

A computerised data base for the diagnosis of rare dysmorphic syndromes

R M WINTER*, M BARAITSER†, AND J M DOUGLAS‡

From *the Division of Inherited Metabolic Diseases, MRC Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, and Kennedy Galton Centre, Harperbury Hospital, Harper Lane, Radlett, Herts WD7 9HQ; and †the Departments of Clinical Genetics and Paediatric Cardiology, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.

SUMMARY A system is described for the computerised storage and retrieval of information on rare dysmorphic syndromes. The clinician can ask a microcomputer for a list of syndromes with any logical combination of physical abnormalities. A descriptive title and full references are also provided on request. Similar systems would be applicable to other medical specialties.

Every medical practitioner occasionally encounters a patient who has a rare syndrome. Diagnosis entails going beyond the standard medical text books and exploring the medical literature, often spanning many decades and hundreds of journals. Systematic aids, such as *Index Medicus*, are available to help, but because articles are only loosely classified under general key words, the process can be extremely laborious. In certain medical specialties, such as clinical genetics, the practitioner is constantly being called upon to diagnose such patients and a substantial amount of time is spent searching the literature.

Seven in 1000 babies are born with multiple malformations.¹ Many of these children will have a malformation syndrome that has been recognised and described in the medical literature, albeit in a single paper. All malformation syndromes are individually rare, some even extremely so, but the correct diagnosis and genetic implications are of great importance to individual families. The average clinician will not be able to remember or recognise the majority of these conditions and even an expert will find difficulty in recalling all of them. A system is described to trace reports of syndromes in the medical literature and to register undiagnosed, unreported cases in such a way that they can be compared with each other and recognised as a new syndrome.

Methods

In designing a computerised system for the retrieval of data several factors have to be taken into account.

Received for publication 3 August 1983.

Accepted for publication 26 September 1983.

These will be enumerated in full, as they are applicable to any specialty.

HARDWARE

The type of computer facility used will depend upon the amount of data that need to be stored, the potential availability of the system, the ease of operation, access, and cost. Where a large amount of data are to be stored and retrieval needs to be rapid (in seconds), a mainframe computer is ideal. However, access for remote users will have to be through a telephone link which can be slow, and frequent use could prove expensive. A microcomputer provides immediate access and data can be transferred between different machines. On the other hand, the amount of data storage is theoretically less and retrieval times are slower. A microcomputer incorporating 1.6 megabytes of data storage on floppy discs was chosen for the present study. The storage space is enough for approximately 750 syndromes and 1500 references, together with retrieval programmes. Extension beyond this amount of data will necessitate the addition of a relatively inexpensive hard disc allowing at least 10 megabytes of storage.

SOFTWARE

Commercial programmes for retrieval and storage of literature references, indexed under 'Key words', are available (for example, Cardbox²). However, the indexing may not be flexible enough for the precise needs of an individual specialty. The alternatives are either to write the programmes from scratch in one of the standard programming languages, such as

Basic, or to adapt a commercially available data base management system. A general data base management system was chosen (dBase II³). This incorporates a high level programming 'language', so that data entry and retrieval programmes can be tailor-made to fit individual needs.

RETRIEVAL STRATEGIES

Syndromes must be indexed under a master list of Key abnormalities. Theoretically, the relative importance of individual abnormalities to the diagnosis of a specific syndrome can be incorporated by assigning a weighted score. In practice, many syndromes are so rare that it is impossible to define which features are most important. It is often useful to have the master list of possible abnormalities subgrouped into different levels. For example, cleft palate might be grouped under disorders of the palate, which in turn could be grouped under oral abnormalities. In this way a 'tight' or 'loose' search can be made for all syndromes with cleft palate or alternatively for all syndromes with palate abnormalities. If no weights are used, the data base can be searched by entering logical combinations of abnormal features, for example (A or B) and C, when all syndromes meeting these criteria will be retrieved.

If physical abnormalities are given weights, then all abnormalities of a specific case must be entered, and the retrieval programme produces a ranked 'score' for each possible syndrome diagnosis.

Because most syndromes of interest are so rare, no weights are used in the present design. The data base was compiled using a master list of 1176 physical features. Each feature is given a three level code, for example, 12.4.1 represents cleft palate. In this sequence 12 represents oral cavity and 4 all conditions affecting the palate. The last number represents clefting of the palate.

Genetic and paediatric journals over the past 15 years were systematically searched for reports of multiple malformation syndromes. Chromosomal abnormalities were excluded. A mnemonic was created for each syndrome and the features were entered using the code. The system allows up to 28 features to be entered per syndrome. A full descriptive title and references were also entered. The data were then entered into a main frame computer, edited, and transferred on to a floppy disc for use in a microcomputer. A retrieval programme was written using the dBase II system. To date there are over 700 syndromes in the data base and approximately 1500 references.

Example

A family was referred where a father and two children

had a striking combination of physical abnormalities. All affected persons had duplication of the big toe (pre-axial polydactyly) together with a very small, supernumerary digit on the ulnar border of the hands (postaxial polydactyly). In addition the father had a large head, as did one child, and all affected persons had unusual shaped heads, consisting of a tall forehead and mild brachycephaly. Because the digital malformations were variable in different family members the computer was asked for all syndromes with pre- or postaxial polydactyly together with an unusual head shape or size: (23.4.1 or 23.4.2) and 3.1. Note that the codes in the brackets indicate pre- and postaxial polydactyly and 3.1 indicates general abnormalities of shape or size of the head.

The following list of syndrome mnemonics was produced:

C-Trigonoceph
Carpenter
Fitch (1976)
Greig
Meckel-G
Robinow-Sorauf
Schinzel (1980)
Smith-L-O.

If necessary, full titles could have been produced, for example, C-Trigonoceph represents the Opitz trigonocephaly C syndrome. Of these eight syndromes, five were well known to the authors and had additional features which did not fit the diagnosis. References were requested from the computer for the syndromes which were not well known, Fitch,⁴ Schinzel,⁵ and Greig.⁶ Consultation of the references indicated that Greig syndrome was the correct diagnosis.

Conclusions

The system is easy to update and two clinicians spending a couple of hours a week can easily achieve this. In addition to a literature data base the same system will store cases with unknown dysmorphic syndromes from the genetic clinic. Other systems have been devised to answer similar questions. The Birth Defects information service from the USA (Tufts)⁷ provides on-line information about dysmorphic syndromes together with a 'diagnostic assist' facility. The system provides a differential diagnosis of potential candidates with an indication of how many significant features were not present and how good the match was for each case.

In the system presented in this report (now called the 'London Dysmorphology data base'), a 'match' is not thought to be appropriate. The computer programme is not designed primarily to make a diagnosis for the clinician, as it is uncertain whether the computer can perform this function any more successfully than the experienced clinician. It will, however, provide for clinicians a means of recalling a manageable list of forgotten, often high risk, genetic syndromes and it will quickly tell them whether they have seen the condition before. Retrieval is 'flexible' in that different search strategies can be tried. The system also has the advantage of being applicable to other branches of medicine provided that a list of signs and symptoms relevant to, say, neurological disorders is drawn up and the literature or previous clinical experience entered. In this way present uncertainty may be partially avoided by using past experience.

References

- ¹ Smith DW. *Recognizable patterns of human malformation*. 3rd ed. Philadelphia: Saunders, 1982.
- ² Cardbox. Caxton Software Publishing, 1982.
- ³ dBase II. Ashton-Tate, 1982.
- ⁴ Fitch N, Jecquier S, Papageorgiou A. A familial syndrome of cranial, facial, oral and limb anomalies. *Clin Genet* 1976;**10**:226-31.
- ⁵ Schinzel A, Schmid W. Hallux duplication, post axial polydactyly, absence of corpus callosum, severe mental retardation, and additional anomalies in two unrelated patients: a new syndrome. *Am J Med Genet* 1980;**6**:241-9.
- ⁶ Fryns JP, Van Noyen G, Van den Berghe H. The Greig polysyndactyly craniofacial dysmorphism syndrome: variable expression in a family. *Eur J Pediatr* 1981;**136**: 217-20.
- ⁷ Buyse M. Centre for birth defects information services. *Birth Defects* 1980; **XVI** (5): 83-91.

Correspondence and requests for reprints to Dr R M Winter, Kennedy-Galton Centre for Clinical Genetics, Harperbury Hospital, Harper Lane, Shenley, Radlett, Herts WD7 9HQ.