

Record linkage in automated systems

ONCE data exist in automated form, the advantages over manually recorded data are enormous. Automation of general practitioner records has made possible the organized collation, storage and analysis of data about each consultation and each service to our patients, whatever the source or time. More often than not we take for granted the quality of the record which results from linking data from successive consultations. However, the quality of this linked record can be no better than the quality and accuracy of the patient identifier to which all other information must be linked. The identifier may be an arbitrary *ad hoc* number or code for each patient. The only practical disadvantage of arbitrary numbers is that they must be generated centrally, so that no two patients have the same number and so that the individual patient can be identified from this number. Also, the patient must have an accessible copy of his identity number at all times.

The identifier may be a number or it may be a code such as the Hogben Number. The Hogben Number is a number constructed from the six digits of the full date of birth, together with the first three letters of the patient's surname and the first letter of the patient's first forename. The original Hogben Number also carried a suffix to indicate the patient's sex. The Hogben Number can be regenerated from the patient's memory at each doctor-patient contact, can be generated by anyone who knows the necessary information at any time and can even be available for retrospective enquiries provided that the original record contains the patient's name and the date of birth is accessible.

Errors

However, people forget their date of birth and make errors when writing it down or on requesting it from others. Even when a record should conventionally carry the patient's full name and date of birth, it may be omitted if the originator sees no point in recording the information. For these and other reasons Boyle (1981) questioned in last month's *Journal* the value of the Hogben Number in an analysis of the quality of the data from the Virginia Family Practice Data System, where a disappointingly high error rate of 11.3 per cent was disclosed in Hogben Numbers used to identify patients. This is a similar result to that found by Acheson (1967), in the Oxford linkage study, which also used the Hogben Number to identify patients.

The Hogben Number was also being used from 1959 onwards as the patient identifier in the Royal College of General Practitioners' Disease Index, which was the basis for the Second National Morbidity Survey in 1970-71. In that study the error rate was reduced to a trivial figure (OPCS, 1974) by observing certain safe-

guards. Everyone participating in the study was aware that patients' names and dates of birth were the basis for the identifier and that it was important that records of these should be accurate. In particular the Hogben Number was generated with particular care when the initial age and sex register card was completed. This set of cards became the basic register for the practice population and the reference point for names and dates of birth. It was also the basis for a central automated version of the age/sex register which was set up at the beginning of the study.

All other data for linkage, accompanied by the appropriate Hogben Number (whether clinical from disease indices or socio-economic from census files), was first matched with this computer file. Non-matching led immediately to a search for the source of error and to its correction. Most often the error was in the new input, but it was occasionally in the original age/sex register. In systems where the matching process takes place a considerable time after the original recording, as in the Oxford and Virginia studies, the discovery of errors comes too late to trace their source effectively or economically. This difficulty is compounded if the new data are linked only to previous data and not first to a master register or index. This was also the method used in the Oxford and Virginia studies.

The consequence of errors

Errors in the identifier have two possible results—non-matching and mis-matching. Only non-matching identifies itself immediately as an error; mis-matching, where new data are linked to the wrong patient's computer file, is a more serious error, for it is impossible to detect. When the identifier is a special number from a given sequence, say from 1-9,999 for practices with more than 999 but less than 9,999 patients, then a wrong entry, provided that it is within the range allowed, will always find a match. Hogben Numbers, because they always contain an enormous potential of redundant information, have a very small chance of finding a mis-match (Crombie, 1973). For example, in a population of 3,000, slightly bigger than the size of the average practice, the chances of an inaccurate Hogben code making a false match are relatively better than one in 33,000. For Hogben Numbers, only non-matching errors are of major significance and these are identified immediately if the first step is matching new input with a basic register of all possible patient identifiers.

Confidentiality

Choosing the identifier involves issues of confidentiality as well as error rates. When identifiers are a set of arbitrary alpha-numeric characters, confidentiality can

be ensured simply by having only one cross-referenced index of identifiers and patients, and keeping it in the hands of the patient's own doctor. However, the automated linkage of data is thereby limited to this one source. Cross-linkage of morbidity data from general practice with other data is possible only when the identity of the individuals is known to both agencies. However, by using the Hogben Number it is possible to cross link two sets of data without breach of confidentiality. The first stage in such a linkage exercise is to establish a population with one of the sets of data which will include all the individuals in the other population. For example, in the case of socio-economic data, one practice will be confined within a geographic area. This area will be included in the census enumeration districts for the geographic area. The Second National Morbidity Survey (OPCS, in press) linked general practice and census data in this way and found that each practice population represented, on average, about 10 per cent of the total census population for the practice area. Once the appropriate Hogben Numbers have been generated, the morbidity and census data can be matched and merged. Only the patient's date of birth is retained; the rest of the Hogben Number is discarded and the patient's identity is suppressed. Provided that the Hogben Numbers for the individuals in the two sets of data are first prepared with care and the numbers from the two sources to be matched and merged at any one time are kept below 30,000, the mis-matching errors will be less than one in 3,300. Non-matching demands a careful search for the source of error in the identifier or the original register of patients at risk.

Boyle's (1981) paper, emphasizing accuracy of the identifier, is timely. The qualities of the Hogben Number may lead us to believe that we can take short cuts in using it. However, the real benefits of the Hogben Number will come only from building on its virtues, in particular its usefulness in maintaining confidentiality. If we do this, we should achieve even better results than can be obtained with a simple numerical identifier.

D. L. CROMBIE
Director, RCGP
Birmingham Research Unit

References

- Acheson, E. D. (1967). *Medical Record Linkage*. London: Oxford University Press.
- Boyle, R. M. (1981). Accuracy of the patient identifier in a family practice data system. *Journal of the Royal College of General Practitioners*, 31, 242-244.
- Crombie, D. L. (1973). Research and confidentiality in general practice. *Journal of the Royal College of General Practitioners*, 23, 863-879.
- OPCS. (1974). *Morbidity Statistics from General Practice. Second National Study 1970-71. Studies on Medical and Population Subjects No. 26*. London: HMSO.
- OPCS. (In press). *Morbidity Statistics from General Practice. Second National Study. Studies on Medical and Population Subjects*. London: HMSO.

Inderal Abridged Prescribing Information

DOSAGE AND ADMINISTRATION: 1. **HYPERTENSION** INITIALLY 80 MG TWICE DAILY INCREASING TO 160 MG TWICE DAILY AFTER ONE WEEK, AS NECESSARY. 2. **ANGINA** INITIALLY 40 MG TWO OR THREE TIMES DAILY INCREASING AS NECESSARY AT WEEKLY INTERVALS. AN ADEQUATE RESPONSE IS USUALLY SEEN IN THE RANGE 120-240 MG/DAY. **CONTRAINDICATIONS, WARNINGS ETC.** 'INDERAL' SHOULD NOT BE USED: 1. IN THE PRESENCE OF SECOND AND THIRD DEGREE HEART BLOCK. 2. IF THERE IS A HISTORY OF BRONCHOSPASM. 3. AFTER PROLONGED FASTING. 4. IN METABOLIC ACIDOSIS (EG IN SOME DIABETICS). 5. WITH VERAPAMIL, AND NEITHER DRUG SHOULD BE ADMINISTERED WITHIN SEVERAL DAYS OF DISCONTINUING THE OTHER. **PRECAUTIONS** 1. SPECIAL CARE SHOULD BE TAKEN IN PATIENTS WHOSE CARDIAC RESERVE IS POOR. 2. BRADYCARDIA (USUALLY LESS THAN 50-55 BEATS/MIN) INDICATES THAT DOSAGE SHOULD NOT BE FURTHER INCREASED. 3. IT IS IMPORTANT THAT A BETA BLOCKING AGENT IS NOT DISCONTINUED ABRUPTLY. 4. AS WITH ALL OTHER DRUGS, 'INDERAL' SHOULD NOT BE GIVEN IN PREGNANCY UNLESS ITS USE IS ESSENTIAL. 5. IF 'INDERAL' AND CLONIDINE ARE GIVEN CONCURRENTLY THE CLONIDINE SHOULD NOT BE DISCONTINUED UNTIL SEVERAL DAYS AFTER THE WITHDRAWAL OF THE BETA BLOCKER (SEE ALSO PRESCRIBING INFORMATION ON CLONIDINE). **ANAESTHESIA** 'INDERAL' MAY CAUSE AN ALTERED RESPONSE TO STRESS AND THEREFORE IT MAY BE NECESSARY TO WITHDRAW THE DRUG BEFORE SURGERY: SEE THE DATA SHEET. **ADVERSE REACTIONS** 'INDERAL' IS USUALLY WELL-TOLERATED. MINOR SIDE EFFECTS SUCH AS COLD EXTREMITIES, NAUSEA, INSOMNIA, LASSITUDE AND DIARRHOEA ARE USUALLY TRANSIENT, RESOLVING ON WITHDRAWAL OF THE DRUG. THERE HAVE BEEN REPORTS OF SKIN RASHES AND/OR DRY EYES ASSOCIATED WITH THE USE OF BETA BLOCKING DRUGS. THE REPORTED INCIDENCE IS SMALL AND IN MOST CASES THE SYMPTOMS HAVE CLEARED WHEN TREATMENT WAS WITHDRAWN. DISCONTINUANCE OF THE DRUG SHOULD BE CONSIDERED IF ANY SUCH REACTION IS NOT OTHERWISE EXPLICABLE. CESSATION OF THERAPY WITH A BETA ADRENERGIC BLOCKER SHOULD BE GRADUAL IN THE RARE EVENT OF INTOLERANCE TO 'INDERAL' MANIFESTED AS BRADYCARDIA AND HYPOTENSION, THE DRUG SHOULD BE WITHDRAWN AND TREATMENT INSTITUTED AS FOLLOWS. **OVERDOSAGE** EXCESSIVE BRADYCARDIA CAN BE COUNTERED WITH ATROPINE 1-2 MG INTRAVENOUSLY, FOLLOWED, IF NECESSARY, BY A BETA RECEPTOR STIMULANT SUCH AS ISOPRENALINE 25 MICROGRAMS INTRAVENOUSLY OR ORCIPRENALINE 0.5 MG INTRAVENOUSLY. 'INDERAL' IS A TRADEMARK FOR PROPRANOLOL HYDROCHLORIDE. FULL PRESCRIBING INFORMATION IS AVAILABLE FROM: IMPERIAL CHEMICAL INDUSTRIES LIMITED, PHARMACEUTICALS DIVISION, ALDERLEY HOUSE, ALDERLEY PARK, MACCLESFIELD, CHESHIRE SK10 4TF. **P.L. NOS** 0029/5064 (40 MG TABLETS). 0029/5065 (80 MG TABLETS). 0029/0103 (160 MG TABLETS). **BASIC NHS PRICES:** 40 MG 250...£9.69, 1000...£36.63; 80 MG 100...£5.82, 500...£27.37; 160 MG 50...£5.82, 250...£27.37.



INDERAL

Propranolol Hydrochloride