

INDIVIDUAL STUDY

Computer analysis of the complete medical record, including symptoms and treatment

R. A. JOHNSON, B.A., M.B., B.Chir., D.P.M.
Moseley, Lancashire

REASONS for the slow progress of medical computing were examined by Ockenden and Bodenham in 1970. Of the problems they identified (*Lancet*, 1970), two have proved most troublesome, namely: the definition of the area of work, and the establishment of a medical data bank. Difficulties in "definition" may be seen in the "small scale experiments" they recommended. In some, only the diagnosis was recorded (Dinwoodie, 1970), in others treatment was added, but portions of the record were not (Greur). In one of the few studies to include symptoms (Morrell *et al.*, 1971), coding was limited to 25 per cent of consultations. A recent symposium (British Computer Society, 1972) confirmed the continuing impact of these obstacles.

A novel solution to both problems, is now proposed. With conventional paper records, the clinician "defines" what is important, by recording it longhand. With the computer records used here, the coding system has been designed for such ease of use, that he may include as much or as little as he judges to be clinically significant in the ordinary way. "Definition" is therefore determined by current practice.

Establishing a medical data bank is less straightforward. Many of the present expensive failures in the computer industry stem from difficulties in retrieving information from such data banks. With this in mind, an exclusively numerical code was developed. In consequence, results that would otherwise be prohibitively expensive in software costs, may be obtained from programmes designed by a full-time general practitioner.

Method

Since the problem of 'definition' required that the whole record be coded, it was essential that the coding system be quicker than, and at least as accurate and detailed as its longhand counterpart. Whereas the *International classification of diseases* (ICD), and the related 'E-Book' codes of the Royal College of General Practitioners are recognised standards; they are both ordinal in character, and explicitly diagnostic in orientation.

They were not designed to cope with the variety of symptoms that arise in medical practice, nor specifically for use in the presence of the patient. For these reasons they were discarded in favour of a more pragmatic basis, namely the frequency with which each symptom had occurred during a preliminary survey of six months general practice.

The opportunity was also taken to redesign the data sheet used. A faceted code (Vickery, 1960) (as opposed to an ordinal one) proved valuable, since it allowed 300 simple codes, to cover up to one million distinct symptoms.

As regards the data bank, a simple illustration will show that a numerical code is mandatory. If the symptom of "pain", for example, is to be entered, it may be punched on a computer card using the symbols "space, P, A, I, N, space". Though this is clear enough to those who speak English, it is relatively expensive in computer terms, since it uses six bytes, and yet contains no details of the whereabouts, severity or duration of the pain. Even mild variations in the same word cause difficulties. Thus: "space, space, P, A,

I, N", "P, A, I, N, S", or "P, A, I, N, F, U, L" have nothing whatever in common, so far as mechanical data processing is concerned. Attempts to provide software to cope with these linguistic details, can be costly.

Use of a numerical code on the other hand, allows the numeral "1" (when inserted in a prearranged place on the data sheet) to stand for "pain", which was the most frequent symptom. In the present study, an 80-column computer card was used, and the code for "pain" was entered in card column (cc) 40. Thus every time this symptom presented, the digit "1" was written in this column.

A second column (cc41) was used to indicate the severity of the pain, on a three-point scale. A third column (cc42) carried information as to the location of the pain, numbering from the head downwards: references to "head" were coded "1", to chest "2", arm "3", abdomen and back "4", and leg "5", all in card column 42.

Almost unlimited detail is available by virtue of the faceted code. With 10^6 combinations, "upper, outer quadrant of the right breast", or "medial border of the left little finger" may be coded with ease. A full description of the construction and use of the coding system is available (Johnson, 1972), but with even the six simple codes mentioned, it is possible to record the presence of pain in any region of the body.

In fact owing to the logical structure of the code, only a further ten similar codes are required before all clinical data may be recorded in compatible form, albeit in general terms. With an average list size, the costs of card punching would be £200-300 a year, for which assistance should be available through ECN 606.

The errors inherent in the method are analogous to those afflicting longhand records. A more serious criticism is that of motivation: it has been suggested (Pratt, 1971) that doctors are unlikely to encumber themselves with a numerical coding of any kind.

However, while a clinician may resist the idea of writing a "1" and a "4" in the appropriate places, instead of his more familiar "A, B, D, O, M, I, N, A, L, space, P, A, I, N": it does not necessarily follow that he will continue to do so, in the face of superior medical data processing. After all, it is the clinician who stands to gain most by coding the patient's record numerically, since all previous references to any pain could be instantly available to him. By comparison, the retrieval of the same information from paper-based records is pedestrian to the point of being hazardous.

Errors will inevitably arise, with respect to the meaning allotted to each code. However, even within the same language, such semantic errors are already causing growing confusion, owing to different regional and eponymic usage (*Journal of the American Medical Association*, 1971). The situation will deteriorate as the context in which the clinician practises becomes increasingly international, both in Europe, Asia and elsewhere. The use of a numerical code can actually reduce these errors, by providing a standardized format of internationally-accepted arabic numerals. The following results are subject to a small margin of error arising from minor programming discrepancies. Developments are in hand, for example, to enable the data file to be accessed by each sub-record, rather than card by card. Analysis by the sex of the patient should also be available for all phases of the programme, soon.

Results

The study took place in a three-man practice of almost 10,000 patients in South Lancashire. Table I gives the age-sex distribution of the patients recorded by one practitioner, from 1 October, 1970 to 30 September, 1971. The age groups essentially follow the *International Classification of Disease*. (W.H.O., 1967).

TABLE I
ATTENDANCES BY AGE AND SEX

	Total	Age groups in years			
		Under 14	15-44 yrs	45-64 yrs	Over 65
Males	4483 (1797)	1291 (501)	1474 (691)	1045 (320)	673 (285)
Females	6882 (2399)	1297 (461)	2966 (1045)	1527 (476)	1092 (417)
Totals	11365 (4196)	2588 (962)	4440 (1736)	2572 (796)	1765 (702)

(Numbers of individuals given in brackets)

The mean rate of attendance was 2.7, with a maximum of 51 for one woman. Symptoms averaged 2.8 per contact, with a total of 12 on three occasions. In the 12 months, there was a mean of 7.4 symptoms per patient, though one man had 71, and one woman 174. Of the 11,365 consultations, 74 per cent were in the surgery, 18 per cent visits, three per cent by telephone, and five per cent by others attending for the patient. (3,124 people accompanied the patient in the consulting room.)

"Pain" was the most frequent symptom, coded 3,517 times (11 per cent) out of 31,110 in all. "Cough" was next, at nine per cent. The commonest anatomical reference was to head, at 1,698, 5.5 per cent of all symptoms. Twenty nine per cent (3,333) of the 12,646 diagnoses were "follow up", with respiratory disease at 2,782 (22 per cent).

The commonest drugs were in the psychotropic and analgesic group, 3,440 or 18.9 per cent of the 18,243 treatments. Antibiotics were next at 3,393 (18.6 per cent). One thousand and twenty National Insurance certificates were issued in the year, of which 581 were for seven days.

The interrelations between symptoms, diagnoses and treatments were tabulated in a printout of 30 tables of 12 columns by 100 rows each, involving 200,000 counting operations. (Only the most frequent diagnoses and treatments were analysed in detail.) "Pain" was analysed in terms of its severity, location, duration, associated symptoms, diagnoses, and treatments. Table II shows both an increase, and a change of distribution with age.

TABLE II
INCIDENCE OF PAIN, BY AGE AND ANATOMY

Anatomy	Total	Age groups in years					Aches	Severe pains
		Under 1	1-14	15-44	45-64	Over 64		
Head	776	9	200	312	167	88	220	15
Chest	552	1	44	260	162	85	120	30
Arm	167	0	12	63	64	28	62	2
Abdomen and back	790	3	135	405	151	96	116	21
Leg	375	1	36	129	121	88	128	3
Total	2660	14	427	1169	665	385	646	71

Of the 242 occasions on which pain was recorded as "getting worse, or relapsing" (compared with 57 on which it was recorded as "improving"), 115 referred to abdominal pain. The distribution of recorded duration was also skewed. Abdominal pains were the

most urgent, half the references were to those lasting two days or less. With head and chest pains, the figure was three days or less. Arm pains however, showed a reverse situation: the majority being one week or longer. Leg pains showed a composite picture, a third recorded as being present for three days or less, whilst almost two thirds were for a week or more.

Respiratory disease was as common a diagnosis as orthopaedic and skin disorders (32 per cent each).

Of the 3,438 treatments for these painful conditions, the majority predictably, were for the psychotropic and analgesic drugs (924 or 27 per cent), used mostly for abdominal and back pains. There were 791 antibiotics, used most frequently with pains in the head. Of the 924 psychotropic drugs, 583 were analgesics, of which 371 were dextropropoxyphene ('Distalgesic'). Diazepam and other tranquillizers were advised on 171 occasions, 73 for head pain and 46 for abdominal pain. Antidepressants were associated mainly with head pains (25 out of 51). Arm pains were half as frequent as leg pains, with almost 50 per cent in the shoulder. Conversely, most leg pains were in the knee. Forty per cent of all abdomen and back pains were in the back (with 51 "worse or relapsing"), while a further 22 per cent related to the lower abdomen (with 19 "worse").

Analysis of the diagnosis "follow up", or "no change from previously recorded diagnosis" is of interest, especially in the oldest age group, since it comprises principally, attendances for repeat prescriptions. Table III gives the symptoms and treatments associated with this diagnosis.

TABLE III
SYMPTOMS AND TREATMENTS ASSOCIATED WITH THE DIAGNOSIS "FOLLOW UP"

Symptoms	Age groups in years						Treatments
	Total	Under 1	1-14	15-44	45-64	Over 64	
General	147	4	6	51	61	25	364 antibiotic
Respiratory	402	27	52	74	138	111	261 respiratory
Mental	431	10	17	131	163	110	1222 mental
Gut	410	50	65	92	82	121	280 gut
Skin & trauma	262	10	35	63	82	72	288 local & orthopaedic
Genito-urinary	279	1	7	243	11	17	157 genito-urinary
Ears, eyes & metabolic	52	1	1	10	24	16	326 ears, eyes & metabolic
Cardiovascular & neurological	60	0	5	16	13	26	418 cardiovascular & neurological
Miscellany	557	3	25	176	199	154	739 misc. & certificates
Totals	2600	106	213	856	773	652	4055
Attendances	3333	75	242	1137	996	883	
as % of all attendances	29%	12%	26%	39%	50%		

"Getting better" was recorded on 1,034 occasions under this diagnosis (out of 1,609 for the entire year), compared with 585 "the same", and 310 "worse". The 1,222 psychotropic

drugs consisted of 374 hypnotics, 390 tranquillizers (of which 145 were diazepam), 263 analgesics (125 of dextropropoxyphene), and 193 antidepressants. Ampicillin (97 times), tetracycline with novobiocin (109 times) and co-trimoxazole ('Septrin') (82 times): all outnumber oxytetracycline (42 times), indicating that these drugs are 'second choice' antibiotics, compared with the latter (which was prescribed 1,011 times overall). Of the 1,262 drugs for the over 65-age group, 462 were psychotropic, and 201 cardiovascular or neurological.

Referrals

The twelfth most frequent treatment in the year, was "referral to hospital (outpatients)". Nevertheless, in view of its wider significance, it may be noted that at an annual frequency of 359, it represented two per cent of all treatments, and occurred on average every thirtieth attendance.

There were 70 hospital admissions, of which 42 were recorded as urgent. These are omitted from this analysis, as are 33 referrals to casualty, and others for investigations.

The predominant symptoms among those referred to hospital were genito-urinary (32 per cent of symptoms coded). These were mainly in women, in the 15-44 age group. The next most frequent symptom code was "musculo-skeletal and skin", at 17 per cent. On 45 occasions "deterioration" was coded, as opposed to 39 "no change", and 18 "improving". The duration recorded for these symptoms was heavily skewed in favour of the longer term: over half were more than two weeks. Pain was not a significant feature.

Discussion

These results indicate the wide-ranging analysis that becomes available by use of a computer-compatible code. Were more clinicians persuaded to record their daily work in this, or a similar numerical code, the results would be impressive. The implications for medical practice even from the present study, are profound. By recording symptoms and treatments in a mutually compatible code, it becomes possible to analyse the treatments prescribed in terms of the symptoms which evoked them. Not only may therapeutic intent be quantified, but also therapeutic efficacy.

Improved medical data processing is already bringing similar benefits to research. South and Rhodes (1971), make the point that their coded data ". . . allows a detailed retrospective inquiry to be made of groups not previously selected". With a comprehensive coding system, all clinical practice would become available for research, and items noted by the observant may later be seen on "retrospective inquiry" to have a relevance that was quite unknowable at the time of recording (*vide* thalidomide). The restrictions and potential ethical hazards of the double-blind system might thus be obviated.

Iatrogenic disease in particular, can only be detected and prevented by improved analysis and dissemination of medical information, which must relate to the individual patient and be readily available at the point of prescribing.

The paper-based recording system is fast becoming a liability in this respect: the more detail that is retained, the more difficult is it physically to retrieve useful information (Johnson, 1971). Persistence of longhand records in paper data-banks necessarily implies iatrogenic disease of increasingly epidemic proportions. Once the clinician concedes a change in his method of recording however (from alphabetic to numerical), an electronic system could be established whose validity and value would grow with each additional detail.

Given the increasing power of modern drugs, and the accelerating rate of change of knowledge (Macrae, 1972), an electronic device that is able mechanically to suggest

appropriate treatments, and advice as to their efficacy and risks with a given patient, will soon be indispensable to responsible medical practice.

Summary

Problems of "definition" with computer applications, and of medical data banks were re-evaluated by means of a flexible coding system designed for use in busy clinical situations. This enabled a general practitioner to code 31,110 symptoms (11 per cent of which were "pain"), 12,646 diagnoses, and 18,243 treatments from 11,365 attendances by 4,196 patients in 12 months.

The ease with which the symptoms could be compared with their associated diagnoses and treatments, and *vice versa*, is illustrated. Quantification of therapeutic intent, obsolescence of double-blind trials, and prophylaxis for increasing iatrogenic disease are among the implications of improved medical data processing.

Acknowledgements

I wish to thank Dr H. H. Greenwood, Mrs A. F. Grundy and the staff of the Computer Centre, Keele University, whose generous help with the data processing made the project possible. The grant awarded by the Department of Health and Social Security is gratefully acknowledged. The support and encouragement given to the project at critical stages by Dr T. S. Eimerl and later by Dr G. K. Matthew was much appreciated. I am also indebted to Dr D. E. Clark and Mr T. C. Sharpe of the Medical Computing Unit, Manchester University, for their timely assistance.

References

- British Computer Society (1972). *Spectrum 71*. London: Butterworths.
- Dinwoodie, H. P. (1970). *Journal of the Royal College of General Practitioners*, **19**, 269.
- Gruer, K. T. (1972). *Journal of the Royal College of General Practitioners*, **22**, 100.
- Journal of the American Medical Association* (1971). **217**, 1464.
- Johnson, R. A. (1971). *The Medol Handbook*. 16c Clough Lane, Grasscroft, Oldham, Lancs.
- Johnson, R. A. (1971). *World Medicine*, **6**, 72.
- Macrae, N. (1972). *The Economist*. Jan. 22.
- Morrell, D. C., Gage, H. G. & Robinson, N. A. (1971). *Journal of the Royal College of General Practitioners*, **21**, 32.
- Ockenden, J. M. & Bodenham, K. E. (1970). *Focus on Medical Computer Development*. London: Oxford University Press.
- Pratt, A. W. (1971). Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland, U.S.A. Personal communication.
- The Records and Statistical Unit (1963). *Journal of the College of General Practitioners*, **6**, 195.
- South, J. & Rhodes, P. (1971). *British Medical Journal*, **4**, 32.
- The Lancet* (1970). **1**, 703.
- Vickery, B. C. (1960). *Faceted classification*. London: Aslib.
- World Health Organisation (1967). *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, p. 471 (6), (iii). Geneva: W.H.O.

SCOTTISH STATISTICS

The following figures are reported from the Scottish Home and Health Department: a fall in the birth rate from 87,335 in 1970 to 86,731 in 1971; a continuing rise in the number of illegitimate births to a rate of 8.1 per 100 during the first nine months of 1971; a continuing rise in the number of abortions to nearly seven per cent of the number of live births.

The total gross cost of the National Health Service in Scotland exceeded £200 million for the first time during the year 1971.

MEDICAL EDUCATION

Participants from 14 member states who are taking part in the Council of Europe's cultural activities have concluded that "human relations are becoming more and more important; if general family medicine is to be developed, more attention must be paid when selecting subjects to behavioural or psychological problems."