

## Inability of rheumatologists to describe their true policies for assessing rheumatoid arthritis

J R KIRWAN,<sup>1</sup> D M CHAPUT DE SAINTONGE,<sup>1</sup> C R B JOYCE,<sup>2</sup>  
J HOLMES,<sup>1</sup> AND H L F CURREY<sup>1</sup>

From <sup>1</sup>The London Hospital Medical College, London E1; and the <sup>2</sup>Medical Department, Ciba-Geigy Ltd, Basle, Switzerland

**SUMMARY** Eighty nine British and Australian rheumatologists took part in a study to discover how accurately they could describe their procedures for measuring disease severity in rheumatoid arthritis. The relative importance they attached to different clinical and laboratory variables showed a very wide variation, and these stated policies were generally poor at predicting their actual judgments when assessing 'paper patients' ( $r^2=39\%$ ). Policies based on equal weighting of all variables, while also poor predictors ( $r^2=41\%$ ), were nevertheless superior to their stated policies for 49 respondents. Policies calculated by judgment (linear regression) analysis were much more successful predictors ( $R^2=73\%$ ). Unhurried, detailed interviews with four experienced rheumatologists provided carefully considered statements of assessment policy, but these also were poor predictors of routine assessments of outpatients ( $r^2=34\%$ ) compared with policies calculated by clinical judgment analysis, even when these were applied to new data ( $R^2=88\%$ ).

**Key words:** judgment policy models, judgment analysis, disease assessment.

Clinical research in rheumatoid arthritis (RA) often requires the assessment of patients and their response to treatment. The basis on which such judgments are made should be clear and explicit if they are to be capable of verification and use by others. Earlier work<sup>1</sup> suggested that two rheumatologists' descriptions of their policies for assessing disease activity did not reflect the actual decisions they made in practice. If this were generally true it would prevent true agreement and uniformity in the selection of cases for clinical research on the cause or the most appropriate means of treating RA. Two investigations were therefore conducted. The first surveyed 89 rheumatologists who stated their policies for judging changes in RA, while the second consisted of separate in-depth interviews with four rheumatologists who attempted to define their own judgment policies with care and in detail. Predicted assessments of RA made from all these stated policies were compared with those actually made in practice.

Accepted for publication 1 July 1985.  
Correspondence to Dr J R Kirwan, Bone and Joint Research Unit, The London Hospital Medical College, London E1 2AD.

### Patients and methods

#### SURVEY OF RHEUMATOLOGISTS

##### *Sample*

The observations for this investigation were taken from a 15% random sample of British rheumatologists and 34 self-selected Australian rheumatologists in previous reports.<sup>2-4</sup> The 89 respondents had rated the importance they gave to 10 clinical variables when evaluating patients' responses to 'disease modifying' drugs such as gold or penicillamine. Each awarded marks out of a total of 100 to indicate the relative importance he or she attached to each variable. Each rheumatologist then assessed 50 'paper patients'<sup>5</sup> (simple forms giving the values for the same clinical variables before and after treatment derived from real patients<sup>2</sup>). The assessments were made on a visual analogue scale and subsequently scored from -55 (worst possible deterioration) to +55 (best possible improvement). These scores are referred to as the rheumatologists' actual judgments. Twenty of the 50 paper patients were replicates in order to allow the reliability of the judgments to be estimated.

*Policy models*

The marks given by each rheumatologist to indicate the importance, as he saw it, of each clinical variable in influencing his assessments were used as weights. The weights of all rheumatologists were standardised to place them on the same scale. The differences between the before and after treatment values of the variables for the 'paper patients' were calculated and were also standardised. For each patient the standardised values of the variables were multiplied by each rheumatologist's standardised set of weights. The 10 products were summed to give 'stated policy predicted judgments', representing the judgment each rheumatologist would have made about each patient had he used the data in the way he had stated.

A similar calculation was made for each patient but giving equal importance to the 10 variables. These scores are referred to as 'equal weight predicted judgments'.

Clinical judgment analysis (CJA)<sup>1 2</sup> was used to derive models of the judgment policy of each rheumatologist. A multiple linear regression equation was calculated taking the rheumatologist's actual judgments as the dependent variable and the changes in value of the clinical measures as the independent variables. This procedure was changed slightly from that in previous reports<sup>2-4</sup> in which before as well as change values were employed, because it enabled the equation to be compared directly with the stated policy of each rheumatologist. The use of the equation removes the random inconsistencies with which the rheumatologist applies his policy in practice.<sup>1 2</sup> The model (equation) for each rheumatologist was then applied to the

'paper patient' data. The resultant (calculated) judgments were termed the 'CJA predicted judgments'.

INTERVIEWS

Four experienced rheumatologists initially saw, assessed, and recorded clinical observations on a series of outpatients suffering from RA (1st set). Their judgments of 'current disease activity' were used to calculate CJA policy models as previously described.<sup>1</sup> Subsequently, further outpatients (2nd set) were seen and assessed (36 patients for doctor 1; 84 for doctor 2; 41 for doctor 3; 62 for doctor 4) and judgments of 'current disease activity' noted. Each rheumatologist was then interviewed separately. Each spent 75-90 min considering in depth his approach to the assessment of 'current disease activity' in patients with RA. This introspection was facilitated by considering 'paper patients', representing eight to 12 patients randomly selected from the 2nd set. Each rheumatologist made repeated attempts to formulate statements or equations representing his assessment policy. At the conclusion of the session each expressed his considered judgment policy in the form of a series of conditional statements or equations.

Both the CJA policy model equations and the considered judgment policy statements were applied to the clinical data from the 2nd set of outpatients, resulting in a series of 'CJA predicted judgments' and 'considered policy predicted judgments' for each rheumatologist. These were compared with actual judgments made by the rheumatologists when they had seen the 2nd set of patients in the clinic.

This procedure is summarised in Fig. 1.

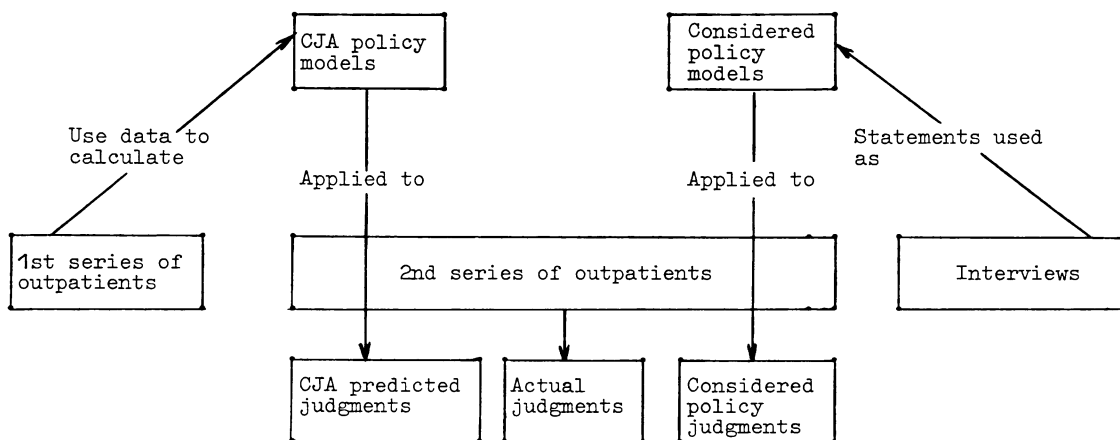


Fig. 1 A summary of the procedure for obtaining 'considered policy predicted judgments'.

## COMPARISONS

The rank correlation coefficient was used to compare judgments derived by the three methods and to compare the sets of replicate judgments. Mean values for stated and calculated judgments were found after Z transformation of the individual correlation coefficients.

## Results

## SURVEY OF RHEUMATOLOGISTS

The relative importance attached to the 10 clinical variables by the 89 rheumatologists is summarised in Table 1. Complete data, including the correlation coefficients for replicates, are available.<sup>6</sup> There was very wide variation. For example, the importance of the ESR ranged from 0 to 21 and that for the erosion count from 0 to 80.

The judgments calculated by the three methods of defining the judgment policies were correlated with the actual judgments each rheumatologist made. The success of each model is shown in Table 2 as the square of the correlation coefficient. This is the proportion of variance in the judgments that can be accounted for by each method.

Stated policies were generally poor at modelling actual judgments (pooled value for explained variance,  $r^2=39\%$ ). The correlation between judgments predicted by stated policies and those actually made failed to reach statistical significance for 14 rheumatologists ( $r^2<16\%$ ) and for 59 was less than 0.7 ( $r^2=49\%$ ).

Policies based on equal weighting of all variables were also poor at predicting actual judgments (pooled  $r^2=41\%$ ), though they were better than stated policies for 49 rheumatologists.

In contrast, CJA policies were usually reasonably good predictors of actual patient assessments (pooled  $R^2=73\%$ ), and for only a single rheumatologist explained less than 50% of the variance. For 19 doctors the variance explained was greater than 80%.

## INTERVIEWS

The final expressions of the considered judgment policy defined by the four rheumatologists interviewed are shown in Table 3. Some were simple equations (e.g., doctor 1), while others required complicated rescaling of the clinical variables (e.g., doctor 3).

The variance in actual judgments explained by the two methods of modelling (considered policy models and CJA models) is shown in Table 4. The pooled values for the four rheumatologists were similar to those obtained during the large surveys and are shown in the bottom row of Table 2. The CJA policy models were clearly superior to the considered policy models at accounting for the judgments actually made by the rheumatologists and explained 80–95% of the variance in their assessments.

## Discussion

The two studies show that, in general, rheumatologists' descriptions of the importance they give to various clinical variables when judging the progress and severity of rheumatoid arthritis are relatively poor predictors of their actual 'paper patient' or real patient assessments. This was true for a large number of rheumatologists in both Britain and Australia and suggests that the assessment policies believed by rheumatologists to be those they employ when classifying the extent or progress of arthritis tell us little about their actual practice.

It may be argued that the method by which these descriptions of perceived policies were collected (scoring the variables on a simple form) was crude and did not provide an opportunity for adequate consideration. Other situations, such as explaining their method of assessment to undergraduate medical students, might result in more accurate statements of policy. It was to investigate this possibility that four rheumatologists were asked to spend as much time as they desired in describing their

Table 1 Summary of stated policies of 89 rheumatologists

Doctor	Relative weights*									
	Asp	ESR	Hb	Eros	EMS	Glob	Pain	FC	Grip	AI
Mean	9	11	9	12	10	11	9	10	8	11
Median	8	11	9	11	10	11	9	10	8	11
Minimum	0	0	0	0	0	0	0	0	0	0
Maximum	100	21	16	80	23	21	20	20	14	35

\* Asp=aspirin consumption; ESR=erythrocyte sedimentation rate; Hb=haemoglobin; Eros=number of erosions on hand x-ray; EMS=early morning stiffness; Glob=patient's global assessment; Pain=pain score; FC=functional capacity; Grip=grip strength; AI=articular index.

Table 2 *Variance explained by different policy models (%)*

<i>Doctor*</i>	<i>Stated policy (r<sup>2</sup>)</i>	<i>Equal weight policy (r<sup>2</sup>)</i>	<i>CJA policy (R<sup>2</sup>)</i>	<i>Doctor*</i>	<i>Stated policy (r<sup>2</sup>)</i>	<i>Equal weight policy (r<sup>2</sup>)</i>	<i>CJA policy (R<sup>2</sup>)</i>
M02	31	30	81	P08	39	36	57
M04	60	56	81	P09	52	51	67
M05	57	61	69	P10	42	42	83
M06	42	34	78	F01	49	54	77
M08	17	19	78	F02	50	53	65
M10	24	34	67	F03	56	47	73
M11	29	26	50	F05	41	36	78
M12	23	31	61	F08	35	31	76
M13	50	44	69	F09	46	45	70
M14	40	38	74	F10	31	38	65
M15	29	30	78	A01	45	53	75
M16	42	49	71	A02	37	60	93
M17	44	54	80	A04	17	20	61
M19	64	57	79	A08	66	66	77
M21	00	56	89	A11	55	49	70
M22	52	57	69	A12	16	20	48
M23	40	48	77	A13	44	40	54
M24	51	57	68	A14	45	51	77
M25	44	34	72	A15	33	29	54
M27	51	46	80	A16	29	23	55
M28	25	18	67	A18	36	37	56
M29	50	52	74	A20	04	05	80
M30	54	49	81	A21	10	09	56
M31	35	32	59	A23	25	32	73
M32	37	38	78	A24	41	38	56
M35	00	03	84	A25	51	58	72
M36	49	37	74	A26	43	43	66
M37	55	58	76	A27	33	37	49
M39	69	74	80	A28	25	25	71
M40	43	51	87	A30	29	31	75
M41	33	51	70	A31	51	47	82
M42	13	10	43	A32	62	59	84
M43	54	60	84	A33	39	49	75
M44	25	28	66	A34	35	40	76
M45	48	54	77	A35	26	26	66
M46	54	56	74	A36	49	53	76
M47	27	28	74	A37	56	59	81
M49	42	44	63	A38	26	23	59
M50	17	18	78	A39	29	37	64
P01	45	41	62	A40	21	25	58
P02	53	49	71	A42	40	39	67
P03	58	57	85	A43	55	55	84
P04	23	18	72	A44	31	31	67
P05	21	20	67	A45	43	66	86
P07	51	44	74				
				Pooled	39	41	73

\*These codes correspond to those employed in references 2 and 3.

Table 3 *The carefully considered judgment policy models of four rheumatologists*

Doctor 1	CDA=JC+EMS+FC*	
Doctor 2	(a) Rescale all clinical variables so that their mean is 5 (b) CDA=2×JC+Pain+EMS+Glob+Grip+FC	
Doctor 3	(a) Rescale EMS so that if EMS > 59 then let E=1 if EMS < 60 then let E=0 (b) Rescale Grip so that if Grip < 51 then let G=4 if 50 < Grip < 101 then let G=3 if 100 < Grip < 201 then let G=2 if 200 < Grip < 301 then let G=1 if 300 < Grip then let G=0 (c) Rescale FC so that F=FC - 1 and if JC > 8 then double the value of F (d) CDA=2×JC+Pain+Glob+E+G+F	
Doctor 4	(a) If EMS > 10×JC then CDA=EMS/3 (b) If EMS ≤ 10×JC then CDA=10×JC/3 } to a maximum of 100	

\*CDA=judgment of current disease activity; JC=the number of currently synovitic joints (joint count); EMS=minutes of early morning stiffness; FC=Steinbroker functional capacity; Pain=patient's pain score (0-4); Glob=patient's own global assessment of the state of their arthritis (0-4); Grip=mean grip strength in mmHg.

Table 4 *Variance (%) in actual judgments explained by the considered policy and the CJA policy models of four rheumatologists*

Doctor	Considered policy ( $r^2$ )	CJA policy ( $R^2$ )
1	33	82
2	41	95
3	44	88
4	8	80
Pooled value	34	88

assessment policies in detail, having been provided with some typical patient data on which to experiment while they attempted to describe their policies. Even in these circumstances carefully considered policies were relatively unsuccessful in predicting actual decisions (Table 4).

In a study of Canadian rheumatologists, who were asked to describe their policies using a series of predefined graphical formats and complex weighting techniques, there was a similar failure to improve on initial brief statements.<sup>7</sup>

By considering all the clinical variables to be equally important it was possible to model assessments as well as (and often more successfully than) using the stated policies of rheumatologists, though the variance explained by equal weight policies remained relatively low.

In contrast, CJA provided a more accurate method for modelling the judgment policies of rheumatologists, in some cases accounting for almost all the variance in their assessments. The

regression models used for CJA did not take account of the tendency of such equations to 'overinterpret' the data from which they have been derived,<sup>1,2</sup> and the equations are likely to be less well suited to new data sets. However, the improvement in predictive power compared with stated and equal weight policies was very large and is unlikely to have been severely affected by this statistical artefact. The CJA policies of the four rheumatologists taking part in the interviews did take this factor into account<sup>2,6</sup> and nevertheless found similar differences between CJA policies and perceived policies. Further, the CJA policies were applied to a data set different from that used to calculate the policy equations (Fig. 1). Thus even when the rheumatologists were provided with every facility for describing their policy and CJA was applied conservatively, it correlated more closely with the actual judgments than did predictions based on stated policies or equal weight policies.

It could be argued that the linear CJA model proved unsatisfactory for some rheumatologists either because their judgments were unreliable, or because they were not actually using linear models but a more complex method of aggregating the information. However, there is a close association between  $r$  for replicates and  $R$  (correlation=0.812); i.e., the evidence strongly suggests that poor fits are mostly accounted for by unreliability of the judge rather than by the inappropriateness of the model. For only four rheumatologists was  $R$  low but  $r$  high, suggesting that the judge had used other than a linear model.

This result differs from that reported in the psychology literature where several examples can be cited of equal weight models (and even random weight models) proving better predictors than calculated regression equations (for a review see reference 8). However, these experiments often centred on unrealistic and unfamiliar judgment tasks performed by non-expert judges. Our results indicate that sophisticated modelling techniques are worthwhile in clinical circumstances.

Our findings also invite caution in interpreting rheumatologists' opinions about the importance of different measures of disease severity<sup>5</sup> and of disease outcome<sup>9</sup> in RA. Opinions expressed in surveys<sup>5 10</sup> and conferences<sup>9</sup> may be different from actual practice.

CJA, by providing the most accurate model of each rheumatologist's behaviour in assessing patients, allows analysis of the underlying reasons for differences between their assessments.<sup>1 2 4</sup> A reduction in interpersonal variation in judgment is an essential prerequisite to collaborative studies or co-operative decision taking and is a major aim of research in this field. The use of CJA to reveal the systematic element of these variations seems to provide an avenue for reaching agreed policies.<sup>11 12</sup> The finding that agreement between rheumatologists was improved by feedback from CJA in one small study<sup>13</sup> is encouraging and invites further progress in this area.

We thank Dr C G Barnes, Dr P Davies, and Dr D W James for their help with this project. Dr J R Kirwan was supported by the Arthritis and Rheumatism Council for Research.

#### References

- 1 Kirwan J R, Chaput de Saintonge D M, Joyce C R B, Currey H L F. Clinical judgment in rheumatoid arthritis. II. Judging 'current disease activity' in rheumatoid arthritis. *Ann Rheum Dis* 1983; **42**: 648-51.
- 2 Kirwan J R, Chaput de Saintonge D M, Joyce C R B, Currey H L F. Clinical judgment in rheumatoid arthritis. III. British rheumatologists' judgments of 'change in response to therapy'. *Ann Rheum Dis* 1984; **43**: 686-94.
- 3 Kirwan J R, Currey H L F. Clinical judgment in rheumatoid arthritis. IV. Rheumatologists' judgments are stable over long periods. *Ann Rheum Dis* 1984; **43**: 695-7.
- 4 Kirwan J R, Currey H L F, Brooks P M. Measuring physicians' judgment—the use of clinical data by Australian rheumatologists. *Aust NZ J Med* (in press).
- 5 Kirwan J R, Chaput de Saintonge D M, Joyce C R B, Currey H L F. Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of paper patients. *Ann Rheum Dis* 1983; **42**: 644-7.
- 6 Kirwan J R. Social judgment theory and rheumatic diseases. Application of clinical judgment analysis to rheumatoid arthritis. University of London, 1983. (MD thesis).
- 7 Anderson G. Measurement of disease activity in rheumatoid arthritis: a modeling approach. University of Toronto, 1980. (MSc thesis).
- 8 Dawes R M, Corrigan B. Linear models in decision making. *Psychol Bull* 1974; **81**: 95-106.
- 9 Bombardier C, Tugwell P, Sinclair A, Dok C, Anderson G, Buchanan W W. Preference for endpoint measures in clinical trials: results of structured workshops. *J Rheumatol* 1982; **9**: 798-801.
- 10 Byron M A, Mowat A G. Corticosteroid prescribing in rheumatoid arthritis—the fiction and the fact. *Br J Rheumatol* 1985; **24**: 164-6.
- 11 Hammond K R, Stewart T R, Brehmer B, Steinman D O. Social judgment theory. In: Kaplan M F, Schwartz S, eds. *Human judgment and decision processes*. New York: Academic Press, 1975.
- 12 Hammond K R, Summers D A. Cognitive control. *Psychol Rev* 1972; **79**: 58-67.
- 13 Kirwan J R, Chaput de Saintonge D M, Joyce C R B, Currey H L F. Clinical judgment analysis—practical application in rheumatoid arthritis. *Br J Rheumatol* 1983; **22** (suppl): 18-23.