SYNDROC: microcomputer based differential diagnosis of malformation patterns

D SCHORDERET AND P AEBISCHER

Department of Paediatrics and Genetics, University of Geneva Faculty of Medecine and the Dialysis Center, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

SUMMARY The differential diagnosis of human malformation patterns is difficult because of their great number and low occurrence. To compensate for deficiencies in experience, the authors have developed a microcomputer based differential diagnosis for dysmorphic syndromes.

Written in PL/I-80 and using a MDBS III data base containing more than 700 signs and 300 syndromes, the algorithm of SYNDROC, the program's name, is built on three different approaches. An initial differential diagnosis is given by a pseudo-Bayesian method. Its refinement is possible through an heuristic approach, and undiagnosed cases are stored for subsequent Boolean retrieval, thus permitting the recognition of new dysmorphic disorders.

The reliability of the program was tested in a retrospective study based on 100 dysmorphic cases. The results showed a 95% agreement between the first two diagnoses suggested by SYNDROC and the diagnosis made by experienced geneticists.

The diagnosis of childhood multiple malformation syndromes represents an intricate problem for the paediatrician. On one hand, the difficulty rests in the multiplicity of the syndromes (a few thousand), and, on the other, in their relatively rare occurrence. With the exception of a few syndromes such as trisomies of chromosomes 13, 18, and 21 or the Marfan, Klinefelter, and Turner syndromes, the probability of a paediatrician encountering these syndromes is low.

We believed that a program containing the different known malformation syndromes and their characteristic signs would be able to compensate, to a certain extent, for deficiencies in experience. We developed a program to help in the differential diagnosis of malformation syndromes: this program, named SYNDROC, runs on a microcomputer. The first part of this article describes the structure and the functioning of the program and the second gives results of a retrospective study of 100 cases of malformation syndromes.

Theoretical aspect

The elaboration of a differential diagnosis does not follow accurate rules. A few theories have been offered but, up to now, no unifying concept has become apparent.[1][2] As a description of all the algorithms would be out of the scope of this article, we will limit ourselves to a short description of those included in SYNDROC.

Descriptive method. The aim is to take maximum advantage of data from patients in hospital.[3][4] Practically, this algorithm can be summarised as follows: the clinical history and physical examination, the diagnosis, and the outcome of each case are stored in a computer. The signs and symptoms of each new patient are compared with previously accumulated patient data by a Boolean operation called the 'and' operation, which consists of finding, given two groups A and B, a subgroup C whose members belong simultaneously to both A and B (the intersection of A and B). The diagnosis can then be derived by reference to the most closely matched subgroup. The advantage of a system like this is its self-learning capacity—each new case described improves the capacities of the program. It is not necessary to know why the signs and symptoms are related in the way they are, but only that their association is reasonably stable.

The disadvantage of the program, however, is its unwieldiness; for optimal efficiency it requires standardisation of the charts, constant updating of the data base, and the acquisition of a large main

frame computer with associated high purchasing and servicing expenses.

Bayesian algorithm. This algorithm has the advantage of being easily programmed as it is based on an equation. Bayes' rule states that the probability P(D/F) that a patient has a disease, given the presence of a particular set of findings, is equal to the probability P(F/D) of the occurrence of this set of findings given the disease, multiplied by the frequency P(D) of that disease in the population and divided by the incidence P(F) of the findings in the general population.[5]

$$P(D/F) = \frac{P(F/D) \times P(D)}{P(F)}$$

This very tempting theoretical algorithm is rarely usable in its original form as most of these parameters are only partially known. In addition, for the equation to be totally accurate, the different signs must be independent, the diagnoses mutually exclusive, and the various probabilities stationary over time. Few situations in medicine meet these criteria. Some approaches partially taking these criteria into account are, however, usable in very limited fields. They are then referred to as pseudo-Bayesian.

Heuristic algorithm. Eddy and Clanton[6] and others[7] have tried to analyse the psychological process by which clinicians solve diagnostic problems. The first step is the collection of all the signs and symptoms of the studied case. Generally their number is too high to allow mathematical processing. Thus, the physician reduces this number by combining sets of findings into aggregate findings (that is, fever + chills + leucocytosis + high erythrocyte sedimentation rate = infection). The number of the aggregates formed depends on the problem's complexity. In a second step, the clinician chooses signs with a high meaning (pivot sign). After selecting one or several pivot signs, he formulates a list of diagnoses. Such a list is possible since the number of diagnoses relating to a pivot sign is limited and since medical education takes this approach. Once this initial differential diagnosis is established, the physician proceeds to a retrograde analysis and tries to explain, for each diagnosis considered, the maximum number of aggregates. The higher the number of aggregates explained by a particular diagnosis, the more probable it becomes. This approach, made of progressive steps, is actually very popular among diagnostic theoreticians.[8]

SYNDROC's algorithm

The differential diagnosis established by SYN-DROC is based on an algorithm which includes, to different degrees, the pseudo-Bayesian and the heuristic approaches and the descriptive method. A first differential diagnosis is computerised with the help of the pseudo-Bayesian approach; by this method each sign is related to a diagnosis by a three-dimensional matrix M(m1,m2,m3). M1 gives the subjective importance of the sign in connection with a given diagnosis. This is assessed on a scale varying from 1 to 100. M2 reflects the frequency of the signs used to describe the syndromes and thus allows a modulation of m1. Theoretically, this frequency should be related to that of the general population but, since this frequency is not well known, m2 was calculated in relation to all the signs included in SYNDROC. M3 represents the frequency of the given diagnosis in the general population. When this frequency is known, m3 is used to rank two or more diagnoses that would have an identical final score. The refinement of the diagnosis is based on an heuristic approach. In this approach, SYNDROC establishes one or many pivot signs. These pivots are determined with the help of the matrix M and represent the signs that have a maximal m1 value (m1=100). Practically, SYNDROC asks for the presence or absence of each pivot sign (if they were not introduced earlier) for every considered diagnosis. This approach allows an improved discrimination between the differential diagnosis already established by the pseudo-Bayesian method. Finally, the descriptive method allows the storage, on a magnetic support (floppy disks), of all profiles studied but not recognised by SYNDROC as having a definite diagnosis. In a case not leading to a diagnosis, SYNDROC will compare this case with all the previous undiagnosed profiles using the Boolean operation 'and'. This way of proceeding should theoretically result in the recognition of new dysmorphic entities.

SYNDROC's initial data base was formed with the help of the profiles of more than 300 diagnoses selected from the specialised literature[9] [10] and original articles. The description of these syndromes relies on more than 700 signs.

SYNDROC is easy to use and computer experience is not necessary. On a practical basis, a complete history and detailed physical examination must be obtained for each patient. The user is then asked to introduce the different signs into the microcomputer by typing them on a visual display unit. A separate alphabetic list of all the signs included is supplied so that the physician can check for correct spelling. (Such a list is available on request.) Once all the signs are introduced, SYN-DROC produces an initial differential diagnosis classified in order of decreasing probability. Each

diagnosis is listed with a certainty coefficient. This coefficient represents a calculated estimation of the probability of the diagnosis being the correct one. It is based on the scores resulting from the different approaches and varies from 0 (excluded diagnosis) to 1 (certain diagnosis). The diagnoses are listed only if their certainty coefficient is between certain values, the upper boundary being the highest certainty coefficient and the lower being the half of this value. Thus, SYNDROC's differential diagnosis can be reduced to a unique diagnosis (the second one would have a certainty coefficient much lower than the first one), a list of possible diagnoses ranked by their certainty coefficients or no diagnosis if the certainty coefficient of the most probable diagnosis is not more than 0·25.

SYNDROC allows a refinement of the first proposed differential diagnosis. With a certain number of questions, SYNDROC tests the presence or absence of the pivot signs which could have been forgotten by the user. During this refinement phase, diagnoses previously unlisted (because of a too low certainty coefficient) can be selected in a second stage.

Hardware

The computer used is a 8 bit microcomputer (Ithaca InterSystems USA) which has a 256K memory and two, double sided, double density floppy disks. It runs under CP/M (Digital Research) and costs $5000 (≈£4237). SYNDROC is written in PL/I-80 (Digital Research) and uses a MDBS III data base (Micro Data Base Systems, USA).

Method

SYNDROC was tested using records from 100 children who were either inpatients in the Geneva University Hospital or had been evaluated at the Medical Genetics Institute of the University of Geneva between 1975 and 1983. Only those in whom a diagnosis belonging to the known syndromes of the data base had been established after a careful physical examination were eligible for the study. Records from the Medical Genetics Institute were chosen in alphabetic order and those of the Geneva University Hospital in chronologic order. The potential occurrence of each sign specified by the program was checked for in the geneticist's written report by one of the authors (PA) (syndromes in SYNDROC were originally coded by DS) and then entered into the microcomputer. After their introduction, the program was stopped and the differential diagnosis was given by the microcomputer. The refinement phase of SYNDROC's algor-

ithm was not used because of the impossibility of answering the program's questions.

Results

The cases analysed are a good reflection of the frequency of these syndromes in children. The 100 cases tested represent 40 different syndromes. Twenty syndromes occurred more than once (recurring syndromes). Table 1 shows their distribution.

The time required by the microcomputer to display the differential diagnosis depends on the number of signs, but never exceeded three minutes.

In 33 cases, SYNDROC submitted only one diagnosis, meaning that a differential diagnosis was unnecessary. In each case there was complete agreement with the clinically established diagnosis. In the other 67 cases, SYNDROC submitted a list of diagnoses classified according to decreasing probability, with a mean of 3·37 syndromes and 8·41 signs for each case. In 33 cases, the agreed diagnosis was listed first; in 29 cases it was in second position (Table 2). In the last five cases, the concordant diagnosis was mentioned twice in third position and once in the fifth. In two cases, SYNDROC did not mention the diagnosis established by the geneticists. One of these was a case of Fanconi pancytopenia, the other an example of the Weill-Marchesani syndrome. A separate analysis of the recurring and non-recurring syndromes leads to 95% diagnostic agreement in first or second position in both groups.

Table 1   *List of the tested diagnoses and their occurrence in the study*

Recurring syndromes				
14 x	*8 x*	*5 x*	*4 x*	*2 x*
Trisomy 21	Marfan	Klinefelter	Noonan	Beckwith-Wiedemann
	Trisomy 18	Rubinstein-Taybi	Vrolik	Carpenter
		Trisomy 13		Cockayne
		Turner		de Lange
				Ehlers-Danlos
				Fanconi thrombocytopenia
				Holt-Oram
				Saethre-Chotzen
				Smith-Lemli-Opitz
				Sotos
				5p-

Non-recurring syndromes (1x)		
Acromesomelic dysplasia	Hallermann-Streiff	Sjogren-Marinesco
Bloom	Kartagener	Thanatophoric dysplasia
Cat-eye	Laurence-Moon-Biedl	Treacher-Collins
Crouzon	Lowe	Weill-Marchesani
Ellis-van Crefeld	Meckel-Gruber	XXXXY
Fetal alcohol	Moebius	Zellweger
Fetal rubella	Seckel	

Table 2 *Position and number of concordant diagnoses proposed by SYNDROC*

Position	Total syndromes		Recurring syndromes		Non-recurring syndromes	
	No	*(%)*	*No*	*(%)*	*No*	*(%)*
1 (alone)	66 (33)	(66) (33)	55 (27)	(69) (34)	11 (6)	(55) (30)
2	29	(29)	21	(26)	8	(40)
3 and more	3	(3)	3	(4)	0	(0)
Not listed (false)	2	(2)	1	(1)	1	(4)
No	100		80		20	

Discussion

The first question arising from any computer based differential diagnosis concerns its reliability. The initial results showed that there was agreement between the two first diagnoses of SYNDROC and the clinical diagnosis in 95% of the analysed cases. The reliability was maximal when the program listed a unique diagnosis (that is in 33% of the cases) and this always agreed with that of the geneticist. In 67% of the cases, however, SYNDROC proposed a differential diagnosis, reflecting the difficulty of choosing a unique diagnosis. In half of these cases, there was agreement between the first diagnosis listed by the microcomputer and that of the geneticist. In 66% of the cases, therefore, there was agreement between SYNDROC's first diagnosis and the clinical one. In 29 cases, the geneticist's diagnosis was listed in second place in SYNDROC's proposed diagnostic list. Half of these cases can be explained by a relatively incorrect weighting of a certain number of diagnoses, especially those defined by a large number of signs. For example, the syndrome 4p-, defined in SYNDROC by more than 40 signs, occurred six times in first position but was never mentioned by the geneticist. It would be desirable therefore to introduce into SYNDROC a modification which would take into account the number of signs defining a syndrome. The last cases, in which the diagnosis differed, may be explained by the use of a retrospective protocol which does not allow the gathering of all the signs, and also by the difficulty of establishing, for a certain number of disorders, a definite diagnosis because of an important overlap of signs. The high rate of agreement between SYNDROC's and the geneticist's diagnosis could be a consequence of the relatively high frequency of recurring syndromes. This is not the case. The separate analysis of the score made by SYNDROC with the recurring and non-recurring syndromes yielded the same results: with 95% agreement.

As opposed to other diagnostic aid programs,[11]

SYNDROC gives not only an exhaustive list of possible diagnoses, but it attempts to propose, by different approaches, the best diagnosis. To our knowledge, SYNDROC is the first program running on a microcomputer which gives a ranked diagnosis of human malformation syndromes.

SYNDROC's attraction lies not only in the proposal of a differential diagnosis, but also as a teaching tool, through the constant repetition of signs and symptoms of malformation syndromes.

We realise that the number of malformation syndromes included in our system is rather limited, but this will increase with further development. In the meantime, the present study shows that microcomputers have a definite place in medical teaching and practice, and that their use should be encouraged.

References

1 Wardle A, Wardle L. Computer aided diagnosis—a review of research. *Methods Inf Med* 1978;**17**:15–28.
2 Wagner G, Tautu P, Wolber U. Problems of medical diagnosis—a bibliography. *Methods Inf Med* 1978;**17**:55–74.
3 McMullin E. Diagnosis by computer. *J Med Philos* 1983;**8**:5–27.
4 Barnett GO. The application of computer-based medical-record systems in ambulatory practice. *N Engl J Med* 1984;**310**:1643–50.
5 Brown GW. Bayes' formula. *Am J Dis Child* 1981;**135**:1125–9.
6 Eddy DM, Clanton CH. The art of diagnosis. *N Engl J Med* 1982;**306**:1263–8.
7 Bergman DA, Pantell RH. The art and science of medical decision making. *J Pediatr* 1984;**104**:649–56.
8 Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medecine. *N Engl J Med* 1982;**307**:468–76.
9 Smith DW. *Recognizable patterns of human malformation.* 3rd ed. Philadelphia: WB Saunders, 1982.
10 Bergsma D. *Birth defects compendium,* 2nd ed. New York: The National Foundation—March of Dimes, Alan R Liss, 1979.
11 Winter RM, Baraitser M, Douglas JM. A computerized data base for the diagnosis of rare dysmorphic syndromes. *J Med Genet* 1984;**21**:121–3.

Correspondence to Dr D Schorderet, Department of Paediatrics and Genetics, University Hospital CH–1211 Geneva, Switzerland.