Section of Epidemiology & Preventive Medicine

President J N Morris FRCP

Meeting May 13 1971
at the Regional Hospital Board,
Old Road, Headington, Oxford

Papers

The Oxford Record Linkage Study as a Medical Information System

by J A Baldwin MD MRCPsych
(*Oxford Record Linkage Study,*
Regional Hospital Board,
Old Road, Headington,
Oxford, OX3 7LF)

The Oxford Record Linkage Study was established in the early 1960s to test the feasibility of linking records of medical research interest from several sources (Acheson 1967). It had been recognized that information of immense potential value for the furtherance of medical knowledge was accumulating in the growing numbers of medical and other records of individual patients for the treatment of specific illnesses or for recording certain events, such as birth or death, but was inaccessible except by very laborious and costly research procedures. The increasing importance of chronic disorder and genetic factors, coupled with the need to obtain morbidity data to supplement mortality statistics, suggested the idea of extracting systematic information from the multitude of records and linking it to form a coherent cumulative personal health record system well fitted to cohort and follow-up studies.

The feasibility of record linkage on a sufficient scale was quickly confirmed and from 1963 the Study has acquired systematic data on all general and psychiatric hospital discharges, birth and death records, and all maternities, whether in hospital or at home, occurring in a population of about 350,000 in a segment of the Oxford Hospital Board Region. From 1966 coverage was extended to a population of about 800,000 yielding an input of about 100,000 records per annum. Record linkage and analysis of the data file on this scale demands computer assistance, and a major effort has been put into developing an operational name-matching procedure and creating a usable cumulative file. This has now been achieved and analyses of the first five years of data are being obtained (*see* paper by Dr A S Fairbairn, p 239). By the end of 1971, the linked files will span eight years, and will provide follow-up periods of increasing interest as time goes by.

Early in the life of the Study it was realized that unlinked systematic data on events such as hospitalization were of great value for health service management and planning activities, and for the selection of samples for clinical research. Indeed, most of the statistical work so far has been in these categories and demand is increasing. There is now a need for up-to-date returns on hospital activity in the form of tables and indexes, at least to maintain a monitor on hospital function, as well as to probe questions of operational interest.

Substantial *ad hoc* enquiries from clinical and administrative staff now number 100 per year, and perhaps the same number of simple enquiries are dealt with as well.

In considering further development of the system, it has been recognized that certain difficulties and drawbacks would have to be overcome, or at least minimized. In administrative eyes, much of the justification for any information system is in its management potential, while hospital staff who contribute to it rightly expect a return which meets their real needs. Among many problems the following seemed most pressing:

(1) The file as it stands is extremely limited, covering only a fraction of even treated morbidity. In some specialties, such as psychiatry, this is a

diminishing and increasingly unrepresentative fraction (Baldwin 1971). Chronic disorder is being treated more and more outside the hospital bed, and indeed outside the hospital service altogether.

(2) The population at risk is really too small for sufficient numbers of cases of many epidemiologically interesting conditions to be obtained in a short period.

(3) Very few data are available on each type of event; clinical data, particularly, are not adequate for many potential studies.

(4) The system was not designed as a rapidly responding, up-to-date file capable of dealing with day-to-day management and clinical enquiries. The computer files cannot be much less than 18 months out of date, and the file organization is not suitable for fast, economical retrieval of small subsets.

(5) Information needs are becoming more varied and subject to more rapid change than hitherto. Specialized interests and topical problems must be catered for, so that the system design must be flexible and capable of accommodating a wide range of data types.

(6) The twin problems of unreliability and underreporting, which are commonly encountered in register systems, are probably the most fundamental issues. While a measure of random unreliability and missing information may be of little consequence for many large scale statistical analyses on such files, their potential benefits lie as much in their ability to meet the needs of a range of users, from area administrative statistics to the clinical interest in small numbers of rare or particular disorders, and even the individual case. Furthermore, error sources are usually found to be far from random, and systematic bias may lead to misinterpretation.

In choosing between possible ways in which the record linkage system might be developed, a number of factors had to be weighed, including potential epidemiological research value, administrative and financial supports available largely within the hospital service, the likelihood of obtaining the co-operation of participating agencies, and the size of the problem in terms of data acquisition, data volumes and processing capability.

It was considered to be most useful to extend the existing coverage of general and psychiatric inpatient, birth and death records to the whole regional hospital board area of 1·9 million population, combining the record linkage approach with the development of the management information system referred to as hospital activity analysis. Coverage of maternity records might also be extended to the same area. In view of the special needs and interests in psychiatry and subnormality, an experimental extension to outpatient and extramural services in a limited area has also commenced.

Acquisition of morbidity data from other branches of the health services presents special problems. The Unit of Clinical Epidemiology has been experimenting with recording systems in general practice for the past five years and development has now reached the point at which it is possible to consider a larger scale system incorporating both general practice and local health authority data. Such a linked file on a large enough population could provide both morbidity and population data of much wider scope than a hospital-based information system. A system of this kind is now being developed under the title of the 'Community Health Project' (*see* paper by Dr J Perry, p 241).

Development on the scale envisaged in the hospital service alone would entail an increase of 2·4 in data volume, and would necessitate redesign of both the file structure and the data acquisition procedure. In trying to tackle the main problems the attempt has been made to create a more versatile, responsive and flexible information system catering for management, clinical and operational research uses, as well as for epidemiology.

The chief principles on which the system is now being developed are:

(1) Data collection is an integral part of service function, not a superimposed and retrospective activity. Information is to be obtained at its point of origin, whether this is on admission to hospital, at discharge, or after. Most of the information required by the system is already acquired for immediate use in the hospital, and it should become available to the system as a by-product. In many cases, the medical record itself should be systematized so that the same data can enter both the clinical record and the information system. Staff engaged in data collection should be trained to the task, and this should include interviewing patients and co-operating with clinical staff.

(2) There must be maximal relevance and pertinence of the information system to the service feeding it. Arrangements must be made to cater for local and special interests, such as alternative diagnostic coding. Some services should be made available, such as routine death clearances for

records departments. The possibility of developing a monitoring scheme for long-term drug therapy is being studied. The system must be capable of giving a timely response to enquiry. Analyses pertaining to the previous quarter year should be available within two months of the end of it.

(3) The system must monitor its own status, with extensive ongoing editing procedures and output of lists of errors for correction. Many types of missing records or parts of records can be detected by an appropriately designed system. Coupled with careful training and supervision of records staff engaged on data collection, these measures can go far to reduce error and underreporting to manageable proportions.

(4) Introduction of new techniques of data preparation would reduce costs and permit dispersal of the whole process of data acquisition into hospitals. Optical character recognition and punched paper tape were both thought to offer greater flexibility and lower cost than the traditional punched card. Plain language could be input and automatic encoding would become a practical possibility, thus improving reliability and probably also reducing costs still further.

A system has now been developed incorporating these principles, and has been successfully operated on a pilot basis for general hospital inpatient data. It is expected that it will be implemented throughout the region in the very near future (Unit of Clinical Epidemiology and Oxford Record Linkage Study 1971).

The scope for further development of such systems in the Health Service is very great and the potential has hardly yet been tapped. The technical ability to link records is now fairly well established both for the cumulative person record, and for family record linkage (Baldwin 1972).

Exploitation of linked files is now becoming worthwhile, but the already high demand for unlinked information must be satisfied in order to justify the system over the long period required to reap the eventual benefits of linkage.

REFERENCES
Acheson E D (1967) Medical Record Linkage.
London, New York
Baldwin J A
(1971) The Mental Hospital in the Psychiatric Service:
A Case-register Study. London, New York
(1972) Aspects of the Epidemiology of Mental Illness: Studies in Record Linkage. Boston (in press)
Unit of Clinical Epidemiology and Oxford Record Linkage Study
(1971) A Medical Information System
for General Hospital Inpatients: Preliminary Assessment
of a Pilot Scheme, Research Report No. 1.
Oxford Regional Hospital Board.

Hospital Admissions Following Common Surgical Operations

by A S Fairbairn MA MB
*(Oxford Record Linkage Study,
Regional Hospital Board, Old Road,
Headington, Oxford, OX3 7LF)*

We now have a magnetic tape file containing numerical particulars of hospital discharges and deaths in Oxford City and County for the five years 1963–67. This is in 'linked' form so that information on successive events happening to the same patient is brought together in sequence on the file. It relates to 90,000 inpatients who between them have experienced 135,000 hospital discharges, drawn from an area whose population is about 800,000. By scanning the tape by computer, defined cohorts of patients can be followed and their hospital readmissions occurring within the five years counted and tabulated. In a population of this size, patients with common conditions who are followed for several years are likely to have appreciable numbers of readmissions purely by chance, that is if each patient stands the same chance of admission as someone of the same age and sex in the general population. The observed number of readmissions for any cause is therefore of importance only if it significantly differs from the expected number. I have looked at the observed and expected numbers of readmissions by cause following five common surgical operations and indicate here some preliminary findings.

Table 1
Number of operations and man-years of observation after operation, Oxford City and County, 1963–67

	GRO code of operations (1956)	*No. of operations*	*Man-years of observation*
Tonsillectomy and adenoidectomy	261–263	6,823	17,823
Appendicectomy	441	3,099	7,783
Hysterectomy	722–724	2,539	6,198
Inguinal herniorrhaphy	402	2,279	5,602
Cholecystectomy	521	955	2,275

The operations are listed in Table 1. Assuming that the operations are evenly spaced and that each patient is followed to the end of 1967, the average length of observation per patient is 2½ years. The actual man-years of observation at all ages, which take into account deaths occurring before the end of 1967 but which disregard the unknown effects of migration, were calculated by computer and are given in the last column of Table 1.

To obtain the expected number of readmissions after operation for any diagnostic cause, it was