ANALYSIS OF IMPROVEMENT IN INDIVIDUAL RHEUMATOID ARTHRITIS PATIENTS TREATED WITH DISEASE-MODIFYING ANTIRHEUMATIC DRUGS, BASED ON THE FINDINGS IN PATIENTS TREATED WITH PLACEBO

HAROLD E. PAULUS, MARLENE J. EGGER, JOHN R. WARD, H. JAMES WILLIAMS, and the COOPERATVE SYSTEMATIC STUDIES OF RHEUMATIC DISEASES GROUP

A composite index for estimating improvement in individual rheumatoid arthritis (RA) patients during trials of slow-acting, disease-modifying antirheumatic drugs (DMARDs) was developed by analyzing the responses of 130 placebo-treated participants in Cooperative Systematic Studies of Rheumatic Diseases studies. If responses in 4 of 6 selected measures were required for improvement (by ≥20% for morning stiffness, Westergren erythrocyte sedimentation rate, joint pain/tenderness score, and joint swelling score, and by ≥2 grades on a 5-grade scale, or from grade 2 to grade 1 for patient's and physician's overall assessments of current disease severity), few placebo-treated patients qualified as improved, whereas significantly more DMARD-treated patients demonstrated improvement. The proposed index appears to be useful in estimating the probability that an RA patient will improve if taking a placebo during a DMARD trial, and may be a useful tool for analysis of DMARD studies.

From the Cooperative Systematic Studies of Rheumatic Diseases group.

Supported by NIH contracts NIAMDD NO1-AM-6-2218 and NO1-AM-1-2204, and by the Ciba-Geigy, Parke-Davis, and Merck Sharpe and Dohme Companies.

Harold E. Paulus, MD: Professor of Medicine, Division of Rheumatology, University of California, Los Angeles School of Medicine; Marlene J. Egger, PhD: Associate Professor of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City; John R. Ward, MD: Professor of Medicine, Division of Rheumatology, University of Utah School of Medicine, Salt Lake City; H. James Williams, MD: Professor of Medicine, Division of Rheumatology, University of Utah School of Medicine, Salt Lake City.

Address reprint requests to Harold E. Paulus, MD, UCLA School of Medicine, Division of Rheumatology, 1000 Veteran Avenue, 32-47, Los Angeles, CA 90024-1670.

Submitted for publication May 30, 1989; accepted in revised form November 21, 1989.

Although the essential human unit in medicine is the individual patient, the responses of individual patients with rheumatoid arthritis (RA) have been difficult to ascertain from published reports of controlled clinical trials with both nonsteroidal antiinflammatory drugs (NSAIDs) and slow-acting, disease-modifying antirheumatic drugs (DMARDs). Rather, the focus of these reports has been the statistical differentiation of 1 therapeutic intervention from another, or from placebo. The traditional assessments of RA attempt to measure some of the cardinal signs of inflammation: swelling, pain, and loss of function. Redness and warmth usually are not quantitated (1). Thus, traditional clinical assessments of joint pain, tenderness, swelling, range of motion, and circumference, grip strength, walking time, morning stiffness, and erythrocyte sedimentation rate (ESR), or other "acute-phase reactants" semiquantitatively record various aspects or effects of the inflammatory process. This is also done by the newer instruments to assess function, health status, and quality of life (2).

The application of standard statistical analysis to each of these measures can determine whether it has been changed more by an intervention than by placebo, or by a comparison intervention. If all of these independently analyzed, but interrelated, measures demonstrate statistical superiority of the tested intervention, one concludes that it is more effective than the comparison treatment, e.g., methotrexate is more effective than placebo (3). However, even in this situation, some patients fail to benefit from the treatment. If only some measures are statistically superior, there is more difficulty deciding whether the treatment is effective.

In some studies, the patient and the investigator

Table 1. Characteristics of *placebo-treated* patients in 4 CSSRD studies on disease-modifying antirheumatic drugs for the treatment of rheumatoid arthritis*

	D-penicillamine	Oral gold/ parenteral gold	Methotrexate	Subtotal	Sulfasalazine/ parenteral gold	Total
Associated with study	55	55	105	215	53	268
At least 1 visit	54	51	98	203	52	255
Eligible after first visit	54	50	94	198	51	249
Withdrawals	14	7	41	62	19	81
Adverse effects	2	1	10	13	5	18
Lack of efficacy	5	2	20	27	10	37
Other reasons	7	4	11	22	4	26
Completing study	40	43	53	136	32	168
Observer changed V_1 to V_L	1	0	5	6	0	6
Efficacy group for joint scores†	39	43	48	130	32	162

* Values are the number of placebo-treated patients. CSSRD = Cooperative Systematic Studies of Rheumatic Diseases; V_1 = baseline visit; V_L = end-of-study visit.
† Values are the same as in previous studies (3,9,11). In the D-penicillamine study (8), 2 placebo-treated patients were eligible at the first visit, but were not included in the results.

are asked to record their overall opinions regarding the efficacy of the treatment. These opinions can provide an intuitive indication of the clinical responses of individual patients, but do not use the carefully recorded traditional clinical assessments. Various composite indices have been proposed; examples include the Lansbury index (4), the Ritchie articular index (5), the pooled index (6), discriminant analysis (7), the Cooperative Systematic Studies of the Rheumatic Diseases (CSSRD) joint count (8), and others (1).

Responses of individual patients participating in a clinical trial may vary from marked improvement to withdrawal for lack of efficacy. Since patients are treated individually, a method of analysis based on individual patient responses should make it easier to extrapolate the results of a controlled clinical trial to the clinical treatment setting.

Reasoning that the responses of the placebo-treated patients in controlled clinical trials of DMARDs represent the background progression of their disease (during treatment with NSAIDs and in some cases with low doses of prednisone), one could attempt to develop a criterion for improvement of individual patients that could be satisfied by no more than 5% of placebo-treated patients. If the responses of a significantly greater percentage of drug-treated patients exceed this criterion, this should indicate that the drug is useful for the treatment of at least some patients with RA. Such a measure could be of substantial value in screening investigational drugs for further research.

PATIENTS AND METHODS

Patients. The CSSRD group has conducted a series of prospective, randomized, double-blind, controlled studies of DMARDs in patients with RA. These studies include 1) comparison of D-penicillamine 500 mg/day, D-penicillamine 125 mg/day, and placebo (8), 2) comparison of auranofin, aurothiomalate, and placebo (9), 3) comparison of D-penicillamine 750 mg/day and azathioprine (10), and 4) comparison of low-dose weekly pulse methotrexate and placebo (3).* A total of 198 patients were randomized to take placebo in these studies (Table 1), passed all study eligibility checkpoints (i.e., were still eligible at the initial visit), and

* The participating clinics for the 5 studies (including the present one), and their clinic directors, were as follows: Boston University, Boston, MA, Edgar S. Cathcart, MD and Robert F. Meenan, MD; Duke University, Durham, NC, Richard P. Polisson, MD; Guthrie Clinic, Sayre, PA, Robert M. Michaels, MD; Johns Hopkins University at Good Samaritan Hospital, Baltimore, MD, Lynn M. Billingsley, MD; Medical College of Virginia/Virginia Commonwealth University, Richmond, Ralph E. Small, PharmD; Arthritis and Rheumatism Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD, Paul H. Plotz, MD and John Klippel MD; Ohio State University, Columbus, Seth M. Kantor, MD; State University of New York, Downstate Medical Center, Brooklyn, David Kaplan, MD and Joyce Z. Singer, MD; University of Alabama, Birmingham, Graciela S. Alarcón, MD; University of California, Los Angeles, Harold E. Paulus, MD; University of California, San Diego, Michael H. Weisman, MD; University of Cincinnati, Cincinnati, OH, Michael E. Luggen, MD; University of Connecticut, Farmington, Arthur Weinstein, MD; University of Michigan, Ann Arbor, Giles G. Bole, MD and William Mikkelson, MD; University of Missouri, Columbia, Gordon C. Sharp, MD; University of South Alabama, Mobile, Joseph G. Hardin, Jr., MD; University of Tennessee, Memphis, Stanley B. Kaplan, MD; University of Utah, Salt Lake City, Cecil O. Samuelson, Jr., MD and Daniel O. Clegg, MD; University of Washington, Harborview Medical Center, Seattle, Robert F. Willkens, MD.

Table 2. Summary of *placebo-treated* patient data base used for analysis of individual improvement*

	D-penicillamine	Oral gold/ parenteral gold	Methotrexate	Subtotal	Sulfasalazine/ parenteral gold	Total
Completed study	40	43	53	136	32	168
Withdrew for lack of efficacy	5	2	20	27	10	37
Available for analysis	45	45	73	163	42	205
Observer changed $V_I - V_L$	1	0	5	6	0	6
Missing values†	14	12	1	27	0	27
Total analyzed	30	33	67	130	42	172

* Values are the number of placebo-treated patients. V_I = baseline visit; V_L = end-of-study visit.
† If any of the paired values for baseline and end-of-study visits were missing for any of the 6 variables used in the analysis, the patient was not used in the analysis of individual improvement.

completed the initial visit. (A few additional patients were found to be ineligible at the first visit, but received some medication, due to late submission of forms to the data center.) Of these 198 patients, 27 withdrew for lack of efficacy and 136 completed the respective studies; 130 of 163 patients who completed their study or dropped out for lack of efficacy were analyzed for the present investigation (Table 2). Thirty-three patients who had missing values, observer changes, or who withdrew for reasons other than lack of efficacy were excluded from this analysis.

Efficacy evaluations. The standard efficacy assessment in each study included measurements of morning stiffness, CSSRD joint pain/tenderness score, CSSRD joint swelling score, patient's overall assessment of current disease severity, physician's overall assessment of current disease severity, and Westergren ESR (8). These 6 measures were arbitrarily selected for use in the development of a criterion for individual improvement.

Baseline evaluations of the patients randomized to receive placebo in the 3 studies differed significantly in a number of respects. Table 3 details this lack of comparability for those placebo-treated patients completing these studies. It is not surprising that these groups are noncomparable, since entry criteria such as previous treatment with other DMARDs differed among the studies.

Statistical analysis. In developing a criterion for improvement of individual RA patients, consideration was restricted to the 6 variables discussed above: morning stiffness, CSSRD joint pain/tenderness score, CSSRD joint swelling score, Westergren ESR, patient's overall assessment of current disease severity, and physician's overall assessment of current disease severity. These data were

Table 3. Comparison of *placebo-treated* group clinical mean values at baseline among 3 studies, for patients who were eligible at and completed the first visit*

Variable	Study			*P*†			
	D-penicillamine (n = 54)	Oral gold (n = 50)	Methotrexate (n = 94)	Overall	DP versus oral gold	DP versus MTX	Oral gold versus MTX
Age	51.0 (52)	48.3 (50)	55.0 (94)	0.0209	0.433	0.052	0.012
Severe disease, %	58 (52)	20 (50)	57 (94)	0.0000	0.001	0.657	0.000
Disease duration, months	116.2 (52)	60.9 (50)	156.1 (94)	0.0000	0.000	0.006	0.000
Functional class	2.5 (52)	2.2 (50)	2.4 (94)	0.0113	0.008	0.662	0.007
Physician assessment	3.2 (52)	2.9 (50)	3.4 (94)	0.0064	0.041	0.375	0.002
Patient assessment	3.0 (52)	3.1 (50)	3.2 (94)	0.1906	–	–	–
Grip strength (right), mm Hg	91.1 (51)	125.1 (50)	90.1 (92)	0.0003	0.003	0.314	0.000
Grip strength (left), mm Hg	91.7 (51)	118.3 (50)	89.0 (92)	0.0029	0.021	0.294	0.001
Joint pain/tenderness score	56.0 (52)	46.6 (50)	55.6 (94)	0.1502	–	–	–
Joint swelling score	37.7 (52)	28.2 (50)	38.5 (94)	0.0120	0.006	0.602	0.011
Joint pain/tenderness count	36.3 (52)	29.9 (50)	33.5 (94)	0.0258	0.010	0.195	0.062
Joint swelling count	26.6 (52)	20.5 (50)	25.2 (94)	0.0131	0.003	0.155	0.068
Morning stiffness, minutes	253.1 (51)	233.1 (45)	205.6 (90)	0.7400	–	–	–

* Values in parentheses are the numbers of patients with valid observations for the indicated variable. DP = D-penicillamine; MTX = methotrexate; – = no significant difference.
† Overall tests for differences among studies were performed using the Kruskal-Wallis nonparametric test. Pairwise comparisons between studies were determined using the Mann-Whitney nonparametric test.

Table 4. Percentage of patients with individual improvement (≥x% improvement in ≥4 variables), incorporating withdrawals for lack of efficacy

Study	n	% improvement ≥20	≥30	≥40	≥50	≥60
D-penicillamine						
125 mg/day	51	18	14	10	10	6
500 mg/day	43	33	23	21	19	19
Placebo	30	3	3	3	0	0
Gold						
Oral gold	56*	27	25	16	14	9
Gold sodium thiomalate	46	37	26	15	15	11
Placebo	33	12	12	9	6	3
Methotrexate						
Drug	57	39	30	30	19	16
Placebo	67	4	4	3	3	3
Azathioprine						
Drug	58	24	17	16	14	12
D-penicillamine 750 mg/day	53	45	42	38	32	26
All placebo patients	130†	6.2	6.2	4.6	3.1	2.3

* Includes 1 patient taking oral gold who withdrew at the third visit because of a flare.
† See Tables 1 and 2 for details.

available for a large number of patients across all studies, in contrast to, for example, measures of activities of daily living. For each variable, the baseline value at the beginning of the study was compared with the value observed at the end of the study to determine the degree of improvement or deterioration that occurred during the study. Patients who withdrew for lack of efficacy did not have end-of-study observations, but were considered to have demonstrated zero improvement for each of the 6 variables.

The percentage of improvement was calculated directly for each of the first 4 variables. However, patient assessment and physician assessment of current disease severity were expressed on a 1–5 scale (1 = asymptomatic, 2 = mild, 3 = moderate, 4 = severe, and 5 = very severe). These 2 variables were accepted as improved only if they satisfied the CSSRD's definition for "important improvement" (3), i.e., improvement by at least 2 grades, or if initially grade 2, improvement to grade 1.

Using the 6 selected variables and calculating improvement as defined above, preliminary analysis was performed to determine the proportions of placebo-treated patients who would be considered to have individual improvement if various criteria were applied. Preliminary definitions required ≥x% improvement in ≥y of the 6 variables, where x was set at 20%, 30%, 40%, 50%, or 60% improvement, and y was set at 2, 3, 4, 5, or 6 variables. Similar calculations were performed for the patients receiving DMARDs in the various studies.

We then tested the individual improvement criterion selected, ≥20% improvement in ≥4 variables, by retrospectively applying it to the 4 completed CSSRD studies used to derive the criterion, plus a fifth study (sulfasalazine versus gold sodium thiomalate versus placebo [11]), to see whether

Table 5. Individual improvement of patients in 4 CSSRD studies*

	Methotrexate		Gold			Azathioprine/D-penicillamine		D-penicillamine		
	Placebo	Drug	Placebo	Oral gold	Gold sodium thiomalate	Azathioprine	D-penicillamine (750 mg/day)	Placebo	Drug (125 mg/day)	Drug (500 mg/day)
Improved, no. (%)	3 (4)	22 (39)†	4 (12)	15 (27)‡	17 (37)	14 (24)	24 (45)§	1 (3)¶	9 (18)	14 (33)
Not improved, no.	64	35	29	41#	29	44	29	29	42	29
Total**	67	57	33	56#	46	58	53	30	51	43
Withdrawals										
Toxicity, no. (%)	10 (11)	29 (31)	1 (2)	4 (5)	20 (24)	20 (19)	29 (28)	2 (4)	9 (10)	18 (21)
Other, no.	11	6	4	8	8	11	7	7	6	4
Total starting study††	94	95	50	81	82	103	102	54	88	86
Improved, % of total starting study	3	23	8	19	21	14	24	2	10	16

* Individual improvement is defined as improvement in at least 4 of the following variables: morning stiffness, Westergren erythrocyte sedimentation rate, Cooperative Systematic Studies of Rheumatic Diseases (CSSRD) joint pain/tenderness score, CSSRD joint swelling score, physician's assessment of disease severity, and patient's assessment of disease severity. The first 4 variables are required to show at least 20% improvement relative to baseline, and the last 2 are required to show "important" improvement by the definition of Williams et al (3).
† $P = 0.00001$ versus placebo.
‡ $P = 0.0483$ versus placebo and versus gold sodium thiomalate.
§ $P = 0.0320$ versus azathioprine.
¶ $P = 0.0073$ versus low-dose D-penicillamine and versus high-dose D-penicillamine.
# Includes 1 patient with a disease flare at the third visit.
** The total number of patients excludes patients with missing values or observer changes for any of the 6 variables used to define individual improvement. Also excluded are patients who withdrew for reasons other than lack of efficacy.
†† This group was eligible at and completed the first visit (see Table 1).

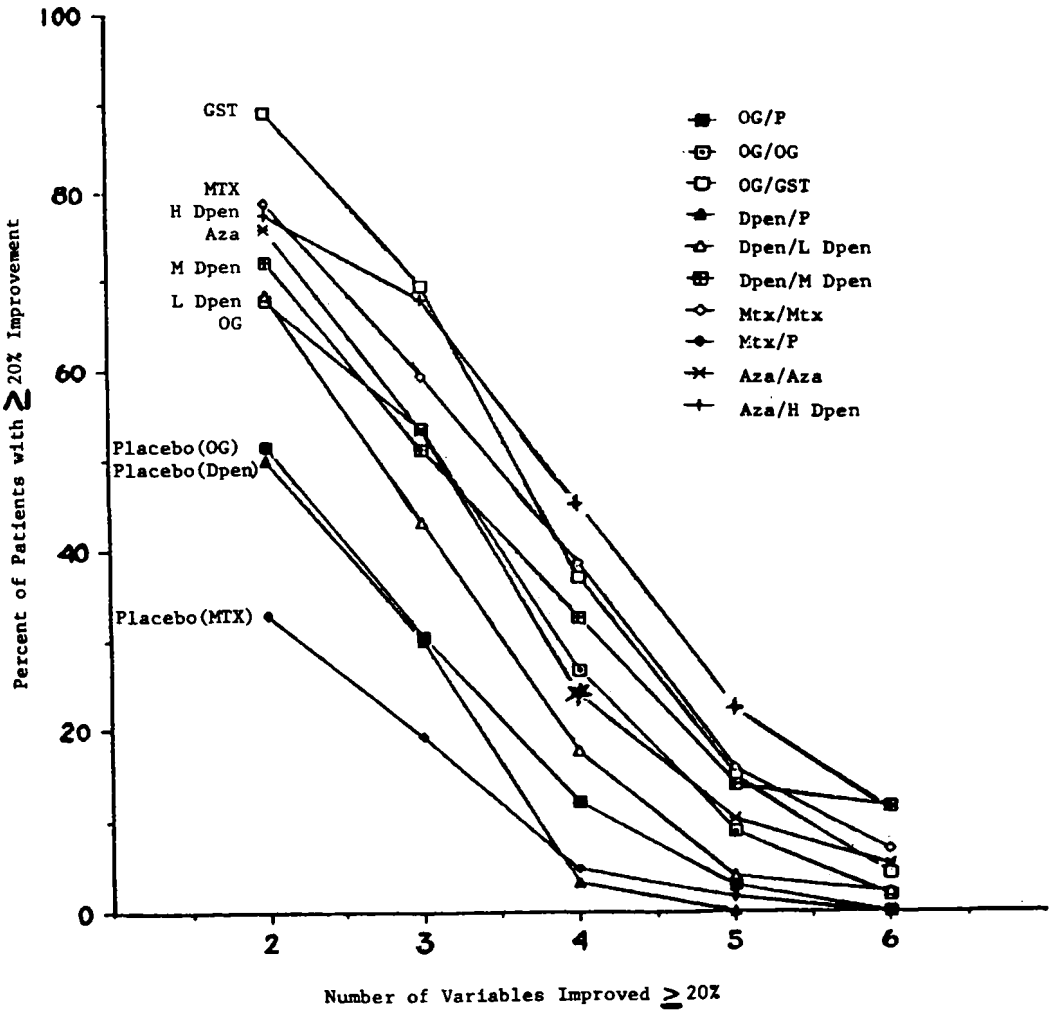

Figure 1. Percentages of patients receiving various drugs or placebo in 4 Cooperative Systematic Studies of Rheumatic Diseases studies who demonstrated ≥20% improvement in 2, 3, 4, 5, or 6 of the measured variables. See Patients and Methods for definitions of variables. GST = gold sodium thiomalate; MTX = methotrexate; H Dpen = 750 mg/day D-penicillamine; Aza = azathioprine; M Dpen = 500 mg/day D-penicillamine; L Dpen = 125 mg/day D-penicillamine; OG = oral gold; P = placebo.

the criterion could distinguish placebo from these standard, accepted DMARDs, while still detecting the required low placebo response rate.

RESULTS

By inspection, a definition incorporating ≥20% improvement in 4 or more variables seemed to produce the largest difference between placebo-treated patients and drug-treated patients, while retaining an acceptably low proportion of placebo-treated patients with individual improvement (Table 4, where y = 4). Results for other values of y were also calculated (data not shown). The average response rate of the placebo-treated patients was 6.2%, with 3%, 4%, and 12% of the placebo-treated patients in the D-penicillamine, methotrexate, and gold studies, respectively, satisfying this criterion for individual improvement, whereas the proportions of drug-treated patients who met this criterion ranged from 18% to 45% (Table 4). Requiring ≥40% or ≥50% improvement in 4 or more variables decreased the proportion of placebo responders, but also substantially decreased the proportions of patients who would be considered to be improved by the various drugs (Table 4). Similarly, if one required ≥20% improvement in ≥5, or in all 6 variables, the proportion of placebo responders decreased, but at the

Table 6. Individual improvement of patients in a CSSRD study on the treatment of rheumatoid arthritis with sulfasalazine, gold sodium thiomalate, and placebo*

	Sulfasalazine	Gold sodium thiomalate	Placebo
Improved, no. (%)†	28 (52.8)	17 (51.5)	12 (28.6)
Not improved, no. (%)†	25 (47.2)	16 (48.5)	30 (71.4)
Total	53	33	42
No. of withdrawals			
Lack of efficacy	6	3	10
Toxicity	11	27	5
Administrative	5	6	4
Total starting study	69	66	51
Improved, % of total starting study	41	26	24

* See Table 5 for definition of individual improvement. Not improved includes withdrawals for lack of efficacy. There were no missing values or observer changes for these data. CSSRD = Cooperative Systematic Studies of Rheumatic Diseases.
† $P = 0.04$, drug versus placebo; $P = 0.03$, sulfasalazine versus placebo; $P = 0.074$, gold sodium thiomalate versus placebo; $P = 0.918$, sulfasalazine versus gold sodium thiomalate.

expense of decreased responsiveness of the drug-treated patients. No placebo patients had ≥20% improvement in all 6 variables, but only 2–12% of the drug-treated patients were responders using this criterion (Figure 1).

Table 5 displays the tests of drugs and placebo for the 4 "training" studies. As expected, this criterion significantly differentiated placebo from active drugs in the methotrexate, gold, and low-dose D-penicillamine studies, but the data from these studies had been used to develop the criterion. However, it also indicates that D-penicillamine (750 mg/day) was significantly more effective than azathioprine, a difference that had not been detected in the original statistical analysis of that study (Table 5).

The selected criterion for individual improvement was tested prospectively in the analysis of the CSSRD study of sulfasalazine versus gold sodium thiomalate versus placebo in RA (11). In that study, traditional analysis had failed to demonstrate significant differences between the drug- and placebo-treated groups, apparently because of a better-than-usual degree of improvement in the placebo-treated patients. However, analysis of improvement in individuals, based on the proposed criterion, suggested that placebo-treated patients were significantly less likely to improve compared with drug-treated patients ($P = 0.04$) (Table 6). Further analysis indicated that individuals treated with sulfasalazine were more likely to demonstrate improvement than those treated with placebo ($P = 0.03$), and that gold sodium thiomalate was marginally better than placebo ($P = 0.074$), but there was no difference between the 2 drugs ($P = 0.918$).

DISCUSSION

The lack of a generally accepted method for estimating improvement in individual RA patients during trials of DMARDs hampers the statistical analysis of controlled clinical trials and makes it difficult to extrapolate the reported results of these trials to the treatment of individual patients. Statisticians lament the large number of somewhat related efficacy variables measured in these trials, and puzzle over how to interpret a trial in which some variables show significant improvement but other related variables do not. In the report of the CSSRD study of methotrexate versus placebo (3), 13 clinical efficacy variables and 4 laboratory efficacy variables were subjected to separate statistical analysis: morning stiffness, right hand grip strength, left hand grip strength, right hand proximal interphalangeal (PIP) joint circumference, left hand PIP joint circumference, walking time, pain analog scale, patient assessment, physician assessment, joint pain/tenderness count and score, joint swelling count and score, hemoglobin level, platelet count, ESR, and rheumatoid factor. Clinicians who read reports of these analyses, and even the scientists who write them, often have difficulty interpreting the clinical importance of the reported mean or median changes in the various measurements of efficacy, and have difficulty applying the statistical conclusions of the trial to the management of their patients. Clinicians would like to know the probability of a good response in a patient about to start DMARD treatment, but published reports usually present only average or median responses of the variables studied, and the analysis generally applies only to those patients who finished the study.

The criterion for measuring individual improvement described in this report was empirically selected so that only a few placebo-treated patients would demonstrate improvement with it. Thus, individual improvement was detected in only 6.2% of 130 placebo-treated patients who were used to develop the criterion. However, patients entering different studies had substantially different characteristics. Thus, the proportion of placebo-treated patients with individual improvement varied from 3% to 12% in the 3 index studies, but was 28.6% in the sulfasalazine/gold so-

dium thiomalate study. Therefore, it may be difficult to compare the unadjusted findings of one study with those in another study. The reader is specifically cautioned against ranking active drugs tested in different studies. Nevertheless, the proportions of patients demonstrating individual improvement with the DMARDs used in the various studies ranged from 18% (D-penicillamine 125 mg/day) in one study to 45% (D-penicillamine 750 mg/day) in another study, and are comparable with the proportion of patients who have been reported to continue to take a DMARD for more than 1 year (12–14).

A liberal definition of individual improvement could be met by large proportions of both placebo- and drug-treated patients. For example, if individual improvement was defined as ≥20% improvement in ≥2 variables, in the oral gold study, 51.5% of the placebo-treated patients, 68% of the oral gold–treated patients, and 89% of the aurothiomalate-treated patients would satisfy this criterion (Figure 1). With a stringent definition, very few patients qualified; for example, if we required ≥30% improvement in all 6 variables, only 4% of aurothiomalate-treated patients, and no patients treated with oral gold or placebo could satisfy this definition of individual improvement (data available on request). Discriminant analysis differentiating drug groups was not performed since placebo responders were found in every study. In addition, discriminant analysis does not provide a decision rule with clinically meaningful cutoff points.

The present analysis includes only those patients for whom a decision about efficacy could be made, i.e., those who completed the study and those who withdrew because of lack of benefit. In evaluating the clinical usefulness of a drug, its toxicity must also be considered as part of the risk/benefit ratio. Withdrawals for toxicity ranged from 2% (placebo-treated patients in the oral gold study) to 31% (methotrexate-treated patients) in the studies evaluated (Table 5). The physician and the patient must weigh the probability of withdrawal for toxicity and the severity of toxic events, against the probability of improvement, when deciding which DMARD to use. The proportions of patients who withdrew because of toxicity vary somewhat among the studies analyzed. For example, 24% of the patients treated with gold sodium thiomalate in the oral gold study withdrew because of toxicity (Table 5), whereas 41% of the gold sodium thiomalate–treated patients withdrew because of toxicity in the sulfasalazine study (Table 6). The toxicity of D-penicillamine appeared to be dose related; withdrawals for toxicity were 10%, 21%, and 28% with doses of 125 mg/day, 500 mg/day, and 750 mg/day, respectively (Table 5).

The proposed criterion for individual improvement of RA patients participating in DMARD studies may be helpful in performing "intent to treat" analyses of these studies. In the "intent to treat" analysis of drug efficacy, the number of patients demonstrating improvement is expressed as a percentage of those who started the study (15). This analysis includes all withdrawals, even those who moved from the area, those who withdrew for other presumably non–drug-related reasons, and those who were not analyzed because of protocol violations, and it most conservatively expresses the probability that a patient starting a drug will benefit from it. Using the proposed criterion for individual improvement in an "intent to treat" analysis, only 8 of 198 patients (4%) starting placebo treatment demonstrated improvement. However, only 10% (those taking D-penicillamine 125 mg/day) to 24% (those taking D-penicillamine 750 mg/day) of patients starting DMARD treatment actually achieved improvement, when the analysis reflected the inclusion of withdrawals for toxicity and administrative reasons (Table 5). This is perhaps unduly pessimistic because we could not calculate individual improvement for patients with missing values, or changes in observer between the baseline and final visits (who actually may have demonstrated improvement if not for this protocol violation). However, it clearly indicates the great need for DMARDs with better risk/benefit ratios.

The methods that we have proposed in this report will be applied prospectively in the analysis of future CSSRD studies and can be verified by independent investigators as well, either prospectively or by reanalysis of completed DMARD trials. Inclusion of an assessment of individual improvement in future reports should help clinicians estimate the probability of a good response in a patient who is about to start treatment, and make it easier to inform the patient about the potential advantages and disadvantages of the treatment. The responses of placebo-treated patients provide a useful standard against which to judge the responses of drug-treated patients. It is conceivable that the placebo-treated patient responses can be used as historic controls for initial studies of the efficacy of new drugs, in which it may be inappropriate to use a placebo-treated group (16). If a sufficiently large data base of placebo-treated patients can be gathered, it may be possible to match important char-

acteristics of the drug-treated patients in these initial efficacy studies to those of historic placebo controls selected from the data base, perhaps compensating for the between-study variability in patient characteristics.

ACKNOWLEDGMENTS

We thank the participating clinics and the clinic directors who performed the 5 studies, and Larry G. Thorderson for statistical assistance.

REFERENCES

1. Buchanan WW, Tugwell P: Traditional assessments of articular diseases. Clin Rheum Dis 9:515–529, 1983
2. Liang MG, Cullen KE, Larson MG: Measuring function and health status in rheumatic disease clinical trials. Clin Rheum Dis 9:531–539, 1983
3. Williams HJ, Willkens RF, Samuelson CO Jr, Alarcón GS, Guttadauria M, Yarboro C, Polisson RP, Weiner SR, Luggen ME, Billingsley LM, Dahl SL, Egger MJ, Reading JC, Ward JR: Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis: a controlled clinical trial. Arthritis Rheum 28:721–730, 1985
4. Lansbury J: Report of a three-year study on the systemic and articular indexes in rheumatoid arthritis: theoretic and clinical considerations. Arthritis Rheum 1:505–522, 1958
5. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakos TG, Grieveson P, Buchanan WW: Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. Q J Med 37:393–406, 1968
6. Smythe HA, Helewa A, Goldsmith CH: "Independent assessor" and "pooled index" as techniques for measuring treatment effects in rheumatoid arthritis. J Rheumatol 4:144–152, 1977
7. McQuire RJ, Wright V: Statistical approaches to indices of disease activity in rheumatoid arthritis. Ann Rheum Dis 30:574–580, 1971
8. Williams HJ, Ward JR, Reading JC, Egger MJ, Grandone JT, Samuelson CO, Furst DE, Sullivan JM, Watson MA, Guttadauria M, Cathcart ES, Kaplan SB, Halla JT, Weinstein A, Plotz PH: Low-dose D-penicillamine therapy in rheumatoid arthritis: a controlled, double-blind clinical trial. Arthritis Rheum 26:581–592, 1983
9. Ward JR, Williams HJ, Egger MJ, Reading JC, Boyce E, Altz-Smith M, Samuelson CO Jr, Willkens RF, Solsky MA, Hayes SP, Blocka KL, Weinstein A, Meenan RF, Guttadauria M, Kaplan SB, Klippel J: Comparison of auranofin, gold sodium thiomalate, and placebo in the treatment of rheumatoid arthritis: a controlled clinical trial. Arthritis Rheum 26:1303–1315, 1983
10. Paulus HE, Williams HJ, Ward JR, Reading JC, Egger MJ, Coleman ML, Samuelson CO Jr, Willkens RF, Guttadauria M, Alarcón GS, Kaplan SB, MacLaughlin EJ, Weinstein A, Wilder RL, Solsky MA, Meenan RF: Azathioprine versus D-penicillamine in rheumatoid arthritis patients who have been treated unsuccessfully with gold. Arthritis Rheum 27:721–727, 1984
11. Williams HJ, Ward JR, Dahl SL, Clegg DO, Willkens RF, Oglesby T, Weisman MH, Schlegel S, Michaels RM, Luggen ME, Polisson RP, Singer JZ, Kantor SM, Shiroky JB, Small RE, Gomez MI, Reading JC, Egger MJ: A controlled trial comparing sulfasalazine, gold sodium thiomalate, and placebo in rheumatoid arthritis. Arthritis Rheum 31:702–713, 1988
12. Dwosh IL, Stein HB, Urowitz MB, Smythe HA, Hunter T, Ogryzlo MA: Azathioprine in early rheumatoid arthritis: a comparison with gold and chloroquine. Arthritis Rheum 20:685–692, 1977
13. Hoh K-F, Dwosh IL, Ford PM, Anastassiades TP, Kelley HG: Drug-related disease remissions in rheumatoid arthritis (abstract). Arthritis Rheum 24 (suppl 4): S72, 1981
14. Currey HLF, Harris J, Mason RM, Woodland J, Beveridge T, Roberts CJ, Vere DW, Dixon AStJ, Davies J, Owen-Smith B: Comparison of azathioprine, cyclophosphamide, and gold in treatment of rheumatoid arthritis. Br Med J 3:763–766, 1974
15. Kirwan JR, Currey HLF: Rheumatoid arthritis: disease-modifying anti-rheumatic drugs. Clin Rheum Dis 9:581–599, 1983
16. Paulus HE: Pilot efficacy studies—identifying their role in drug development and rationale for their use: the Paulus model, Proceedings: Early Decisions in DMARD Development. Atlanta, Arthritis Foundation, 1989