Methotrexate, Hydroxychloroquine, and Intramuscular Gold in Rheumatoid Arthritis: Relative Area Under the Curve Effectiveness and Sequence Effects

STACEY HURST, MICHAEL J. KALLAN, FREDERICK J. WOLFE, JAMES F. FRIES, and DANIEL A. ALBERT

ABSTRACT. ***Objective.*** The use of disease modifying antirheumatic drugs (DMARD) for rheumatoid arthritis (RA) is predicated on the expected value of the treatment course. Most clinical data are generalized from randomized controlled trials (RCT), which may result in estimates that are discordant with clinical experience and cannot address the effects of sequence of drugs. We computed estimates of relative DMARD effectiveness from a large observational database using area under the curve (AUC) data.

Methods. We examined data collected over a 20 year period on 1160 patients who were followed at the Wichita Arthritis Center. We utilized Health Assessment Questionnaire (HAQ) disability index data to quantify the effect of methotrexate (MTX), hydroxychloroquine (HCQ), and injectable gold (gold) on subsequent patient outcome. Using an AUC analysis, we compared length of treatment course, total disability averted, annual disability averted, and percentage of possible disability averted across drugs, and examined differences between first courses of therapy in DMARD naïve patients and subsequent courses of the same and different DMARD in patients.

Results. Patients treated with MTX, HCQ, and gold improved at a rate of –0.33, –0.18 and –0.38 annualized HAQ area units, respectively. Since duration taking drug was greatest for MTX, then HCQ, then gold, the cumulative improvement was greatest with MTX (–1.07) versus gold (–0.74) versus HCQ (–0.47) in disability unit years. All 3 drugs were better cumulatively with earlier disease (MTX –1.74 for < 1 yr vs –0.95 for > 1 yr; HCQ –0.68 vs –0.43; gold –1.71 vs –0.49). A second trial of the same drug was far less effective than the first course. On a percentage of possible improvement basis, these drugs were nearly equal since HCQ is given to less severely affected patients.

Conclusion. MTX cumulatively is the most effective DMARD of these 3 because of the length of the therapeutic segment. In terms of disability averted, none of the agents decrease disability by more than 25% of the theoretically possible improvement. We documented that effectiveness of RA treatment is a function of drug sequence, duration of disease, whether it is a first or second course, and severity of disease. None of these clinically relevant observations have emerged from clinical trials. These methodologic approaches provide important quantitative comparative data and will be useful in further assessment of the relative effectiveness of present and future DMARD. (J Rheumatol 2002;29:1639-45)

Key Indexing Terms:
RHEUMATOID ARTHRITIS THERAPY DMARD
AREA UNDER THE CURVE OBSERVATIONAL

From the Division of Rheumatology, University of Pennsylvania, Philadelphia, Pennsylvania; The National Databank for Rheumatic Disease, University of Kansas, Wichita, Kansas; and Stanford University, Stanford, California, USA.

Supported by a grant from the NIH to ARAMIS (Arthritis, Rheumatism, and Aging Medical Information System) AM43584 and a Clinical Research Grant from the Arthritis Foundation.

S. Hurst, Department of Rheumatology, University of Philadelphia, currently, Emory University, Atlanta, GA; M.J. Kallan, MS, Department of Rheumatology, University of Pennsylvania; F. Wolfe, MD, National Databank for Rheumatic Disease; J.F. Fries, MD, Stanford University; D.A. Albert, MD, Department of Rheumatology, University of Philadelphia.

Address reprint requests to Dr. D.A. Albert, University of Pennsylvania, Division of Rheumatology, 5 Maloney, Suite 504, 3600 Spruce Street, Philadelphia, PA 19104-4283.

Submitted August 8, 2000; revision accepted February 2, 2002.

Clinical management of rheumatoid arthritis (RA) is largely based on the results of small relatively short term, randomized controlled trials (RCT) of the efficacy of alternative therapeutic agents[1]. Based on these trials, rheumatologists must choose between a number of disease modifying drugs (DMARD) that include methotrexate (MTX)[2,3], injectable gold (gold)[4], sulfasalazine (SSZ)[5], cyclosporin A (CSA)[6], and the newer agents etanercept[7,8], infliximab[9-11], and leflunomide[12,13]. Several other agents that are useful include hydroxychloroquine (HCQ)[14], minocycline[15,16], staph protein A columns[17], and the immunosuppressive agents azathioprine (AZA)[18] and cyclophosphamide[19].

However, efficacy studies have been limited for a variety of reasons: (1) they cover too short a period to estimate

outcomes over an entire treatment course (i.e., they are right censored); (2) they do not permit estimation of cumulative effects [i.e., they seldom use cumulative or area under the curve (AUC) measures]; (3) they do not reflect the residual effects of prior therapy; (4) they do not examine repeated courses of the same therapy; (5) they underestimate the impact of compliance; (6) they do not account for the effects of co-morbidity; (7) they seldom have more than a single comparator; and (8) patient selection characteristics often limit the generalizability of the results to other patient groups. The clinician needing guidance in complex situations thus has little evidence upon which to base a strategy. Perhaps as a result, clinical effectiveness seems less robust than expected[20-23]. Newer strategic approaches to the treatment of RA emphasize early and consistent use of DMARD[24,25], often including use of combinations of agents. Accumulated trial data (RCT) support these concepts[26-28]. In an earlier study, we utilized 3 different estimates of drug effectiveness to develop a sequential drug strategy using Markov modeling techniques[29]. The major conclusions of that study were: (1) The most rapid way to achieve remission is to utilize the most effective agent first. With the DMARD we examined, this was MTX. (2) Based on published measures of efficacy, assuming each drug's effect is independent, most patients should be improved by the third or fourth agent. (3) There wasn't a striking difference in efficacy of the DMARD we investigated with the exception of MTX, which was substantially better. (4) If we modeled length of time on each DMARD, the pattern of DMARD use mimicked the sawtooth pattern that has been suggested as a strategic approach to RA treatment.

We examined treatment course therapeutic segments in 3 commonly used DMARD and developed an analytic methodology. We sought to test the results of our prior modeling study by examining a large cohort of patients followed for many years, with documentation of DMARD utilization and measures of drug effectiveness. We examined the following questions: (1) what is the relative cumulative effectiveness of several DMARD in common use; (2) is a DMARD therapeutic segment independent of prior treatment, or are there predictable sequence effects; (3) does the time between original diagnosis and DMARD use have consequences for the overall effectiveness of the DMARD used; (4) is a second course of the same DMARD as good as the first; and (5) are drugs given initially in DMARD-naive patients more efective than when used subsequently in these patients.

MATERIALS AND METHODS

Patient selection. Patients in this series represented a 100% sample of all patients with RA attending the Wichita Arthritis Center outpatient clinic in Wichita, Kansas, from July 1980 through February 1999[30]. These patients were seen as part of their ordinary clinical care. The details of this data set have been described[20]. All patients satisfied American College of Rheumatology criteria for RA[31]. The demographic characteristics of this patient cohort are similar to reported statistics of RA patients: 72% were female, mean age of 54.9 years, with disease duration of 6.7 years. These patients were predominantly (93.9%) Caucasian.

Demographic and clinical variables. Demographic and other variables were captured at each clinic visit, using a method of data capture and entry described by Wolfe[20,32-34]. Briefly, by detailed self-report questionnaire and interview, we questioned patients about changes in clinical status at each visit. The CLINHAQ questionnaire was administered at each clinic visit. This instrument contains the Health Assessment Questionnaire (HAQ) disability index[35,36], and a variety of other self-reported disease and clinical measures.

The original database consisted of 1853 patients with more than 26,000 observations. We restricted our analysis to patients entered and followed after July 1, 1980, to assure uniform availability of HAQ disability scores, resulting in a core set of 1160 evaluable patients with RA. For each of these patients disease duration, patient demographics, and medication information (nonsteroidals, corticosteroids, DMARD) was available. Patients could have been taking one or more of the following DMARD at any visit: MTX, HCQ, gold, AZA, SSZ, or D-penicillamine (D-Pen). There were also several other combinations of drugs taken, but the number of patients in most groups was small; therefore we focused on MTX, HCQ, and gold therapeutic segments.

Each patient visit with all corresponding data was considered one observation. New variables were created to analyze each patient's clinical course by treatment segments for each DMARD taken. We defined a significant treatment segment as the continuous use of a DMARD for 3 months or greater. There were gaps in DMARD use: a patient taking a particular DMARD for one or more visits might stop for a time and then start again. We considered a gap of ≤ 6 months for HCQ and gold, and a gap of ≤ 3 months for MTX, as one continuous treatment segment. Patients rarely stopped drugs before 3 months; thus observations here include virtually all patients who started a drug. If no DMARD was taken for a period of 3 months or longer it was considered a "no-drug segment."

The "therapeutic segment" concept was proposed by Fries[23] for ARAMIS studies (Arthritis, Rheumatism and Aging Medical Information System) in which assessments were by mailed questionnaire at 6 month intervals, but is equally applicable to clinical studies without fixed intervals between assessments. The concept is familiar as a "treatment course" and the term "segment" reminds us that RA treatment strategies generally include prior and subsequent treatments. The therapeutic segment method is one way of quantifying longitudinal data on effectiveness.

The prescribing physician (FW) generally used HCQ as the first drug when RA was judged to be mild. In patients with more severe disease, it was used as a subsequent drug when other DMARD failed. Gold was used in more severe cases from 1980 through 1983; thereafter it was replaced by MTX. MTX was originally used in severe RA as a second line drug, but became the first line DMARD for severe disease around 1990. It was used increasingly as the first line drug both in mild and severe RA during this decade. Because of the non-random assignment to treatment, roughly following the rules set out above, conclusions regarding the overall or relative effectiveness of the various treatments will have limitations as to validity, and should be accepted cautiously. While there were data available on the dose of drugs taken, these were in standard ranges and were titrated toward optimal levels for individual patients.

Several different subsets of the data were examined for the various analyses. One subset included information for each patient's first DMARD trial. Another subset consisted of data for the first time a given drug was used, whether the drug was the first DMARD or not. This subset was then examined to exclude first DMARD in order to analyze sequence effects. A third group contained data for the second or greater trial of a drug for all patients. Lastly, we examined the information for first DMARD use between patients who took their first DMARD within one year of RA diagnosis and those who took it later in the course of their disease.

The primary outcome measure we used was HAQ disability[37]. An eligible difference score was one in which there was an initial HAQ score

within the first 3 months of starting a treatment segment and a final HAQ score within 3 months of completing a treatment segment and before starting another DMARD or combination. The effect of a treatment segment was calculated by examining the AUC of HAQ disability (AUC-HAQ).

AUC-HAQ. AUC-HAQ is calculated by using the baseline HAQ score as a reference point, while assuming that no response is a flat line when HAQ score is plotted against time. It is described in mathematical terms as: AUC-HAQ = (area above baseline HAQ reference line) – (area below baseline HAQ reference line), and the magnitude of improvement or worsening is the difference of the AUC-HAQ from 0.

Annualized disability avoided. The AUC annualized disability avoided is the total disability averted, divided by the mean length of the segment in years.

Percentage of possible disability averted. The percentage of possible disability averted is the total disability averted divided by the total disability, if the baseline values were continued for the length of the segment × 100 (Figure 1).

The change in AUC-HAQ disability score, rate of change per month, and initial score were calculated for each drug for all patients and for various subsets. The HAQ scores are adjusted for (1) age, (2) baseline HAQ, (3) chronologic date of entry, (4) education level, and (5) prednisone use. The data were analyzed using SAS (version 6.12) software[38].

RESULTS

Table 1 shows the overall effectiveness measures for each of the 3 DMARD, including the average length of time patients received each DMARD. Although there was variability, the average time taking a drug was greatest for MTX (3.23 yrs), next HCQ (2.61 yrs), and lastly gold (1.96 yrs), and each of the durations is statistically different from the others ($p < 0.05$).

Of interest, the average baseline HAQ disability level was substantially higher for MTX and gold than for HCQ ($p < 0.001$), consistent with the general guidelines for treatment allocation used. AUC total disability averted was greatest with MTX, then gold, then HCQ ($p < 0.05$). When annualized, however, gold was nonsignificantly better than MTX, both of which were better than HCQ ($p < 0.01$), reflecting the influence of the more lengthy MTX treatment courses as compared with gold.

The percentage of possible disability averted is a newly coined variable that avoids the floor effects of other measures; floor effects work against the agent HCQ, given to less severely ill patients. None of the agents decreased disability by more than 25% of the theoretically possible improvement.

In analysis stratified by initial HAQ score there were significant differences among drugs. The higher the baseline HAQ score (worse disease), the greater the reduction in HAQ scores (more disability averted) for MTX (–0.138 average annual HAQ effect for entry HAQ < 1.0 vs –0.630 for patients with HAQ scores > 2.0). The same trend was present for gold, with a yearly reduction of only 0.060 units for the mildest patients and 0.842 for the most severe. Interestingly, HCQ, which is generally given to patients

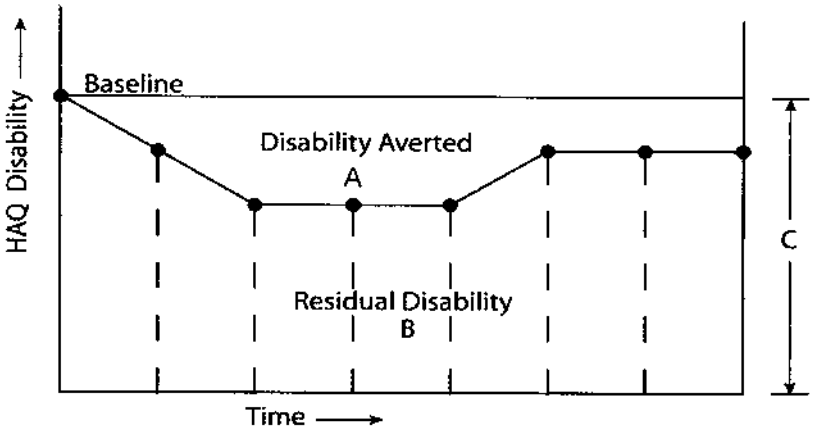

Figure 1. Area under the curve disability. The x-axis measures HAQ score and the y-axis shows time. Each closed circle represents a clinic visit when HAQ is determined. Baseline is a horizontal line constructed from the initial HAQ score. The solid line is the actual plot of HAQ scores. A: disability averted; B: residual disability; C: total potential disability.

Table 1. Responses to treatment courses of MTX, HCQ, and gold.

Drug	N	Average Baseline HAQ	SE	Average Length, yrs	SE	Average Disability Averted, AUC	SE	Annualized AUC Disability Averted	SE	% Possible Disability Averted
MTX	452	1.60	0.03	3.23	0.13	1.07	0.12	0.33	0.03	21.2
HCQ	267	1.18	0.04	2.61	0.17	0.47	0.13	0.18	0.03	16.0
Gold	205	1.52	0.05	1.96	0.17	0.74	0.16	0.38	0.04	24.1

with mild disease, works less well in patients with HAQ scores < 1.0 (–0.044/yr) than in more severely affected patients (HAQ scores 1.0–2.0, –0.423/yr), but not as well in the most severely affected patients (HAQ = 2.0, –0.179/yr). The poor response of lower levels of disability is statistically significantly different from the effect of HCQ on moderate or severely ill patients; this is likely a floor effect similar to that seen in Table 1, where HCQ performs better when based on the percentage of possible disability averted.

Only 44 patients received a second or greater course of one of the 3 DMARD studied, but the results were strong and consistent. The second course falls well short of the first (Table 2) (MTX 0.34/yr vs 0.11, $p < 0.01$; HCQ 0.19/yr vs –0.03, $p < 0.05$; gold 0.39/yr vs 0.05, $p < 0.005$). Only a small proportion of the possible disability was averted (–2 to 7.5%). Table 3 compares effectiveness of a drug if it is the first DMARD used in a patient in whom the same drug is given later in the course. Gold is dramatically less effective by all measures, if not used first, by a factor of 2 [e.g., annualized disability averted 0.40 to 0.22 ($p < 0.05$)]. HCQ averts comparable levels of disability when used first or later; however, the later HCQ patients had much higher baseline disability, and their percentage of possible disability averted was similar. MTX lost very little effectiveness when used later in the treatment sequence; the differences were nonsignificant for all variables examined.

As seen in Table 4 the effect of all 3 DMARD on disability is more pronounced in early, compared with later, disease. This was most evident with gold, which was about 3-fold more effective in reducing disability in patients within the first year of disease compared with those treated later in the course of their RA. However, the same discordance of effect in early versus late disease was noted with MTX and HCQ.

DISCUSSION

Observational databases provide a useful complement to RCT. While databases cannot match RCT for unbiased estimates of drug efficacy, they can theoretically provide better predictions of therapeutic effectiveness. The tradeoff is between the generally better internal validity of the trial and the sometimes better external validity of the observational study. Some limitations, such as regression to the mean, are similar for both types of studies. Some confounders are more typical of observational studies, such as secular time trends, whereas other confounders are more typical of RCT, such as the effects of prior therapy. To predict a treatment response with knowledge right censored at 6–24 months of

Table 2. Responses to treatment courses of MTX, HCQ, and gold (first vs second or later course).

Drug Course	N	Average Baseline HAQ	SE	Average Length, yrs	SE	Average Disability Averted, AUC	SE	Annualized Disability Averted	SE	% Possible Disability Averted
MTX	452									
First	425	1.61	0.03	3.31	0.14	1.12	0.13	0.34	0.03	21.7
Second or later	27	1.50	0.11	1.95	0.39	0.21	0.26	0.11	0.08	7.5
HCQ	267									
First	256	1.17	0.04	2.60	0.17	0.49	0.14	0.19	0.03	17.4
Second or later	11	1.50	0.23	3.00	0.71	–0.10	0.19	–0.03	0.06	–2.0
Gold	205									
First	199	1.52	0.05	1.97	0.17	0.76	0.16	0.39	0.04	24.5
Second or later	6	1.29	0.21	1.47	0.53	0.08	0.19	0.05	0.12	3.8

Table 3. Responses to treatment courses of MTX, HCQ, and gold (as first DMARD vs later DMARD).

Drug Course	N	Average Baseline HAQ	SE	Average Length, (yrs)	SE	Disability Averted, AUC	SE	Annualized Disability Averted	SE	% Possible Disability Averted
MTX	452									
First DMARD	283	1.58	0.04	3.15	0.17	1.09	0.16	0.35	0.04	22.9
Later DMARD	169	1.63	0.05	3.35	0.22	1.03	0.19	0.31	0.04	18.8
HCQ	267									
First DMARD	221	1.13	0.05	2.64	0.19	0.48	0.15	0.18	0.03	17.7
Later DMARD	46	1.46	0.10	2.49	0.35	0.41	0.31	0.17	0.08	10.6
Gold	205									
First DMARD	172	1.50	0.06	2.06	0.20	0.82	0.18	0.40	0.05	25.8
Later DMARD	33	1.60	0.11	1.44	0.24	0.32	0.14	0.22	0.06	13.0

Table 4. Responses to first treatment course of MTX, HCQ, and gold by duration of disease from initial diagnosis.

Drug	N	Average Baseline HAQ	SE	Average Length, yrs	SE	Average Disability Averted, AUC	SE	Annualized Disability Averted	SE	% Possible Disability Averted
MTX (1st course & DMARD)	283									
< 1 year from diagnosis	50	1.77	0.09	2.81	0.34	1.74	0.37	0.62	0.08	35.3
≥ 1 year from diagnosis	233	1.54	0.04	3.23	0.18	0.95	0.18	0.29	0.04	20.1
HCQ (1st course & DMARD)	221									
< 1 year from diagnosis	47	0.97	0.08	2.52	0.41	0.68	0.23	0.27	0.06	33.0
≥ 1 year from diagnosis	174	1.17	0.05	2.67	0.22	0.43	0.17	0.16	0.04	14.8
Gold (1st course & DMARD)	172									
< 1 year from diagnosis	47	1.52	0.10	2.28	0.40	1.71	0.49	0.75	0.09	46.5
≥ 1 year from diagnosis	125	1.49	0.07	1.98	0.23	0.49	0.17	0.25	0.05	16.3

treatment, in first drug courses, often in DMARD-naïve patients, and in patients who meet the criteria and accept the study is not without hazard[39]. Here we document that effectiveness of an RA treatment is a function of drug sequence, duration of disease, whether treatment is a first or second course, and depending on the severity of disease; moreover, the effects of these covariates are different for different medications and can be large. None of these clinically relevant observations have emerged from clinical trials.

We were fortunate to have longterm longitudinal data, with many data points for the HAQ Disability Index, our measure of effectiveness[40-42], and good information on covariates. A limitation is that treatment decisions were made by a single physician, but his guidelines were similar to those of many other rheumatologists, and we are finding similar effectiveness in other data sets.

We studied 3 DMARD with widespread use in North America, but these do not represent the universe of drugs of interest. Hence, we endorse a methodology for comparison of effectiveness that is clinically relevant and applicable to newer drugs as data become available.

AUC analyses are an important methodological refinement, yet they have seldom been employed. They are more stable and more sensitive to differences among patients than end-of-study or first/last scores because they are based on many observations in each patient over time. More important, they measure cumulative disability, which is a more meaningful outcome than point disability. There are many curves that connect first and last data points, including early progression, insidious progression, and late progression, and they can result in very different AUC scores. Such analyses should become standard.

A problem with outcome measures including AUC is that improvement in patients with less severe disease is systematically underestimated because there is little room to improve. We introduce here another variable, percentage of possible improvement, also an AUC concept, which is more sensitive to less severely affected patients, and which naturally complements AUC analysis using absolute values of improvement. We suggest that this analysis should also become standard, although similar percentage disability averted may not be exactly comparable from different points on the HAQ scale since HAQ scores may not be a linear metric. Figure 1 illustrates each of the area concepts. The use of the baseline value to estimate expected disability is conservative, given the expected rise in disability with increasing duration of disease.

Patients had the greatest overall benefit with MTX and continued taking MTX longer than the other DMARD. Additionally, our rules of inclusion probably underestimated the effect of MTX. We permitted an initial HAQ score after the drug was started; thus the rapid onset of action of MTX could have generated an improvement before the first HAQ score. MTX is frequently used in conjunction with other agents, so that our analysis, which did not include MTX combination DMARD therapy, systematically underestimates the actual length of time taking MTX. Interestingly, gold is as good as MTX in terms of the rate of improvement; however, patients continue gold for less time, possibly because of greater toxicity, necessitating drug withdrawal, or difficulty of monitoring. We had too few patients in the AZA, D-Pen, SSZ, and combination groups in this cohort to make detailed analyses. HCQ was about one-third or half as effective as gold or MTX.

If gold is used later in a sequence of DMARD, it is much less effective than as first DMARD. This effect has been observed before for gold[43], but interestingly not for MTX or HCQ. Our data suggest that if gold is to be used it generally should be reserved for patients with severe disease and only if they have not previously received another DMARD. Similarly, HCQ generally should be used for moderately ill patients primarily within the first year of their disease.

On occasion, patients are given second trials of a particular agent, perhaps because a particular toxicity does not preclude a retrial. These data suggest that this is an ineffective therapeutic maneuver for all 3 of these drugs. It is clear that patients do much better if treated within one year of diagnosis, with rates of improvement that range from about

2-fold to almost 4-fold that of patients treated later in the disease course. Early RA generally improves but remains active and rarely goes into drug-free remission[45]. Thus, DMARD is a lifelong therapy.

If one defines a patient responder as a patient who is better at the end than at the start of the treatment trial, then all of the DMARD evaluated give about the same chance of response. However, the magnitude of response differs among the agents. Our findings are very similar to response rates reported in comprehensive metaanalyses[1,46] and in our previous analysis[29].

These data support the observation that any DMARD used first is at a selective advantage. This may be due in part to the relative effectiveness of drugs earlier in the course of disease, but the advantage is seen even when one restricts the analysis to patients treated later in the course of their disease, suggesting that DMARD responsiveness defines a subset of patients[47]. More important, the baseline for subsequent DMARD, absent a washout period, reflects the partial disease control from a prior DMARD[43]. Our data are consistent with strategies that aim for early control of disease activity[28,48,49].

It will be interesting to use these techniques to examine data currently being acquired on the newer agents etanercept[8,12], infliximab[9-11], and leflunomide[13,14], and on combination DMARD therapy[50].

This study generates a methodologic approach to evaluate old and new drugs, combinations of drugs, and sequences of DMARD. We hope that this approach will assist in improving treatment strategies for rheumatoid arthritis.

REFERENCES

1. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses. Arthritis Rheum 1990;33:1449-61.
2. Williams HJ, Willkens RF, Samuelson CO, et al. Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis. A controlled clinical trial. Arthritis Rheum 1985;28:721-30.
3. Weinblatt ME, Coblyn JS, Fox DA, et al. Efficacy of low-dose methotrexate in rheumatoid arthritis. N Engl J Med 1985; 312:818-22.
4. Hart FD, Lewis-Faning E. Gold therapy in rheumatoid arthritis. Report of a multi-centre controlled trial. Ann Rheum Dis 1960;19:95-119.
5. Pinals RS, Kaplan SB, Lawson JG, Hepburn B. Sulfasalazine in rheumatoid arthritis. A double-blind, placebo-controlled trial. Arthritis Rheum 1986;29:1427-34.
6. Yocum DE, Klippel JH, Wilder RL, et al. Cyclosporin A in severe, treatment-refractory rheumatoid arthritis. Ann Intern Med 1998;109:863-9.
7. Weinblatt ME, Kremer JM, Bankhurst AD, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med 1999;340:253-9.
8. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. Ann Intern Med 1999;130:478-86.
9. Elliott MJ, Maini RN, Feldmann M, et al. Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. Lancet 1994;344:1125-7.
10. Elliott MJ, Maini RN, Feldmann M, et al. Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis Rheum 1993;36:1681-90.
11. Kavanaugh A, St. Clair EW, McCune WJ, Braakman T, Lipsky P. Chimeric anti-tumor necrosis factor-alpha monoclonal antibody treatment of patients with rheumatoid arthritis receiving methotrexate therapy. J Rheumatol 2000;27:841-50.
12. Smolen JS, Kalden JR, Scott DL, et al. Efficacy and safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double-blind, randomised, multicentre trial. European Leflunomide Study Group. Lancet 1999;353:259-66.
13. Weinblatt ME, Kremer JM, Coblyn JS, et al. Pharmacokinetics, safety, and efficacy of combination treatment with methotrexate and leflunomide in patients with active rheumatoid arthritis. Arthritis Rheum 1999;42:1322-8.
14. Anonymous. A randomized trial of hydroxychloroquine in early rheumatoid arthritis: The HERA study. Am J Med 1995;98:156-68.
15. Kloppenburg M, Breedveld FC, Terwiel JP, Mallee C, Dijkmans BA. Minocycline in active rheumatoid arthritis. A double-blind, placebo-controlled trial. Arthritis Rheum 1994;37:629-36.
16. Tilley BC, Alarcon GS, Heyse SP, et al. Minocycline in rheumatoid arthritis. A 48-week, double-blind, placebo-controlled trial. MIRA Trial Group. Ann Intern Med 1995;122:81-9.
17. Felson DT, LaValley MP, Baldassare AR, et al. The Prosorba column for treatment of refractory rheumatoid arthritis: a randomized, double-blind, sham-controlled trial. Arthritis Rheum 1999;42:2153-9.
18. Paulus HE, Williams HJ, Ward JR, et al. Azathioprine versus D-penicillamine in rheumatoid arthritis patients who have been treated unsuccessfully with gold. Arthritis Rheum 1984;27:721-7.
19. Anonymous. A controlled trial of cyclophosphamide in rheumatoid arthritis. N Engl J Med 1970;283:883-9.
20. Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: a 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. Arthritis Rheum 1998;41:1072-82.
21. Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. J Rheumatol 1998;25:2108-17.
22. Pincus T. Rheumatoid arthritis: disappointing long-term outcomes despite successful short-term clinical trials. J Clin Epidemiol 1988;41:1037-41.
23. Fries JF. Reevaluating the therapeutic approach to rheumatoid arthritis: the “sawtooth” strategy. J Rheumatol 1990;17 Suppl:12-5.
24. Fuchs HA, Kaye JJ, Callahan LF, Nance EP, Pincus T. Evidence of significant radiographic damage in rheumatoid arthritis within the first 2 years of disease. J Rheumatol 1989;16:585-91.
25. van der Heijde DM, van Leeuwen MA, van Riel PL, et al. Biannual radiographic assessments of hands and feet in a three-year prospective followup of patients with early rheumatoid arthritis. Arthritis Rheum 1992;35:26-34.
26. Borigini MJ, Paulus HE. Innovative treatment approaches for rheumatoid arthritis. Combination therapy. Baillieres Clin Rheumatol 1995;9:689-710.
27. O’Dell JR, Haire CE, Erikson N, et al. Treatment of rheumatoid arthritis with methotrexate alone, sulfasalazine and hydroxychloroquine, or a combination of all three medications. N Engl J Med 1996;334:1287-91.
28. Boers M, Verhoeven AC, Markusse HM, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1997;350:309-18.

29. Albert DA, Aksentijevich S, Hurst S, Fries JF, Wolfe F. Modeling therapeutic strategies in rheumatoid arthritis: Use of decision analysis and Markov models. J Rheumatol 2000;27:644-52.
30. Wolfe F, Hawley DJ, Cathey MA. Termination of slow acting antirheumatic therapy in rheumatoid arthritis: a 14-year prospective evaluation of 1017 consecutive starts. J Rheumatol 1990; 17:994-1002.
31. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
32. Wolfe F, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. Arthritis Rheum 1998;41:1571-82.
33. Wolfe F, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. Arthritis Rheum 1999;42:1797-808.
34. Wolfe F. Health status questionnaires. Rheum Dis Clin North Am 1995;21:445-64.
35. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
36. Wolfe F, Cathey MA. The assessment and prediction of functional disability in rheumatoid arthritis [published erratum appears in J Rheumatol 1991;18:1774]. J Rheumatol 1991;18:1298-306.
37. Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. Arthritis Rheum 1996;39:616-22.
38. Garrido K, Reeve K, Repole W. SAS fundamentals: A programming approach. Cary, NC: SAS Institute; 1966.
39. Wolfe F, Albert DA, Pincus T. A survey of United States rheumatologists concerning effectiveness of disease-modifying antirheumatic drugs and prednisone in the treatment of rheumatoid arthritis. Arthritis Care Res 1998;11:375-81.
40. Wolfe F. The prognosis of rheumatoid arthritis: Assessment of disease activity and disease severity in the clinic. Am J Med 1997;103:12S-18S.
41. Wolfe F. Critical issues in longitudinal and observational studies: purpose, short versus long term, selection of study instruments, methods, outcomes, and biases. J Rheumatol 1999;26:469-72.
42. Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000;43:2751-61.
43. Fries JF, Williams CA, Singh G, Ramey DR. Response to therapy in rheumatoid arthritis is influenced by immediately prior therapy. J Rheumatol 1997;24:838-44.
44. Wolfe F, Pincus T. The level of inflammation in rheumatoid arthritis is determined early and remains stable over the longterm course of the disease. J Rheumatol 2001;28:1817-24.
45. Wolfe F, Hawley DJ. Remission in rheumatoid arthritis. J Rheumatol 1985;12:245-52.
46. Felson DT, Anderson JJ, Meenan RF. Use of short-term efficacy/toxicity tradeoffs to select second-line drugs in rheumatoid arthritis. A metaanalysis of published clinical trials. Arthritis Rheum 1992;35:1117-25.
47. Anderson JJ, Wells G, Verhoeven AC, Felson DT. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum 2000;43:22-9.
48. Egsmose C, Lund B, Borg G, et al. Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. J Rheumatol 1995;22:2208-13.
49. van der Heijde D, Jacobs JWG, Bijlsma JWJ, et al. The effectiveness of early treatment with "second-line" antirheumatic drugs: A randomized controlled trial. Ann Intern Med 1996;124:699-707.
50. Pincus T, O'Dell JR, Kremer JM. Combination therapy with multiple disease-modifying antirheumatic drugs in rheumatoid arthritis: a preventive strategy. Ann Intern Med 1999;131:768-74.