and for manpower forecasting.

A general axiom, agreed during the meeting, was that, wherever possible, data should be collected only once, and used for as many purposes as possible. In general, data collected for a primary purpose and used for secondary and tertiary purposes were more likely to be accurate than data collected only for tertiary purposes. However, some reservations were entered by the Working Group in connexion with the multiple uses of data. These stated that care has to be exercised to ensure that the accuracy and validity of the data is understood in connexion with the purpose for which it is to be used.

2.3 Problems in data bank establishment and operation

In setting up a data bank, the objectives are often poorly defined. Better definition comes only after consultation following the production of preliminary results. This means that there must always be considerable flexibility at the design stage of the system.

In many medical data banks, the real cost of data collection is the most expensive part of the system, in some cases amounting to 75-80% of the total cost of data processing. This proportion is reduced when data are collected as part of an on-going service in which they are routinely utilized.

Problems in data bank operation depend to a large extent on the problems in standardization of recording, in classification and definition. It was felt that good definition of terminology used was of more importance in the operation of a data bank than exact comparability of data between data banks.

In considering any of the uses of a data bank, care must be taken that the biases inherent in the collection of data are well understood. If data are collected over a period of years it must be remembered that both the meaning and validity of the data may change with time, a fact which must also be taken into account in interpretation.

There was a consensus that when data are transmitted from one data bank to another they should be in unit form, i.e., they should relate to individual data units rather than be transmitted as aggregates, as this provides far more flexibility for programming, processing and analysis. On the other hand, aggregates may be more suitable for transmission, if computers are not readily available for any further analysis required.

The need to maintain adequate confidentiality of data is considered in more detail later in this report. There was general agreement that data relating to people or patients should be either linked or linkable, although this should not always necessitate the retention of the name of the individual concerned since in many instances, a single registration number would be sufficient. All data banks should, wherever possible, be relatable to a known population at risk.

Although some types of data, for example on vital statistics registration, have a basis common to all developed countries, other types, such as those dealing with manpower or the population served by particular health facilities, may differ considerably according to the organizational pattern of a particular health service. For example, in the United Kingdom and the Netherlands, the need for the population to register with a general practitioner for primary care purposes provides virtually complete and regular population groups for each doctor or group of doctors. Such regular population groups are impossible to obtain in Belgium and Switzerland where patients visit the general practitioner or specialist of their choice.

2.4 Conclusion

Strong emphasis was laid on the need for data banks to be cost/effective and for their objectives to be realistic. A proper degree of realism in health computing is now more prevalent. In setting up a data bank it is often wise to consider what to leave out rather than what to include.

3. TYPES OF HEALTH DATA BANK

3.1 Hospital administrative data banks

Two papers presented to the Working Group dealt specifically with hospital inpatient data banks. The first dealt with the *Gestion administrative et médicale des Malades* (GAMM system) used in the public assistance hospitals in Paris, the second, a more specifically statistical system developed for the *Stichting Medische Registratie* (Foundation for Medical Data Collection) in Utrecht, Netherlands.

The GAMM system is used for hospital management in the preparation of admission records, lists of patients admitted to, being treated in and discharged from hospital. Individual summary patient records are also prepared. For medical services, laboratory results sheets, logs, and registers are prepared, as are reports of X-ray film consumption together with an activity report summarizing the various procedures undertaken on patients, with the aim of obtaining invoices for payment either by the social security bodies or the patient. Bed occupancy rates and other statistical data are prepared for the central administration.

Data are first collected as the patient arrives at the hospital and from these, pre-punched cards are printed as well as a set of self-adhesive labels. The cards are then used to record later examinations, etc., with the data being prepared as close to its course as possible. They can be prepared by

punching through masks or, in some instances, recording in alphanumeric form for later punching. The cards are collected from the ward three times a day and transmitted to the hospital computer terminal. The central file is then updated and the results of some examinations are printed out, together with relevant earlier results. Documents necessary for hospital management, such as bed states, are printed out, while other data are held for future production of invoice statistics, etc. Laboratory tapes are transported by van to the central computer.

Although consideration was given to the use of special-purpose computers within the individual hospitals, detailed study showed that only a large central computer was suited for the task (two IBM 370/145 with 1024 bytes of central core).

In general, few administrative problems were reported in establishing the system and there has been good acceptance from medical and nursing staff, patients and the hospital administration. Data on individual patients are either linked or linkable but the medical data are separated from the patients' identification with files being recognized by number. Further development is expected in the future, including the inclusion of more hospitals and possibly a move towards a real-time system.

This is a patient-oriented system; its one main drawback is considered to be the doubt as to the total size of the population to be covered.

The Netherlands system, on the other hand, which covers 90% of discharges from general and teaching hospitals, is population-based and is used for epidemiological, administrative and planning purposes. Utilization rates are calculated according to the area of residence of the patients and for individual diseases. Similar calculations for specialties are invalidated because of the presence of overlapping areas of treatment, as in general medicine and cardiology. In this system, it is not possible to link separate admissions of the same patient, although it is hoped that such procedures will become possible in time. The system is similar to many others now coming into being at national and regional level throughout Europe.

3.2 Community-oriented data banks in health care

Until quite recently, the concepts of health were largely of medical or biological origin. Lately, however, social factors have increasingly been considered in the concepts of health and illness; there is now a tendency to regard health as "a state of equilibrium between the individual and his environment". This has had the result that social wellbeing is regarded as an important part of a person's physical and mental health.

In order to understand and plan the achievement of such social wellbeing, links have to be built between data recording the health of a population and data relating to social conditions. Such linked data banks may be termed community-oriented. Finland is one country where these have been developed.

Community-oriented data banks have two general areas of application:

- (1) investigation of links between health and social conditions;
- (2) coordination of health and social care and the allocation of resources between health and social policy sectors – this can be for either national or regional purposes, or both.

Community-oriented data banks consist of sets covering data on health, socioeconomic facts, and administrative and economic matters.

Health data, which can be anonymous, should include patients' area of residence and occupation, so that health data can be related to other social data in the same area. The socioeconomic data include data on employment, education, housing, etc., whereas the administrative data set includes data on the economic status of the community, and the resources available for dealing with health and social matters.

A community-oriented data bank can then be used to develop models in order to discover the factors affecting certain planning problems and the relationship between them. For example, such a model has been developed in Finland to estimate future demand for hospital beds. Using such models it might be hoped to optimize the use of health care and social resources.

3.3 Data banks on absence through ill-health

In the same way as absence from work on grounds of sickness represents only a part of total morbidity, it also forms only one part of work total absenteeism. Nevertheless, it is of value both from the point of view of measurement of medical care and from the point of view of work administration, in industrial and other enterprises. Much work in Poland has gone into the development of sickness absence data where, for historical reasons, a large proportion of primary care is provided by factory doctors.

Usually, absence from work through sickness must be confirmed by a doctor. Data on such absence can therefore be used to a varying extent, depending on whether the data cover only one factory or enterprise or whether they cover all persons in a given region or country. Much of their value relies on whether there is a ready availability of appropriate populations at risk; without such a group, meaningful analysis becomes difficult if not impossible.

Data available usually contain personal identification and demographic status and include information on place of work, type of employment conditions together with data on period of sickness absence. The cause of sickness absence is also available but, for reasons of confidentiality, is not usually made known to industrial management. If the data bank is also to serve purposes of current medical care, it should ideally contain data on pre-employment medical evaluation, the results of medical examinations, laboratory tests and treatment records, together with other family illness, etc.

Such data banks have considerable value at local level but in Poland it is considered necessary only to transmit records of absences of at least six days to regional or national level for, although such absences represent only 30% of the number of absences, they accounted for 70% of days lost.

Completeness of the data in sickness absence data banks presents a particular organizational problem. In Poland, where the sickness absence certificates are given by the industrial doctor, it is difficult to record tests and treatment given outside the place of employment. Another difficulty reported is the maintenance of confidentiality in a system which has to serve both medical and industrial management purposes.

In Sweden, a sickness absence data bank exists for the whole population, and is kept for insurance purposes. At the time of its establishment it was decided not to collect diagnostic data although it would probably have been collected if the scheme had been established at the present time.

Although data from data banks for work absence for reasons of ill-health present difficulty in use, it is often cheaper to employ them than to obtain information by setting up specially designed epidemiological studies.

3.4 Administrative data banks on ambulatory patients

Data banks on hospital inpatients, particularly those concerned with statistical uses, are in use in many European countries at national, regional and local level. In recent years, however, there has been a shift of interest to the ambulatory care segment of the health care system. There is a growing recognition of the fact that ambulatory care is absorbing a significant and increasing share of the total resources used for health care and this has created an increased need for data on the provision and utilization of ambulatory care services in the population, both in outpatient departments and in general practice.

As with all other patient-oriented health data banks, ambulatory care data should, wherever possible, be person-specific and population-based, although, as explained elsewhere in this report, the structure of the health service is often a principal factor in deciding if these ideals are attainable. Wherever possible, data should also be provider-specific, not only in terms of classifying the type of health professional who provide care to particular patients, but also in identifying, for example, the particular physician so that variations between doctors can be studied, as can the relationship of doctor characteristics to treatment patterns, etc.

The Group was informed of several types of health data banks existing in Sweden. The first of these related to a data bank compiled from a sample of all contacts between dentists and their patients. The sample is chosen to include only those persons born on specific days of a month, thus allowing linkage of records of visits over a period of time. The data bank is used to study the efficiency of the insurance scheme. In Sweden the data can also be linked with income data to study differences in the utilization of the service.

Some outpatient systems are used to generate data for the financial transactions that take place between the providers of health care and the insurance systems, and such systems, of course, provide a readily available source of data for statistical analysis of outpatient and other ambulatory care work. Some systems are also used for administrative routines and other matters, but there are apparently no systems yet functioning in Sweden where there is total integration of administrative, financial and medical aspects of data banks, although work is under way towards such an objective in Malmo and Stockholm counties.

At Tierp, north of Uppsala, a "population laboratory" has been established for epidemiological and health services research purposes. A comprehensive health information system is being established with extensive record linkage of information from all sources of data. The patient's national identification number is used as the basis for this linkage. A form "travels" with the patient on any visit to the health centre on which the study is based. This acts as a primary data source and enables information to be collected on the reason for the visit, diagnoses, laboratory tests, treatment, etc. Data on visits to private practitioners are collected via the health insurance system. The study has been useful in deriving statistical information to show the proportion of the population using different kinds of health service.

3.5 Data banks in general practice

As mentioned in section 3.4 above, morbidity measurement in primary care is being recognized as forming an increasingly important component of health planning. Data banks on this aspect of practice have the general objective of eliciting the number and type of complaints presented to the general practitioner and the manner of treatment. They are also required to give an insight into the quality of work done, thus making it possible to develop criteria for medical audit, which is seen mainly as a possible application for use in postgraduate education.

Since 1970 the Netherlands Institute of General Medical Practice, in cooperation with the department of the Chief Medical Officer of Health and the Ministry of Health, has been collecting data from over 50 general practitioners, all volunteers, spread throughout the Netherlands in such a way that there is approximately one sentinel-station doctor, as they are called, for about 250 000 population. Data are gathered on weekly reports, and cover items of interest to the public health administrator, to the participating doctors and to those interested in research into a particular topic. Data can be collected for short or long periods or continuously. The only constraints on the type of question that may be asked are that it should be capable of being well defined, uniformly interpreted and gathered in a simple way. Examples of some of the questions asked are numbers of:

- (a) new cases of influenza-like illness;

- (b) cases of tonsillectomy;
- (c) requests for abortion;
- (d) cervical smears;
- (e) prescriptions for tranquillizers;
- (f) sterilizations performed, whether for males or females;
- (g) consultations for hypertension.

In the Netherlands, individuals register for primary care with a particular general practitioner. This means that the data collected can be related to a known population by sex, age, urban area and province, and that extrapolation of the results to the total Dutch population is also possible, provided care is taken over interpretation.

Drawbacks to the system lie chiefly in the fact that the representativeness of the sample of doctors or their patients is always in question, and that the conditions reported always cover only a small proportion of morbidity in general practice.

One of the difficulties of morbidity recording is the lack of a good disease classification for use in general practice. The Working Group noted that the Ninth Revision of the International Classification of Diseases will go some way towards meeting this need but felt that further development of the classification will be necessary before it meets the requirements of general practice. Participants were also told of the classification produced by the World Organisation of National Academies, Colleges and Academic Association of General Practitioners/Family Physicians (WONCA) which has recently published a classification for general practitioners (the WONCA code), and of the list of OXMIS problem codes produced at Oxford, England. The Group felt that no adequate problem-oriented classifications exist at the moment and suggested that WHO should take steps towards their further development, taking into account known experience.

3.6 Child health status data banks in the United Kingdom

In the United Kingdom computer hardware for health administration has been standardized, and this has permitted the development of several standard systems, including one for payroll purposes. The Working Group was told of the development of a "sun/satellite" system for the recording of child health status. The "sun" is represented by a register containing identification data on all children residing in the area from birth to the date of leaving school. So far, three satellite systems have been developed: for immunization, preschool and school health. If feasible, further systems will be developed to cover primary care, inpatient, outpatient and adult screening events.

The immunization package attempts to ensure a high immunization rate. All parents are invited to agree to a course of immunization for their children

and, if consent is given, the computer then sends for the child according to the programme agreed. Completion of the appropriate immunization programme is recorded in the computer, which then arranges, if applicable, for payment to the general practitioner (in accordance with the United Kingdom National Health Service fee-for-service procedure), and also updates the necessary statistics.

The preschool health system schedules attendances of children at specified intervals for examination by specially trained doctors and nurses. The system records the development of the normal child, and provides a register of handicaps which may be highly relevant to the child's future education and welfare. It was noted that a much higher proportion of parents accept preschool health examinations than agree to the immunization system. The system least accepted in the United Kingdom is the one concerned with school health, although it operates in a similar fashion to the preschool health examination system. The reason for this possibly lies in the reluctance of education authorities to change from their well tried manual methods.

In general, the introduction of an immunization system has resulted in a 10-15% rise in the immunization rate where the manual system was achieving 60% cooperation. The preschool system achieves over 90% acceptance in some areas.

All systems permit a considerable degree of flexibility in scheduling. Considerable effort has been expended in the development of training programmes which have further improved the acceptability of the package to the appropriate health authorities. This training package was directed towards physicians, nurses and clerical staff. The designers of the system did not feel that explicit education of the public would be cost/effective and thus concentrated their efforts on those providing the service.

One advantage of the British system is that it was originally developed by one authority acting as a "centre of excellence". It then proved possible to transplant it at small cost to other areas as hardware was standardized. No other authority is compelled to adopt the developed system, but good acceptance was ensured through the saving in development costs together with the inbuilt flexibility of the system.

3.7 Disease registers

The Working Group briefly considered the status of disease registers as examples of health data banks. The commonest of these relates to cancer. The main objectives of the cancer registration data banks include:

- (1) the establishment of a register of cancer cases;
- (2) registration of cases of suspected cancer;
- (3) the follow-up of cases of cancer of all degrees of severity;

- (4) feedback of information to appropriate health institutions;
- (5) a basis for epidemiological research; and
- (6) an estimation of resources used in the treatment of the disease.

A method for inviting patients to attend for follow-up examination has been developed in Czechoslovakia as part of their oncological data bank. This is similar to the one used for the child health status data bank in the United Kingdom and described earlier in this report.

Not all registers have the same objectives. For example, in the United Kingdom the main objectives of cancer registration have been reduced to those used for incidence studies, calculation of survival rates or as a basis of case-finding for further research.

The Group recognized, however, that data banks based on a single disease register might have only a limited value apart from the purposes for which they were originally established. Because of problems of both under- and over-reporting, care has to be exercised in the use of data from such registries unless assurance can be given that the problems involved have been overcome.

Because of the time and expenditure of resources needed to ensure exact reporting, together with standardization of diagnostic and other criteria, some members of the Working Group questioned the cost/effectiveness of such registers, suggesting that it was preferable and cheaper to use more general files from which to generate specific files relating to individual diseases and conditions. Some countries, however, such as Czechoslovakia, preferred to have better data on more important groups of diseases.

3.8 Health manpower data banks

The Working Group recognized that manpower data banks are a vitally important element of a total health information system.

In a labour-intensive industry such as the health services, the most important resource is manpower, and this is often more difficult to obtain and control than finance. Efficient personnel management and manpower planning need readily available data with which to work, in the form of both "stock" and "flow" data. The former relates to personnel matters at a particular time, the latter to movement into and out of the health system as well as to movement within it, and also includes data on personnel in training. This type of data needs to be qualitative, i.e., to include educational background and postgraduate skills and qualifications, as well as to be quantitative, i.e., to provide the number and distribution of various grades of staff.

Data for personnel management and manpower planning is often a secondary use of the data. Primary data may be collected for payroll purposes or for basic personnel records. Both the primary purposes have much in common but there is much to be said for the personnel data bank being separate from, though linked to, payroll data banks. Health manpower data

and the form of data banks that can be developed depend to a very large extent on the pattern of health services organization in a particular country. A totally nationalized health service will find it easier to obtain manpower data than one in which private medicine is still widely practised. However, in all systems there is a need for data on both trainees and qualified staff. It was noted, however, that in many countries information for the former is the responsibility of an education ministry rather than a ministry of health. Manpower data is, of course, required for all health service occupations, not just the medical, dental and nursing professions.

Health data banks are required not only to assist the control of the stock and flow of members of the health professions but also to ensure that they are geographically distributed according to health needs. In some countries this can be achieved by direct management, in others by control of posts, in others by the ebb and flow of market forces.

Control over the number of entrants for postgraduate training in particular specialties has been found to be essential in many countries to avoid dissatisfaction over career opportunities, severe surpluses and shortages.

Efficient design and use of health manpower data banks will be found to improve utilization of resources by enabling data on groups or individuals to be readily available to health service managers and planners.

3.9 Other health data banks

Not all health data banks relate to population, patients or health personnel. During the meeting, mention was made of data banks containing environmental health information, health equipment used for stock control procedures, drug data banks, and the data bank established within WHO for the exchange of information between Member States and WHO, and for the central internal exchange of health information.

The WHO Information System

The WHO Information System is an umbrella title, covering a number of existing and planned information subsystems which, though they differ as regards purpose, users, design and operation, are all related with one another in that they service technical cooperation programmes between WHO and individual countries and the international exchange of health information ("information transfer").

To meet these objectives, the WHO Information System is divided into two main parts:

(1) the Programmes Information System, consisting of subsystems to support technical cooperation by providing the information required for programme planning, management and evaluation; and

(2) Special-purpose Information Subsystems, mainly technically and/or scientifically oriented subsystems supporting "information transfer" which consists of machinery using WHO as a neutral ground for absorbing, distilling, synthesizing and disseminating information of practical value for countries in solving their health problems — in terms of its volume of data/information, the Special-Purpose Information Subsystems as a whole account for the greater part of the WHO Information System.

The Programmes Information System mainly acts as support for the following matters or at the following levels:

- (a) country project management and evaluation;
- (b) intercountry project management and evaluation;
- (c) interregional project management and evaluation;
- (d) country health programming;
- (e) medium-term programming, both at regional and headquarters echelons;
- (f) policy and executive levels and reporting to the Executive Board and the World Health Assembly.

The Programmes Information System is based on a minimum of well-defined information elements, referred to as a *profile*, which is used throughout WHO for formulating, and reporting on programmes and projects. The profile principle ensures: (a) that the information collected on related programmes or components is compatible and comparable, and (b) that a user of the system can get more information out than he has put in.

On the other hand, the Special-Purpose Information Subsystems embrace widely differing kinds of systems, but ones which can be categorized with a view to rationalizing existing subsystems and establishing whether a need exists for developing new ones. It is planned to study:

- (1) directories and non-bibliographical reference systems;
- (2) systems to maintain and process country health and health-related information, including a rationalization of the collection of information from Member States;
- (3) the processing of requests for information received by WHO;
- (4) information dissemination; and
- (5) standardized procedures and support for scientific surveys.

The design of the WHO Information System carefully avoids amalgamating these Special-Purpose Information Systems with the Programmes Information System. Instead, the design requires and caters for a liaison between them. Such a liaison may take the form of a transaction in one

system triggering an information communication to another, or simply of a direct cross-reference between systems.

It is clear that within the WHO Information System large volumes of the data/information are common to several users and uses, and that not all future uses and requirements can be foreseen at the design stage. There is therefore a fair measure of the desired independence of the data/information from present uses which might call for technological support for the WHO Information System that is largely data-base oriented.

Another property of the WHO Information System is its emphasis on decentralization, in line with the organizational structure and programme of WHO. This is reflected in the claim that information should be stored, as far as feasible, where it is more widely used, i.e., at a project site, in a Regional Office or at headquarters respectively.

The strategy which has been followed, is to proceed with the implementation of the Programmes Information System and, when this gains momentum, to turn to the Special-Purpose Information Subsystems. By the time that it has been completed, the development of a technological master-plan for the support of the WHO Information System will have passed through eight phases, as follows:

- I Data analysis,
- II Functional analysis,
- III Technological requirements (analysis),
- IV Technological environment (specification),
- V Detailed planning of subsequent implementation phases,
- VI Preliminary design of the technological master-plan,
- VII Detailed design of the technological master-plan, and
- VIII Full implementation of the technological master-plan.

Phases I and II are currently being carried out for the Administration and Finance Information Subsystem, which is part of the Programmes Information System.

An Organization-wide survey of the Special-Purpose Information Subsystems is under way. Its findings will be analysed to develop a strategy for the rationalization of existing Special-Purpose Information Subsystems and the development of new ones as required. Approval for these survey proposals was obtained and implementation in collaboration with the programme focal groups concerned began in January 1977.

By 1 January 1978, the Programmes Information System and a few Special-Purpose Information Subsystems are expected to have been completely redesigned and largely to have become operational. Subsequent implementation of other components of the WHO Information System would then proceed incrementally.

Drug data banks

The Working Group also discussed drug data banks. This is a particularly relevant part of health information systems because of the large and increasing number of active drugs and the multiplicity of information sources, not all of them reliable.

Drug data banks can be classified from the viewpoint of the users: industrial, therapeutic, or research. They may alternatively be classified according to the nature of the data: whether it summarizes existing knowledge or is used for the reporting of factual data on behaviour of the drug including its side effects.

The Group was told of the French BIAM system (*Banque d'Information automatisée sur les Médicaments*) which is designed to help physicians in prescribing drugs. The data bank consists of two sets of data, one of active principles, and the other for the proprietary preparations marketed in France. Its objectives are to provide up-to-date information on drugs and to allow selection compatible with characteristics of the patient, the disease and the drugs. It operates through special centres, telephoned by the physician and linked by a terminal to the data bank. The system is now in the final stage of evaluation.

Another system at Stanford, California, known as MEDIPHOP provides information on drug interaction.

Since the major task in the establishment of these systems lies in the initial collection of information, it was suggested that efforts should be made to avoid duplication of effort, so that further systems can be designed reasonably cheaply, probably at national level.

Data banks concerned with drug effectiveness or adverse effects rely on ongoing collection of information. Adverse effects may occur only rarely or only after long periods of use. Such data banks necessarily need to be organized on a large scale, usually at national or international level, such as the WHO Centre for International Monitoring or the Boston programme in the USA, because: (a) the reporting of adverse effects is voluntary; (b) there is a lack of knowledge of normal values and reactions; and (c) data collection is often unstructured, making validation and follow-up difficult or impossible.

The Swedish system, covering a limited area, has linked patients' names to the prescribing of drugs, and has been used in exploring the relationship between *Rauwolfia* and carcinoma of the breast.

It was suggested that there was a need for further education of the medical profession, standardization of terms and concepts, improvement of data collection networks and procedures and, if possible, linkage of central drug data banks with other computerized medical record systems. In addition, it was considered important for collaboration to be fostered between central drug data banks and those organized by pharmaceutical firms. The latter often have data on drug usage which can act as a useful complement to information on adverse reactions.

4. PRIVACY AND CONFIDENTIALITY

Throughout the meeting the issue of data protection arose in discussions. All countries have a need to protect the confidentiality of contacts between the physician and his patient, but the development of computers has been thought by many to put that privacy at risk. On the other hand, recent development of data protection legislation has tended to place restrictions on the handling of medical data so that bona fide clinical and epidemiological research is liable to suffer.

The members of the Working Group agreed that although no computer system can ever be rendered totally immune to unauthorized access, records held on such a system were always safer in this respect than written ones.

The inherent danger in computerization is that it can be linked with other records in a way which patient or physician may find objectionable. Discussion on issues of data protection has been very controversial, with one extreme argument being that no medical data must be stored or transmitted. At the other extreme, there exists a wish to neglect completely the requirements of confidentiality and privacy. The Group agreed that both extremes are wrong, the correct objective being a middle course that maintains the secrecy of individual information while offering the convenience of a health information system and suitable for epidemiology, health administration, management and research uses.

In many countries, laws have been passed, or are in the process of being drafted, to secure the citizen generally against risks resulting from improper use of storing data about him. Although this has been the common objective, no single method has been adopted, if only because of the varying existing social and health systems.

Data protection can be achieved by three general methods:

- (1) hardware precautions,
- (2) software techniques, and
- (3) organizational methods.

The first two methods are largely technical; in general, it can be said that the security of hardware and software can be as good as the investment of resources in them. Given this, security is no longer a problem, although some data banks have to be larger than would otherwise be the case, because of the need to label each item of data with its security classification. On the other hand, organizational provisions to protect the confidentiality of data depend to a large extent on the legal and ethical requirements of the country and its institutions. There are various categories of confidentiality and of types of persons authorized to use data. Each type of user should have access only to data which correspond to his needs. For example, at the local level physicians

should have access to most detailed data, whereas at central or national level, restricted data may be available often only in unidentified form.

There was general agreement that identified medical data held in health data banks should not be transmitted to data banks for other disciplines. On the other hand, there should be encouragement for anonymous data from various types of data bank to be aggregated for statistical purposes in a meaningful way.

The Working Group recommended that the different countries still in the process of drafting data protection legislation should take into account the needs both of patient care and medical research. It should be realized that for this purpose it is often necessary to use personal information on a patient, including identification.

5. COMMUNICATION WITH USERS

The Working Group agreed that the utilization of many health data banks has so far been disappointing. While, in some instances the reason for this may be in the nature of data collected or in the design of the system, too often there exists a communications gap between those responsible for the data bank and those who wish to use it or ought to wish to do so.

When health information was first computerized there was a certain tendency to attempt to deal with all available information regardless of its value. By now, though, the situation should be different. Given the state of computer technology and information handling, data banks should hold information in such a way as to make it available to those who need it quickly and in readily usable form. This applies both to hard data, such as statistics, and less hard data, for example information on clinical practice. However, many users remain disappointed with the information they obtain from data banks.

In analysing the problem, it has to be realized that the background and professional views of those concerned with the building and use of information systems do not help them to communicate easily with one another. Computer scientists, providers of health care, administrators and epidemiologists are all involved. Each complains of a lack of understanding from the other, with resulting evidence of a conflict between different power groups, involving access to information by different professions or subgroups. This involves both a conflict of competence and a conflict of professional interests. While most people agree that a well-organized information system ought to be a useful tool for decision-making, nevertheless many seem afraid to leave the use of this tool to others. In other words, there is a lack of trust. Sometimes it even seems that the ultimate aims of a health service are lost in such debates and multi-professional cooperation is forgotten.

If lack of understanding and trust is one source of problems, an excess of information can be another, particularly when it cannot be related to the people or populations concerned. Too often, people prefer manual systems because the available output from data processing schemes appears to be directed to the wrong questions.

During the Group's discussions, mention was frequently made of the importance of employing a specialist whose role it is to interpret the users' needs to staff employed on the information system itself, and also to do the converse, i.e., to interpret the needs of information systems to their users. Such an individual could do much to improve relationships and increase the value of both data banks and health information systems.

It was suggested that, in developing health information systems it is important to keep in mind the criteria that they should be as simple as possible, that they should be problem-oriented, and when they relate to people, that they should be person-specific and population-based.

Health service information systems may be defined as "those whose primary purpose is to provide intelligence for use in decision-making by organizations and individuals responsible for health care". With regard to health data banks, it is not only the storage and retrieval of data which matters but also the purposeful reduction of data re-ordering and interpretation so as to produce useful and usable information in as simple a form as possible.

6. ORGANIZATION AND TECHNIQUES OF INPUT INTO DATA BANKS

Much thought should be given to the techniques by which data are collected before being entered in data banks. Data collected outside normal recording procedures do not generally achieve the same accuracy in a long-term programme, and consideration should be given to methods by which the usual procedures can be improved or at least maintained. At the same time, the standard of data collection has to be supervised and maintained by clear instructions. Wherever possible, data should be collected at its place of origin but should be transmitted to the appropriate computer centre according to a prearranged routine. This applies equally whether there is local batch processing of data or input via remote job-entry terminals.

Before data is accepted in a data bank it has to be checked for both validity and practical usefulness. Ideally, the user should be provided with a feedback to show what has been entered in a computer, so that he can himself check its accuracy. This is not always possible, however, in which case the use of validity and feasibility programmes assumes greater importance. Similar

principles apply to the "conversational mode" of real-time input, but here the provision of feedback to the individual entering data is much simpler to arrange.

7. DATA BANK CONCEPTS AND CONSIDERATIONS ON HARDWARE

The Working Group emphasized that a data bank is only one part of an information system. Although there was some debate over terminology, participants arrived at a consensus that a data bank is an integrated collection of data that are related in a meaningful way and that can be reaggregated in different logical orders by several independent users and applications, but where each data item is stored only once. There was also agreement that, wherever possible, the structure of existing files, procedures and organizations should be taken into account in the establishment of data banks as this will reduce the real cost of data collection.

Data bank concepts and data base management systems represent a major step forward; nevertheless problems and disadvantages still remain. For example, with the reduction of redundant data, a data bank is now more vulnerable to destruction through machine malfunction, human error, or deliberate tempering. The other problem area is that of security, where data is available on request to a large community of users.

In a traditional file system, data is held or processed separately for each application, but with a data-base approach all data are managed through the same software, and the design of the latter determines the successful operation of the data bank.

In order to achieve this a data base management system must have control and integration of data, coupled data structuring, retrieval and maintenance of data, data independence, concurrent usage and privacy controls.

The administration of a data base is important, in order to exploit fully the advantages of technology, to satisfy the user and to fulfil his requirements quickly and accurately. For this, a data administrator is required to liaise with users, decide the information content of the data bank, the storage structure and access strategy. He must also define appropriate validity and feasibility checks, and monitor performance. Finally, he has to be able to respond to a change in user requirements.

Despite the progress that has been made in the last decade, data-base technology is still in its "first generation". The performance of many data-base management systems is poor because of bad design or because of the complexity of the software.

There was some discussion on the hardware to be used in data bank operation. Opinions differed as to whether the best future prospects lie in using several special-purpose mini-computers for data banks or whether it is better to employ one (or more) large computer systems. No clear-cut recommendation was possible except to call on designers of data banks to take recent technical developments into account, and to consider the cost/effectiveness of the system architecture available in relation to resources and the objectives of the project.

8. TRAINING

The Group considered that the training of health personnel of all types in the use of statistics, information systems and computers is still inadequate. They felt that general information on these topics should be included in undergraduate training for medical and other health professions, with the objective of providing an insight into the importance of these topics in their profession.

In postgraduate training, emphasis should be put on cross-fertilization between the various professions involved in the design, construction operation and use of information systems and data banks. This is particularly important in improving cooperation between different groups of specialists and in prevention of the growth of the type of communication barriers discussed elsewhere in this report.

It was suggested that postgraduate training might be organized at various levels:

- (a) basic: general background information;
- (b) intermediate: the need to ensure that all concerned in the operation of data banks understand the purposes and methods used;
- (c) advanced: preparation of the specialist to advance knowledge and to take part in teaching.

9. EXCHANGE OF INFORMATION

There was a general discussion on the need to exchange information about the nature, functions, and design of medical data banks. There is often a lack of knowledge about work going on even in the same country and there is known to be no coherent description of the various medical data banks in Europe, while there is more knowledge of the situation in the USA.

An IFIP project was set up during 1975 to gather information on medical data banks, but progress to date has been disappointing, possibly because the task is so large. Doubt was also expressed on the feasibility of such schemes because of lack of cooperation from users in submitting detailed information when the resultant feedback is of little direct interest. There was general agreement that cooperation is best fostered by exchange of knowledge between interested individuals and groups. WHO and other international organizations and societies have a role to play in this by the calling of conferences, meetings, etc.

10. TERMINOLOGY

While no clear-cut agreement on terminology was reached by the Working Group there was a consensus that a data bank is part of an information system (see p. 6) but there is no standard terminology available for terms such as "data bank", "information system", "data base", "files", etc. While such standardization is desirable, it is probably an unrealistic objective at present. It is therefore all the more important that emphasis should be laid on unambiguous and clear-cut definitions of the terms used in any particular project, and that these should always be given in descriptions of the project.

11. CONCLUSIONS AND RECOMMENDATIONS

The Working Group arrived at the following conclusions and recommendations.

(1) The general objectives of health data banks should be consistent with the overall objectives of the health services. To achieve this it is wise to consider the application of criteria that data should be parsimonious, problem-oriented, and (when they relate to individuals) person-specific and population-based.

(2) In considering the design and implementation of health data banks, the specific objectives must first be delineated and the possibility of more general applications considered should the occasion arise. These objectives will necessarily be considerably affected by the type of organization of the health service in the country concerned. Close attention should be paid to the need for realism and cost/effectiveness.

(3) The Working Group emphasized that a "data bank" is part of an "information system". In an attempt to define a data bank the Working

Group arrived at a measure of consensus that it is an integrated collection of data which is interrelated in a meaningful way, can be accessed in different logical orders by several independent users and for different applications and has each data item logically stored only once. Wherever possible, the construction of data banks should take into account the use of existing files, procedures and organizations as this will reduce the real cost of data collection.

(4) In designing health information systems it is necessary to take into account recent technical developments and to consider the cost/effectiveness of the system's architecture in relation to available resources and the objectives of the project.

(5) Great care has to be taken to see that the accuracy, reliability and validity of the data is understood in the context of the questions asked, particularly when the information is being used for multiple purposes. It should be remembered that it may be both easier and much cheaper to use data already stored than to collect data for a specific study, sometimes *ab initio*.

(6) Health data banks which contain data relating to persons should, if possible, be related to the total population of the region or country or to other well-defined population groups under review in order to provide denominator data. Population-oriented health data are needed at the regional and national levels of administration in order to give guidelines for the allocation of resources and in the planning of health systems. These data are of great importance for epidemiological studies.

(7) Many other types of data banks relate, for example, to drugs or the environment, which may not depend on person- or population-based information. The data of other, equally relevant, data banks relating to the social or economic status of persons should be compatible with health data banks.

(8) The Working Group considered data protection, e.g., confidentiality, security and privacy, on several occasions throughout its deliberations. It considered that data accumulated in data banks are often more protected against misuse than data filed in a traditional way. All the members were concerned to ensure that the restrictions placed on the use of computerized data banks should not unduly restrict the use of patient data for patient care and research purposes.

The situation existing in different countries prevents the formulation of detailed recommendations. The Working Group nevertheless recommends that when legislative authorities are drafting privacy laws they take into account the needs of patient care and medical research, realizing that for this purpose the use of personal information about the patient, including personal identification, is often necessary.

It was the view of the Group that health data banks should remain as separate entities, i.e., person-related information should not be transferred to other non-health data banks.

(9) It was recognized that, whereas in several countries hospital inpatient information systems are operational, the same cannot be said of information systems covering either outpatient departments or general practice. The Group recommended that every encouragement should be given to the collection of data on ambulatory care. For example, data should be collected on primary care in order to provide information which would be of value in policy making, research, management and education.

(10) Although there has been a tendency to try to distinguish different types of health data banks, it was appreciated that these distinctions are very difficult to draw clearly and indeed that there should be encouragement given towards the linking of health data banks for which data should be compatible. In order to obtain more efficient health care and health information systems the same data can often be collected by clinicians and used by them as well as by other health professionals.

It is recommended that, wherever feasible, data should be collected once and used for as many relevant purposes as possible. The compatibility and linkage of primary care registrations; hospital data; preventive, social and health-environmental information systems; and health care delivery systems is recommended.

(11) There was general agreement that when data is transmitted from peripheral to more central data banks it should be in a non-aggregated form, as this allows for greater flexibility in programming, processing and analysis.

(12) There was considerable discussion on the meaning of terms such as data bank, information system, data base, files, because there is no standardized terminology available. Such standardization is desirable and the Working Group therefore recommends that:

(a) in discussions on the general topic of health information systems clear definitions, together with descriptions and criteria of the terms used, should always be given; and

(b) WHO/IFIP and other organizations should consider in more detail whether a standard terminology is possible, taking into account the views of health personnel, management staff and computer specialists.

(13) The Group also felt that standardization of definitions, classifications and possibly coding of data used in health information systems should at least be attempted within each Member State, and the Group recommended that WHO should direct its standardization efforts in this direction.

(14) General information and elementary training on significant aspects of information science should be included in the curricula of undergraduate training in health matters with the objective of providing education on the importance, uses and impact of these topics in relation to the various fields of health.

Postgraduate training should be given to ensure that there is appropriate cross-fertilization between the different professions involved in health information systems, both as users and providers.

(15) The Group appreciated WHO's efforts towards making the Ninth Revision of the International Classification of Diseases more applicable to general practice and primary care, but felt that further attention to this problem would be necessary in the future. It also recognized that no adequate problem-oriented classifications exist in several fields, and recommended that WHO should take steps towards the further development of such classifications, taking into account known experience.

(16) The Group considered how WHO, its Member States and IFIP could assist in disseminating information on the experience of different countries in the design, implementation and use of health data banks. It is recommended that the exchange of such information should be encouraged by all possible means, e.g., through publications, meetings, conferences, fellowships and professional societies.

Annex

LIST OF PARTICIPANTS

Temporary Advisers

Dr B. Baikushev, Chief, Department of Information Systems and Computer Centre, Research Institute of Social Hygiene and Public Health, Sofia, Bulgaria

Dr J. Boukal, Director, Department of Therapeutic and Preventive Care, Ministry of Health of the Czech Socialist Republic, Prague, Czechoslovakia

Professor K. Bauknecht, Institute for Data Processing, Zürich University, Switzerland

Dr C. Bruins, Netherlands Institute of General Medical Practice, Utrecht, Netherlands

Professor J. Červenka, Chief, Department of Epidemiology, Institute of Postgraduate Medical Education, Bratislava, Czechoslovakia
(*Chairman*)

Dr W. Duncan Dolton, Regional Specialist in Community Medicine (Service Planning), Mersey Regional Health Authority, Wilberforce House, Liverpool, United Kingdom

Dr O. Fokkens, Director, Medical Data Bank Foundation, Utrecht, Netherlands

Dr M. Goldberg, Senior Lecturer, Medical Data Processing Service, Pitié-Salpêtrière Hospital, Paris, France

Professor G. Griesser, Institute of Medical Statistics and Documentation, University of Kiel, Federal Republic of Germany

Dr M. Heasman, Director, Information Services Division, Scottish Health Service, Common Services Agency, Edinburgh, United Kingdom
(*Rapporteur*)

Professor J. Martin, Medical Data Processing Unit, Medical Faculties A and B, University of Nancy, France

- Mr J. Massué, Scientific Counsellor, Council of Europe, Strasbourg, France
- Dr J. Mathiot, Director, Department of Research and Data Processing, Public Assistance Hospitals of Paris, France
- Dr G. Penzel, Organization and Computer Centre, Carl Gustav Carus Medical Academy, Dresden, German Democratic Republic
- Dr F. Roger, Medical Data Processing Centre, Catholic University of Louvain, Belgium
- Dr M. Santrucek, Deputy Director, Department of Health Statistics, Institute of Social Medicine and Organization of Health Services, Prague, Czechoslovakia
- Professor F. Sawicki, Chief, Department of Medical Statistics, National Institute of Hygiene, Warsaw, Poland
- Associate Professor B. Smedby, Department of Social Medicine, Academy Hospital, Uppsala, Sweden
- Dr I. Väänänen, Chief Administrator, Helsinki University Central Hospital, Finland
- Mr J. Valenta, Director, Department of Information Systems and Automation, Ministry of Health of the Czech Socialist Republic, Prague, Czechoslovakia
- Dr K. Zaruba, Director, Department of Health Statistics, Institute of Social Medicine and Organization of Health Services, Prague, Czechoslovakia

International Federation for Information Processing

Technical Committee 4 – Information Processing in Medicine

- Mr J. Roukens, SAZZOG Foundation, Arnhem, Netherlands (*Co-Chairman*)
- Dr J. van Egmond, Department of Medical Data Processing, Gent, Belgium (*Co-Secretary*)

World Health Organization

Regional Office for Europe

Dr Z. Brzezinski, Regional Officer for Epidemiology (*Secretary*)

Dr D. Sokolov, Chief, Strengthening of Health Services

Dr A. Weber, Regional Officer for Health Information

Headquarters

Mr J. Dowd, Health Statistical Methodology

Dr S. Mandil, Information Systems Programme