Adverse reactions in ten years' general practice, computer analysed

R A Johnson MRCGP PhD (medical computing) *General Practitioner, Oldham, Lancashire*

Summary

A total of 167 264 symptoms were recorded in numerical shorthand during a ten years period in a general practice. Computer analysis yielded 1535 adverse reactions with the drug prescribed. Especially prominent were reactions from three antibiotics (13% of all reported reactions), most of which were alimentary symptoms. Ampicillin also induced rashes reported at a rate of 5.2 per thousand prescriptions. Reactions to an oral contraceptive (Minovlar) were the most frequent to a single named drug, being 25.9% of prescriptions for that drug. A worldwide total of around one trillion symptoms seems likely before AD 2000. Computers could be used to detect the unpredictable side effects, as with practolol. Various 'costs' involved are noted, together with wider questions.

Introduction

Five hundred million is a reasonable estimate of the number of symptoms presented annually to general practitioners in Britain. In the 12 months from 1 October 1970, 31 110 symptoms were recorded[1], which on average is around 20 000 annually per general practitioner.

Three recent editorials have called attention to the major problem of adverse drug reactions[2-4]; and iatrogenic disease has been recorded at 30% in a hospital setting[5]. The detection of unpredictable adverse effects, best typified as the 'practolol-problem', is still more difficult. Only a serendipitous chain of events linked practolol with its now well known noxious effects.

There is in principle a more reliable method, which relates to the yet untapped annual harvest of half a billion symptoms. If all these symptoms could be laid out in chronological order, counting day one as the day a patient first received a given drug, then all adverse effects would peak in the ensuing time period, whether of days, weeks or decades. This paper is preparatory to such an exercise. It paves the way for a more rigorous demonstration of the more reliable antidote to the practolol-problem.

Coding clinical data for computers has not been popular. In reality, however, it is more reliable to persuade flexible humans to reorganize their information numerically (as they already do their telephonic addresses). When results from coding data are seen to be superior, it will no longer be a question of whether or not to encode clinical data, but of how best to do so.

Method

The singular advantage of a segmented or faceted code[6] over an ordinal one such as the International Classification of Diseases, is that far fewer codes can cover many more eventualities, making life much easier for the originating recorder, who alone can improve the validity of data.

The phrase 'cough hurts chest' almost reflects natural language, and falls naturally into three segments or facets. Table 1 shows the general 'Medol' symptom codes devised by the author (unpublished); by using these, all clinical medicine can be coded in outline. As with all information systems, increasing detail incurs increased costs, if only in terms of user fatigue. Reading from Table 1, 'cough hurts chest' would require '20-10-20', though at this simple level this only covers 'a respiratory ailment with pain somewhere in the chest'. The code '21-13-22-1W' adds more precision, though it is more complicated: the phrase now reads 'cough with severe pain, at the rib edge or costal margin for one week'.

The simplicity of the 16 general codes may seem to offer too coarse a sieve to convey the more intricate parts of clinical medicine. However, the three

Table 1. The 16 general symptom codes

System (physiological)			*Reaction*		*Area (anatomy)*	
			10	pains	10	head
20	respiratory	(26%)			20	chest
30	mental	(21%)			30	arm
40	alimentary	(18%)	40	'better'	40	trunk
50	skins, trauma and locomotory	(18%)	50	'no change'	50	leg
60	genitourinary	(10%)	60	'worse'		
70	ear, nose, eyes and metabolic	(4%)				
80	cardiovascular, neurological and allergy	(3%)				

(% = frequency of occurence in a decade of general practice)

0141-0768/86/030145-04/$02.00/0

columns in full run from '00' to '99', giving a total of 300 two-digit codes to cover a million uniquely identifiable symptoms. Moreover, the simplicity is essential to ensure that the overall framework may be readily grasped. Later, each 'system' code can be subdivided 500 times over, giving in excess of 50 000 'system' codes. In effect this means there is room for 500 varieties of 'cough', or even, should this prove needful, for 500 types of diarrhoea, or any of the 100 other 'system' codes, some of which appear below. The idiosyncracies of all medical specialties can thus comfortably be accommodated in a common code structure, allowing full compatibility between all clinical users. The duration code is self-explanatory, since time long ago submitted to digitization, down to the nearest nano-second.

Results

Every clinical item that occurred during normal general practice between August 1970 and February 1981 was recorded consecutively by one practitioner. A total of 167 264 symptoms were recorded, of which 1535 were adverse reactions spontaneously reported, together with the drug noted at the time.

Table 2 shows the ten most frequent symptoms reported with the drug named. Table 3 lists the ten drugs most frequently mentioned. Again, analysis of the changes from one year to the next is available from the data. The proprietary names are shown (except for Septrin, Distalgesic, and Mogadon) since this is what was written on the prescription. The recent 'limited list' has revealed a remarkable attachment by some patients to a given brand: they insist the generic equivalent is not as efficacious. Most regrettably, lack of funding prevented recording to document and thereby investigate the extent of this 'brand affection'.

Minovlar received most comments, especially amenorrhoea, though head pains also featured. Indomethacin had the next highest rate percent prescribed, a telling portent of the problems which later befell similar drugs.

Table 4 indicates the detail in which the various reported reactions were recorded. Alimentary reactions were reported more frequently than antibiotics, rising to 42 per thousand for co-trimoxazole. Ampicillin induced a higher rate of rash reporting than co-trimoxazole, though some reported the determatological effects as 'spots' which takes code S52.

The reliability of the data cannot be higher than that supported by the method of data entry, and data verification. Punched cards were notorious as sources of error; only 'real time' data entry[7] and validating could reduce the error levels. Nevertheless, three 'check' columns were included in the symptom fields, and these seem to have stabilized the entry data processing of symptoms to an error level of around 10%. The dates of attendance and of birth were less well protected, and seem to have suffered rates two or three times higher. The majority of these could undoubtedly be retrieved were the erroneous data fields to be scanned 'manually' – a task for which resources are not currently available.

Table 2. Adverse reactions by symptoms reported

Symptom	*No. (%) of symptoms*	*Most frequent drug associated (with numbers found)*
S35 tired	204 (14)	Diazepam (49), antidepressants (33), antibiotics (22)
S44 vomiting	143 (9.5)	Antibiotics (34)
S68 period troubles	140 (9.3)	Oral contraceptive (139), oxytet (1) [less period (16), heavier period (15), no period (21)]
S92 pain	111 (7.5)	47 'head' (20 with oral contraceptive), 30 trunk
S46 diarrhoea	102 (6.8)	Antibiotics (40) [including ampicillin (18), oxytet (6)]
S14 dizziness	75 (5)	Antibiotics (17), oral contraceptive (9), diuretics (8)
S43 nausea	75 (5)	Antibiotics (20), indomethacin (10)
S47 constipation	70 (5)	Iron preparations (16)
S51 rash	69 (5)	Antibiotics (34) [ampicillin (15), co-trimoxazole (3)]
S94 'troubled by'	62 (4)	No obvious pattern [antibiotics (8)]

Table 3. Adverse reactions by drug

Number prescribed	*Drug reported*	*Total (%) reactions reported*	*% of drugs prescribed*	*Commonest symptom (no. of reactions)*	
558	Minovlar	145 (9.5)	25.9	Period troubles	(68)
2964	Diazepam	91 (6)	3.0	Tired	(49)
2879	Ampicillin	77 (5)	2.6	Diarrhoea	(18)
3931	Oxytet	75 (5)	1.9	Vomiting	(11)
665	Co-trimoxazole	47 (3)	7.0	Vomiting	(10)
2420	Co-proxamol	42 (3)	1.7	Tired	(9)
416	Indomethacin	37 (2.5)	8.8	Nausea	(8)
1405	Albamycin-T	33 (2)	2.3	Diarrhoea	(11)
1827	Nitrazepam	30 (2)	1.6	Tired	(5)
938	Navidrex-K	27 (2)	2.8	Vomiting	(11)

Table 4. Detailed reactions to three antibiotics

'System' code	Reactions: Ampicillin (2879 prescriptions)		Oxytetracyline (3931 prescriptions)		Co-trimoxazole (665 prescriptions)	
S30 mental						
S31 insomnia	2	0.7%%				
S32 stress						
S33 anxiety						
S34 irritable	1	0.3%%				
S35 no energy	4	1.4%%	7	1.8%%	7	10.5%%
S36 depressed			1	0.2%%		
S37 phobic						
S38 rambling	2	0.7%%	2	0.5%%		
S39 addiction						
S40 alimentary						
S41 anorexia						
S42 dysphagia					2	0.3%%
S43 nausea	3	1.0%%	5	1.3%%	5	7.5%%
S44 vomiting	6	2.0%%	11	2.8%%	10	15.0%%
S45 dyspepsia	3	1.0%%	3	0.8%%	3	4.5%%
S46 diarrhoea	18	6.0%%	6	1.5%%	2	0.3%%
S47 constipation	4	1.3%%	5	1.3%%	3	4.5%%
S48 'piles'						
S49 faeces						
S51 rash	15	5.2%%	3	0.7%%	3	4.5%%

(%% = rates per thousand prescriptions of named drug)

A total of 84 481 cards were counted and screened in about ten minutes, using a less than optimized algorithm.

It is intended to implement other analyses as facilities become available, notably the 'practolol-problem' already mentioned and also the incidence of aches and pains (14 501 in all) and of mental symptoms (13 788 counted so far).

Discussion

The thrust of this brief paper is that the digital computer can be harnessed far more effectively to the service of medicine than is generally supposed, provided the data are digitized. The actual results shown are more illustrative than definitive, since over any 10-year span substantial changes are bound to occur, both in practice and personal circumstances.

The thesis proposed is that by using a method such as this, it is perfectly feasible to analyse virtually unlimited quantities of clinical data in a reasonable length of time.

The question of costs has two aspects. In terms of equipment, magnetic media have for some time been cheaper than paper or cardboard. For example, all the copies of this journal for twenty years could comfortably be stored on a standard 180-minute videocassette (cost £4 each, with 1 minute per megabyte) making the cost of each copy only 2 pence.

The one cost which is rising is that of human labour. Computers can only be cost effective in terms of time when the clinical output improves. However, there are other considerations which may influence the outcome, not least the vast, almost astronomical quantity of clinical data which cannot otherwise be scrutinized. The cost of medical ignorance is iatrogenic disease.

Extrapolating from the figure of 500 million symptoms mentioned above, gives an annual average of 10 symptoms per person. Worldwide this amounts to around 40 billion for 1985, rising to some 60 billion annually by the year 2000, approaching one trillion symptoms (10^{12}) in cumulative total by then. By that time, parallel processors would need only 10 seconds to screen for a given item. Analysing this amount of free text, even were it legible, would not be feasible. Only digital processes move fast enough to cope with such figures, and in deference to this it would seem prudent to digitize medical data.

Since earliest times, writing has consisted of attributing human meaning to marks inscribed on a plain surface. This system breaks down when the number of such symbols exceeds say 100 million, for the sheer mountains of physical material frustrate any reliable retrieval. The digital computer handles many times this quantity of data with speed and precision, always with the proviso that the data first be converted into digits or numerals. One trillion symptoms (at 4 bytes each) could be stored today in a small box of 12 inch optical disks.

Symptoms are the very life-blood of medical practice; they are what patients bring for us to alleviate, and also what some of our treatments inflict upon them. Is the profession prepared to remain ignorant of such a large quantity of them? Or will it adapt its recording techniques so as to avail itself of these facilities? What else are medical records for? And if their contents cannot be analysed, why keep them? These are some of the challenges posed by computers.

Acknowledgments: I wish to thank Professor E T Powner for his support, and also the late Professor P Byrne. I am grateful too to Mrs A Hart, Mr G Dineley and Mr P Watts for their help.

References

1 Johnson RA. Computer analysis of the complete medical record including symptoms and treatment. *J R Coll Gen Pract* 1972;**22**:655–60

2 Anonymous. Non-steroidal anti-inflammatory drugs: Have we been spoilt for choice? *Lancet* 1984;i:141–2

3 Turner P. Long-term assessment of drug safety and efficacy. *J R Soc Med* 1984;**74**:93–4

4 Rawlins MD. Postmarketing surveillance of adverse reactions to drugs. *Br Med J* 1984;**288**:879–80

5 Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *N Engl J Med* 1981;**304**:638–42.

6 Vickery BC. *Faceted classification*. London: ASLIB, 1960

7 Johnson RA. *Real time retrieval of clinical data*. PhD Thesis. Manchester University, 1980

(Accepted 22 November 1985. Correspondence to 16c Clough Lane, Grasscroft, Oldham OL4 4EW)