

AN EPIDEMIOLOGICALLY-ORIENTATED COMPUTERISED
CLINICAL CASE RECORDING SYSTEM

Michael Thrusfield*

This paper describes a computerised system for recording and querying case record summaries, which can be used for epidemiological investigations, in the Small Animal Practice Teaching Unit [SAPTU] at the Royal (Dick) School of Veterinary Studies, Edinburgh.

Epidemiologists require information on animal disease morbidity and mortality in either the total population or defined cohorts. The latter is required to elucidate and quantify causal components by measuring the strength of association between determinants and disease, so that effective control strategies can be developed.

Information on several intrinsic factors [e.g. age, sex and breed] is collected routinely in the course of clinical consultations in small animal practice. Data relating to extrinsic factors [e.g. diet] also may be collected if they are relevant to a particular case. These data are usually stored in longhand as part of animals' case notes. Correlation of factors using many longhand records can be a laborious and time-consuming exercise. Edge-notched item cards and centre-punched feature cards (Jolly 1968) have been used to record veterinary data (Griner 1980). However, they are bulky and difficult to handle when multiple correlations from many records are needed.

Computerised data storage and retrieval techniques also have been applied to veterinary case recording [e.g. the Veterinary Medical Data Program (Tjalma *et al.* 1964)]. Many of these projects are based on the 'systems analysis' approach to computing which considers the application [output] to be central. Many are also the successors of the batch job processors of the 1960's which used an 'individual record' structure. These two constraints limit output; programs are written to produce defined outputs; if new outputs [e.g. different determinant correlations] are required then new programs must be written.

*Department of Animal Health, University of Edinburgh, Royal (Dick) School of Veterinary Studies, Veterinary Field Station, Easter Bush, Roslin, Midlothian EH25 9RG, Scotland.

Ideally, information relating to any possibly significant determinants should be recorded. However, it is naive to assume that busy clinicians are willing to record routinely a large amount of data, only justified by the unlikely possibility that they may be of use in future retrospective surveys to test currently unformulated causal hypotheses. A recording system which is of use to the epidemiologist needs to be flexible to allow *ad hoc* correlation of factors. In addition, it must be attractive to the clinician who regularly is providing the information - above all it must be easy to handle and clinically useful. Therefore, a balance needs to be struck between the requirements of the epidemiological purist and of the clinician.

Clinicians' Requirements

SAPTU's clinicians required a record of patients' breed, sex and age. They also wanted a summarised record of data relating to specific periods of consultation, namely:

- diagnoses [suspected or confirmed];
- treatments [satisfactory or unsatisfactory];
- status of an animal at the end of a period of consultation [alive or dead].

There was also a need to record symptoms as well as interpretative diagnoses, which would be of value when a firm diagnosis could not be offered - for example if suitable diagnostic techniques were not available and if unusual syndromes were encountered. Other data requirements included the recording of auxiliary diagnostic tests [e.g. radiography] and broad categories such as 'referral' and 'health check'. The output requirements included the display of individual record summaries and a listing of cases for specified breeds, diagnoses, treatments and ages. Inputting of data needed to be simple with suitable safeguards against errors. Output was required rapidly in a readily identifiable form - preferably in plain English text. A simple English query language was similarly necessary.

THE SYSTEM

The case recording system was implemented in 1980 on an ICL2980 mainframe computer, to which the Teaching Unit is connected via a video display unit [VDU]. This gives clinicians immediate access to the data base and also allows staff members in other departments to interrogate the system via any peripheral terminal. The main reason for choosing a mainframe, however, was the relatively low cost: that of a single VDU. Design, implementation and maintenance costs are notional as part of Edinburgh Regional Computing Centre's service to the University. The system was transferred to a new ICL2988 computer in August 1982, with some minor modifications [*vide infra*].

The Recorded Data

Details of a patient's breed, sex and age [date of birth] are recorded at the beginning of a period of consultation and a six-digit case number is allocated to the record. After a period of consultation the case record is summarised, each summary including a maximum of four diagnoses and/or symptoms, recorded as either suspected or confirmed, and a maximum of two treatments, recorded as either satisfactory or unsatisfactory. Broad categories such as 'referral' and 'vaccination' are recorded as diagnoses. Auxiliary diagnostic techniques are defined as treatments. By convention, a diagnosis is 'confirmed' only when it is supported by auxiliary investigation. Finally, the patient is recorded as either alive or dead.

Inputting Data

The summarised data are converted to code: a two-digit numeric breed code [an 'X' suffix indicating a crossbreed]; a four-digit numeric diagnosis code [the last digit being a check digit]; a three-digit numeric treatment code [the last digit similarly being a check digit] and a two-character alpha sex code. Diagnoses/symptoms are defined as either 'confirmed' or 'suspected', treatments as either 'satisfactory' or 'unsatisfactory' and the animal as either 'alive' or 'dead' by placing an 'X' in the appropriate box on the code form.

A code form is printed onto the longhand case record envelope [23 coded summaries/envelope] for ease of transcription. Lists of the codes with their plain text equivalents are present in all clinic consulting rooms as wall charts, arranged by organ system. Data are input interactively by SAPTU's secretarial staff as lines of predominantly numeric code. A summary is rejected either if an incorrect code is entered or if the summary is inconsistent with previous records relating to the same animal. A consistency check is not attached to 'sex' because this can change, for example when animals are neutered. If cases are reviewed, then records can be modified by deletion and subsequent re-entry.

Querying the Data Base

The data can be flexibly queried in plain text. This is facilitated by the 'data base' structure of the system. The older 'systems analysis' approach to data handling is now being replaced by the 'data base' approach which considers the data as central and the application [output] as peripheral and constantly changing. There are three main data base models (Date 1965): hierarchical, network and relational. In the hierarchical model the various components of a record are viewed as a branching tree with access to certain components [the 'roots'] being gained via more major ones [the 'stems']. In a network, all of the components can be cross-linked as required. The relational model views data in terms of their

relations; this is conceptually closer to the way in which the mind considers information.

The SAPTU system is a network data base [Integrated Database Management System (ICL 1979a,b)].

The record components are stored as *specifier types*:

BREED;
SEX;
MONTH [and year] of consultation;
CONFIRMED diagnosis;
SUSPECTED diagnosis;
AGE at time of consultation;
TREATMENT;
ALIVE [YES or NO].

The specifiers can be queried either alone or in combinations, as categories, using various *connectors*:

AND;
NOT;
OR;
WITH;
THEN.

AND connects specifiers relating to a single summary. NOT excludes specifiers. OR connects specifiers. WITH connects specifiers of more than one case record summary. THEN requires specifiers following it to occur in later summaries than those which preceded it.

Several *commands* can be used:

DISPLAY;
LIST;
LIST BRIEF;
COUNT;
COUNT SIMULTANEOUS;
MAXIMUM;
RANK*.

DISPLAY enables a record to be retrieved by case number and date. LIST prints all records in a given category. LIST BRIEF prints only the case numbers and dates of consultation of records in a category. COUNT generates counts of specifiers and categories. COUNT SIMULTANEOUS builds up counts of various category definitions in relation to a defined category. MAXIMUM determines the most commonly occurring value of a specifier type within a category. RANK creates a list of specifiers in a category in decreasing order of occurrence [zero occurrences are not printed].

Examples: LIST BREED=NOT ALSATIAN CROSS OR ALSATIAN AND SEX=MALE OR

*This command was added in response to requests from clinicians when the system was mounted on the ICL2988 computer.

MALE NEUTER AND AGE \geq 4 AND CONFIRMED=HIP DYSPLASIA END generates a list of case record summaries of all animals which have confirmed diagnoses of hip dysplasia and are either male or male neuters, four years of age or older, but excluding alsatians and alsatian crosses.

RANK CONFIRMED OVER BREED=SIAMESE AND AGE \leq 9MTH AND MONTH \geq 82/01 AND MONTH \leq 82/07 END produces a list and counts of occurrences of confirmed diagnoses in Siamese cats nine months of age or less between January and July 1982, in decreasing order of prevalence. [See Thrusfield and Hinxman (1981) for further examples].

DISCUSSION

The system facilitates the generation of certain epidemiological information including general and specific clinic morbidity rates, and the correlation of diseases, treatments and intrinsic determinants such as age, sex and breed.

There are, however, several disadvantages to the scheme. The use of the term 'confirmed' in relation to a diagnosis implies, erroneously, certainty, whereas the definition more correctly refers to a diagnosis which has been investigated by auxiliary techniques. There is a maximum of four diagnoses [including symptoms] and two treatments per summary. Sometimes cases are presented with a number of symptoms, diagnoses and treatments in excess of this maximum, forcing clinicians to decide subjectively which to include in the summary record and which to exclude from it. The codes are consecutive numbers, with a maximum of 1000 diagnoses and 100 breeds and treatments. These code registers were not filled when the system was implemented but, as clinicians add more codes, the maximum for each register may soon be reached. There is no facility for recording extrinsic determinants such as nutrition, and no facility for recording non-encoded information relevant to some cases such as previous history. It is difficult to add new concepts to a standard computerised record. The clinicians were encouraged to accept a more comprehensive record format but rejected the idea. However, retrospectively, they appreciate the value of a more detailed record.

The success of the system has resulted from its wide acceptance by all of the clinicians, ensuring collection of data relating to all of the clinic's cases, even though the record may be considered epidemiologically deficient in some areas. Several of the reasons for lack of motivation have been obviated. These include onerous methods of coding, the collection of seemingly redundant information and lack of confidentiality [clinicians' names are not recorded].

A common criticism of many data collection projects is that they become an 'end in themselves' with little extraction and analysis of the recorded data. At the time of writing, the SAPTU system has been

fully operational for 15 months, during which time approximately 14,000 case record summaries have accumulated. The clinicians use the data base frequently; detecting sequelae, finding cases of particular personal interest and locating material for teaching. Little epidemiological use has been made of the data base so far. The detection of determinant correlations frequently requires a large number of records, especially if disease prevalence is low and if determinants are not strongly associated with a disease. The data base is still being 'primed' for epidemiological use.

THE FUTURE

The SAPTU data base demonstrates that it is possible to devise a case recording system which is epidemiologically useful and which also is attractive to clinicians. A more widespread data collection scheme, pooling data from many clinics, including general veterinary practices, would have obvious epidemiological merit: it would decrease bias and provide a larger body of data for analysis. Modern computer and communications systems allow the accumulation and transfer of considerable amounts of data. The choice between a single mainframe system and individual microcomputers is a complex one: if it is not necessary to analyse all of the accumulated data simultaneously, then micro systems using identical software and hardware would provide the most economic means of capturing data (Hinxman and Nolan 1981). An essential part of an integrated system is a standardised case record, comprising basic 'core' data (Thrusfield 1981). This would require inclusion of data of varying specificity to accommodate different interests and degrees of diagnostic reliability. The easiest way of achieving this is by coding information hierarchically. Several veterinary hierarchical codes have been devised for diagnoses and treatments [e.g. the *Standard Nomenclature of Veterinary Diseases and Operations* (Priester 1964, 1971)] and symptoms (White and Vellake 1980). None of these is entirely satisfactory to all users: the codes may be too complex (Cordes *et al.* 1981) or may lack specificity (Slocombe 1975) depending upon the uses to which the data are put. The development of a widely acceptable coding system would therefore necessitate a considerable amount of co-operation between all potential users.

REFERENCES

1. Cordes, D.O., Limer, K.L. and McEntee, K. 1981. Data Management for the International Registry of Reproductive Pathology Using SNOMED Coding and Computerization. *Vet. Pathol.* 18: 342-350.
2. Date, C.J. 1975. *An Introduction to Database Systems*. Addison Wesley Publishing Company, Philippines.

3. Griner, L.A. 1980. Storage and Retrieval of Necropsy Records at San Diego Zoo. In: The Comparative Pathology of Zoo Animals (ed. R.J. Montali and G. Miguki), pp.663-667. Smithsonian Institution Press, Washington D.C.
4. Hinxman, A.I. and Nolan, A.D. 1981. Options and costs for a computerised veterinary clinical data base. In: A computerised veterinary clinical data base. Proceedings of a symposium, Edinburgh, 30th September 1981, pp.37-40. Department of Animal Health, Royal (Dick) School of Veterinary Studies, Edinburgh.
5. ICL. 1979a. VME/B IDMS: FORTRAN Interface, RP0044. 2nd ed. Putney.
6. ICL. 1979b. VME/B IDMS, Part 2: Data Base Establishment, 7P6924. 1st ed. Putney.
7. Jolly, J.L. 1968. Data Study, p.33. Weidenfeld and Nicolson, London.
8. Priester, W.A. 1964. Standard Nomenclature of Veterinary Diseases and Operations. USHEW/PHS Nat. Cancer Inst., Bethesda, Md. 1st Ed. revised 1966. 2nd (abridged) ed. 1975.
9. Priester, W.A. 1971. Coding Supplement to Standard Nomenclature of Veterinary Diseases and Operations. USHEW/PHS Nat. Cancer Inst., Bethesda, Md. Revised 1977.
10. Sloccombe, J.O.D. 1975. Surveillance for parasitism in domestic animals. In: Animal Disease Monitoring (ed. D.G. Ingram, W.F. Mitchell and S.W. Martin) pp.129-135. C.C. Thomas, Springfield.
11. Thrusfield, M.V. 1981. Proposals for an integrated veterinary clinical data base. In: A computerised veterinary clinical data base. Proceedings of a symposium, Edinburgh, 30th September 1981, pp.41-47. Department of Animal Health, Royal (Dick) School of Veterinary Studies, Edinburgh.
12. Thrusfield, M.V. and Hinxman, A.I. 1981. A computerized system for storing and querying clinical case record summaries. J. small Anim. Pract. 22: 669-679.
13. Tjalma, R.A., Priester, W.A., Adelstein, E.H. and Zwaag, R.V. 1964. Clinical Record Systems and Data Retrieval Function in Veterinary Medicine. A proposal for systematic data programming. J. Am. vet. med. Ass. 145: 1189-1197.
14. White, M.E. and Vellake, E. 1980. A coding system for veterinary clinical signs. Cornell Vet. 70: 160-182.