VARIATIONS IN RESPONSES TO NON-STEROIDAL ANTI-INFLAMMATORY DRUGS

D.L. SCOTT¹, S. RODEN², T. MARSHALL¹ & M.J. KENDALL¹

¹The Medical School, University of Birmingham, Birmingham B15 2TJ and

²Selly Oak Hospital, Birmingham B29 6JD

- 1 This paper assesses the variability in the response of patients with rheumatoid disease to a group of non-steroidal anti-inflammatory drugs (NSAID).
- 2 Thirty six patients were divided into three groups and each group received a set of four NSAID. The different drugs were each given three times daily for a week and the treatment order was randomised.
- 3 The study demonstrated marked variability and unpredictability in patient responses and cast doubt on the value of comparative trials of different NSAID.

Introduction

The large number of non-steroidal anti-inflammatory drugs currently available makes it difficult to select the most suitable drug for an individual patient. All these drugs are effective in reducing the inflammatory component of rheumatoid disease, and as analgesics (Hart, 1978). They are usually evaluated in short term trials against known comparators, although some studies have compared multiple drugs or monitored their use over long periods (Aylward *et al.*, 1975; Huskisson *et al.*, 1976; Pinals *et al.*, 1977). However, in evaluating these short term trials there are many problems. These include assessing the extent of the placebo effect, the lack of objective end points, the choice of appropriate doses, the variability of the disease and the marked differences between patients. The impact of the last two variables may be so great as to mask relatively small differences between drugs. Nevertheless if the commonly used non-steroidal anti-inflammatory drugs, when given at equivalent dosages, could to some extent be ranked in order of efficacy and tolerability this would help the prescriber very considerably.

The aim of this study was to determine whether a ranking of these drugs is possible or whether the variation between individual patients in the response to these drugs is so great that it obscures any possible differences due to the drugs. In evaluating the effectiveness of therapy we have used two of the most important indices of inflammatory synovitis—joint pain and morning stiffness.

Methods

Thirty-six patients with definite or classical rheumatoid arthritis (by the criteria of the American Rheumatism Association, 1958) were entered into a 4 week prospective study. Patients who had recently started gold, penicillamine or steroid therapy or who had changed their dose of these drugs within the preceding two months were excluded. There were 30 women and 6 men of mean age 59 ± 8 years; the duration of their rheumatoid arthritis ranged from 6 months to 32 years (mean 5 years). Each patient took four different drugs and was randomly allocated to one of the following treatment groups: *Group A* ibuprofen (400 mg), benorylate (1500 mg), indomethacin (50 mg), fenoprofen (600 mg); *Group B* ibuprofen (400 mg), sulindac (100 mg), naproxen (250 mg), ketoprofen (50 mg); *Group C* ibuprofen (400 mg), aspirin (enteric coated 1200 mg), diclofenac (50 mg), flurbiprofen (100 mg). Each drug was given with or after meals in a similar dosage of three times daily for 7 successive days, and the four drugs were given over a consecutive 4 week period in random order. A supply of paracetamol tablets was given in case additional analgesia was required.

Every day, patients recorded the severity of their joint pain (in four categories: none; minimal; moderate; severe) and the duration of their morning stiffness (in four categories: none; less than 30 min; 30–60 min; greater than 60 min). Patients for whom data was completely absent for one or more drugs (because of refusal to take them) were excluded from

the analysis (2 in Group A, 1 each in Groups B and C. Other missing values (mostly reflecting incomplete compliance due to adverse reactions) occurred in 18/128 drug-patient exposures with respect to morning stiffness, and 23/128 exposures with respect to pain (a drug-patient exposure is the exposure over (up to) 7 days of one patient to one drug). The difference between the number of incomplete exposures on each index reflects a failure to record one of them. In all these instances (i.e. whether failure to take the drug and/or failure to record a response), missing values were estimated as suggested by Huitson (1966). There was no evidence that previous exposure to a drug influenced the results.

Methods

Pain and morning stiffness have been treated as two separate outcome indices, and have been analysed independently for each of the three treatment groups. The two main factors to be assessed were *drugs* and *patients*. The drugs were selected to be of specific interest, and their effects were assumed to be fixed. In contrast, the patients could be regarded as a sample from the whole population of patients with rheumatoid disease, and hence their effects were assumed to be random. Records on different days of the week were used as 'replicates'. In this way, it was possible to study the existence of an interaction effect, i.e. whether the response by an individual to a drug could be predicted from the main 'patient' effect and the main 'drug' effect, or not.

The main method of analysis was by an analysis of variance, using patient responses as a measured variable. Basically, an additive model has been proposed, of the form

$$X_{ijk} = \mu + D_i + P_j + (DP)_{ij} + E_{ijk}$$

Where X_{ijk} = the response of patient j to drug i on day k

μ = the overall mean

D_i = the difference between the overall mean and the i 'th drug mean (i.e. the i 'th drug effect)

P_j = the difference between the overall mean and the j 'th patient mean (i.e. the j 'th patient effect)

$(DP)_{ij}$ = the difference between the overall mean and the (ij) 'th drug/patient mean, after allowing for the i 'th drug effect and the j 'th patient effect.

E_{ijk} = the day-to-day (within patient) variability of patient j on drug i .

Testing the validity of this model by an analysis of variance splits the total variance into four

components: that due to variation between drugs; that due to variation between patients; that due to variation between different drug/patient combinations; and that due to within patient/within drug variability (the 'error' variance in this study).

Results

The four categories for recording morning stiffness and joint pain were labelled 0-3, and these labels were subsequently used as scores in the analysis. Mean scores for each drug (across all patients) and each patient (across all drugs) are given in Tables 1 and 2.

Table 1 Mean drug scores for morning stiffness and pain.

		Morning stiffness			Pain		
		Group			Group		
		A	B	C	A	B	C
Drugs	1	2.35	2.08	1.84	1.93	1.79	1.62
	2	2.58	1.96	2.07	2.28	1.52	1.74
	3	2.43	1.95	1.82	1.71	1.73	1.73
	4	2.56	2.21	2.22	2.18	1.62	1.97
		A			B		
Key	1	Indomethacin		Sulindac	Diclofenac		
	2	Benorylate		Ketoprofen	E.C. Aspirin		
	3	Fenoprofen		Naproxen	Flurbiprofen		
	4	Ibuprofen		Ibuprofen	Ibuprofen		

Two features are worth noting. Firstly, *within* a group differences between patients far exceed differences between drugs, on both indices. Secondly, there are marked differences between the patterns of response for morning stiffness and joint pain. One third (43/128) of patient-drug exposures gave a maximum total score (i.e. all seven days added

Table 2 Mean patient scores for morning stiffness and pain.

<i>Morning stiffness</i>			<i>Pain</i>		
<i>Group</i>			<i>Group</i>		
<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
2.63	1.84	2.54	2.29	1.66	2.03
3.00	0.25	1.66	2.50	0.61	1.45
2.88	3.00	0.62	2.84	2.79	1.61
2.86	2.21	3.00	1.63	1.25	1.75
2.79	1.21	1.36	1.54	1.36	1.27
3.00	1.00	2.89	2.18	1.15	2.41
0.86	2.25	2.39	1.09	2.46	2.00
2.14	3.00	1.18	2.18	2.00	1.57
2.00	3.00	2.07	2.00	2.39	1.83
2.48	2.67	1.21	2.03	1.24	1.14
	2.11	2.95		1.39	2.35

together) for morning stiffness (a maximum score indicating the most severe stiffness) compared with only 7/128 maximum scores for joint pain. On the other hand, in only 5/128 exposures was stiffness completely abolished, and pain in only 3/128 exposures. This suggests that in future a finer scale for assessing stiffness should be used.

The results of an analysis of variance are presented in Table 3. In this table, if the null hypothesis were true (that drugs, patients and interaction all had no effect), the F-ratio would in each case be equal to 1. The statistical significance of an effect is determined by the extent of the departure of the F-ratio from 1 in relation to the degrees of freedom involved. In each analysis the drug effect was not significant, showing that little of the considerable variation in morning stiffness and pain could be attributed to consistent differences between the drugs. It is therefore not possible to rank the drugs in any consistent order of efficacy.

In contrast, every analysis gave highly significant patient effects for both morning stiffness and joint pain ($P < 0.005$ in all cases); and the same was true for the interaction effect. This effect is due to some patients responding in an unpredictable way to some of the drugs they received when compared to their overall response to all the drugs and to a particular drug's effect on all patients. The main findings can thus be summarised as follows:

1. There are differences between the scoring systems for the two different outcome indices, the grading for morning stiffness being less discriminatory than that for pain.
2. The analysis of variance suggests that differences between drugs are of no importance in an overall sense compared with other sources of variation studied.
3. The same analysis suggests that patients vary widely in their responses to the drugs under test.
4. Likewise, there is a substantial interaction effect. The practical interpretation of this finding is that some patients are particularly suited to some drugs, and particularly *not* suited to others.

Power

Because of the essentially negative finding with respect to the possible existence of a drug effect, the power of the study is important. For a given significance level (we have used 5%), power is defined as 1 minus the probability of failing to find an effect if it really exists. As with all such calculations, the main problem is in specifying the non-centrality parameter: in practice, this reduces to deciding the minimum extent of any genuine drug effect it is felt would be of clinical significance.

This has been studied in terms of the relationship between the between-drug variability and the within-

Table 3 Analysis of variance of pain and morning stiffness in the three treatment groups

		Group A			Group B			Group C		
		F ratio	Degrees of freedom	P value	F ratio	Degrees of freedom	P value	F ratio	Degrees of freedom	P value
Joint pain	Drug effects	2.8	3,27	NS	0.8	3,30	NS	1.1	3,30	NS
	Patient effects	33.8	9,211	< 0.005	78.0	10,240	< 0.005	19.1	10,249	< 0.005
	Interactions	7.8	27,211	< 0.005	8.6	30,240	< 0.005	6.1	30,249	< 0.005
Morning stiffness	Drug effects	1.6	3,27	NS	1.9	3,30	NS	1.4	3,30	NS
	Patient effects	105.0	9,213	< 0.005	192.3	10,252	< 0.005	87.6	10,247	< 0.005
	Interactions	4.4	27,213	< 0.005	7.3	30,252	< 0.005	7.9	30,247	< 0.005

NS = not significant

patient/within-drug variability. The results offered here have been calculated retrospectively, are approximate, and are based on the work of Pearson & Hartley (1951) supplemented by Winer (1971). The implications of two assumptions or alternative hypotheses have been studied.

- (a) that the between-drug variation should be not less than the within-patient/within-drug variation. (This is an extremely weak assumption. Realistically, if variation between drugs were no greater than that obtained from day-to-day variation by the same patient to a single drug, one would not be very interested). Here, the power of the test is 94%.
- (b) that the between-drug variation should be not less than twice the within-patient/within-drug variation. In this case, the power is over 9%.

Discussion

Our methodology raises two points. Firstly, only short courses of treatment were given, but these should be adequate to assess the effect of drugs on symptoms; Huskisson & Scott (1978) have shown

that these drugs are effective within a brief period of time. Secondly, in the absence of reliable data on equi-anti-inflammatory doses, an informed estimate had to be made based on the literature available and current clinical practice, but to all intents and purposes the drugs were given at equivalent dosages.

The major finding of this study is that the variation between patients in their response to non-steroidal anti-inflammatory drugs is very large whereas there is no discernable difference between the drugs. Huskisson *et al.* (1976) reported a similar finding in a study of four propionic acid derivatives. This result has two implications. Firstly, it is probably impossible to rank NSAID in terms of therapeutic efficacy. Secondly, attempts to assess new NSAID by comparing them with one of the standard agents such as indomethacin may have little meaning or clinical relevance.

Side effects to these drugs and the unpredictable responses given by some patients to some drugs means that in clinical practice each physician must be familiar with a number of different non-steroidal anti-inflammatory drugs and be prepared to change the prescription if the patient's clinical response is not satisfactory.

References

- AYLWARD, M., PARKER, R.J., HOLLY, F., MADDOCK, J. & DAVIES, D.B.S. (1975). Long term study of indomethacin and alclofenac in the treatment of rheumatoid arthritis. *Br. med. J.*, **2**, 7.
- HART, F.D., (1978). *Drug treatment of the rheumatic diseases* Lancaster: MTP Press
- HUITSON, A. (1966). *The analysis of variance*. London: Griffin.
- HUSKISSON, E.C. & SCOTT, J. (1978). Sulindac. Trials of a new anti-inflammatory drug. *Ann. Rheum. Dis.*, **37**, 89–92.
- HUSKISSON, E.C., WOOLF, D.L., BALME, H.W., SCOTT, J. & FRANKLYN, S. (1976). Four new anti-inflammatory drugs: responses and variations. *Br. med. J.*, **1**, 1048.
- PEARSON, E.S. & HARTLEY, H.O. (1951). Charts of the power function for analysis of variance tests derived from the non-central F-distribution. *Biometrika*, **38**, 112–130.
- PINALS, R.S., RUNGE, L.A., JABBS, J.M. & MADDI, V.I. (1977). New non-steroidal anti-inflammatory agents. Prospective survey of relative merits. *New York State J. Med.*, **77**, 1268.
- WINER, B.J. (1971). *Statistical principles in experimental design*. Tokyo: McGraw-Hill Kogakusha.

(Received March 16, 1982,
accepted June 30, 1982)