Toward a Definition and Method of Assessment of Treatment Failure and Treatment Effectiveness: The Case of Leflunomide versus Methotrexate

FREDERICK WOLFE, KALEB MICHAUD, BARBARA STEPHENSON, and JOSEPH DOYLE

ABSTRACT. ***Objective.*** Time to treatment discontinuation (TTD) is an accepted method of assessing treatment effectiveness in the community, but is susceptible to channeling bias and secular and cohort effects. In addition, TTD does not consider the addition of new disease modifying antirheumatic drugs (DMARD) to insufficiently effective therapies. We expand the definition of treatment failure to include discontinuation or addition of a second DMARD (1) to examine leflunomide (LEF) versus methotrexate (MTX) effectiveness in clinical practice; (2) to obtain an estimate of overall clinical effectiveness; and (3) to identify factors associated with treatment successes and failure. In addition, (4) we test the feasibility of performing a clinical trial using a longitudinal data bank.

Methods. Using the National Data Bank for Rheumatic Diseases longitudinal data bank, 1431 patients with rheumatoid arthritis (RA) who began taking LEF or MTX as part of their routine medical care were followed from 1998 through 2001. None of the 1431 patients had received either treatment previously. Patients were assessed at 6 month intervals for periods up to 36 months by mailed questionnaires concerning DMARD therapy and demographic and RA severity factors. Kaplan-Meier survivor functions and Cox regression analyses were used to assess treatment failure, defined as time to discontinuation or to the addition of a second DMARD.

Results. For 756 patients taking LEF, the failure rate was 55.5 per 100 patient-years, and the median time to failure was 15 (95% CI 13, 17) months. For 675 patients taking MTX the failure rate was 57.3 per 100 patient-years, and the median failure time was 14 (95% CI 12, 18) months. These differences were not statistically significant. The overall rate of discontinuation was 68.7% of the failure rate. Discontinuation was predicted by adverse effects [hazard ratio 1.76 (95% CI 1.51, 2.04)] and by clinical status prior to starting DMARD, and these results were not affected by specific DMARD treatment. Discontinuation was more common with LEF, and addition of a second DMARD was more common with MTX. More than 77% of treatment failures, defined by use of additional therapy, resulted in starting anti-tumor necrosis factor treatment rather than a conventional DMARD.

Conclusion. In an observational clinical trial using a contemporary longitudinal data bank, with time to treatment failure as the outcome, LEF and MTX had equal effectiveness as measured by time to treatment failure. Treatment failure rates were substantially greater than noted historically. Given the availability of many efficacious additional treatment options, this increase in failure rate appears to reflect a greater propensity to discontinue and/or add therapy. (J Rheumatol 2003;30:1725–32)

Key Indexing Terms:
LEFLUNOMIDE METHOTREXATE EFFECTIVENESS DISCONTINUATION
TERMINATION PREDICTION RHEUMATOID ARTHRITIS

The length of time patients remain on treatment has been shown to be an excellent measure of drug effectiveness or ineffectiveness[1-6]. Time on drug provided a validation or refutation of the results of clinical trials, as patients and physicians voted with their feet as to drug effectiveness. There are many instances when controlled trials demonstrated or failed to demonstrate efficacy where actual effectiveness measured in the community by time on drug yielded contrary results[7-12]. This is particularly true for methotrexate (MTX), which was not better than other disease modifying antirheumatic drugs (DMARD) in meta-analyses[11], but became the central treatment of rheumatoid arthritis (RA) and the one that RA patients stayed with the longest[1-3].

In another sense, time on drug can be thought of as a measure of treatment failure, in that patients stop therapies that are insufficiently effective or have unacceptable side effects. However, duration of therapy and its discontinuation can be influenced by a number of factors. All other things

From the National Data Bank for Rheumatic Diseases and University of Kansas School of Medicine, Wichita, Kansas, USA.

Supported by a grant from Aventis Pharmaceuticals, Inc.

F. Wolfe, MD, National Data Bank for Rheumatic Diseases and University of Kansas School of Medicine; K. Michaud, MS, National Data Bank for Rheumatic Diseases; B. Stephenson, RN, PhD; J. Doyle, RPh, MBA, Aventis Pharmaceuticals, Inc., Bridgewater, New Jersey, USA.

Address correspondence to Dr. F. Wolfe, Arthritis Research Center, 1035 N. Emporia, Suite 230, Wichita, KS 67214.
E-mail: fwolfe@arthritis-research.org

Submitted July 30, 2002; revision accepted January 16, 2003.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

being equal, the key factor is the balance between efficacy and tolerability. Tolerable drugs may be continued even if they have limited efficacy, and drugs that cause adverse events or side effects may be tolerated if efficacy is sufficient. The optimum drug, of course, has much efficacy and few side effects.

Despite its attractiveness and simplicity, duration of therapy has many problems, for other factors may also play a role in duration of therapy. For one, confounding by indication and channeling bias can lead to erroneous results. These biases, which are most prominent just after drug release, describe a condition in which patients with the worst prognosis are assigned the most effective drugs, with the result that the drugs do not appear to be as effective as they seemed in clinical trials[13-15]. These biases usually disappear after a few years as the pool of severe and nonresponder patients complete their exposure to the new agent. Another crucially important factor is the availability of other therapies. For example, studies showing that patients stayed on MTX longest were performed at a time when there were no treatments perceived as being more effective, in contrast to the current era, when anti-tumor necrosis factor (TNF) agents are available. It seems likely, then, that time on MTX would be reduced at this time compared to a decade ago, even though its actual effectiveness has not changed. This switching behavior is modified further by drug cost, formulary status, availability, and current physician and patient beliefs about treatment efficacy.

Perhaps the most important factor that limits the usefulness of time to discontinuation, and one not previously considered, is the current tendency to add additional DMARD therapy in the face of insufficient efficacy rather than simply discontinuing treatment.

The consequence of all of these influencing factors is that studies cannot compare drugs used in one time period with drugs used in another, cannot simply consider treatment discontinuation as the endpoint, and must be appropriately matched or adjusted for demographic and patient severity characteristics.

In this report we use this new definition of treatment failure, the discontinuation or the use of additional DMARD, (1) to examine leflunomide (LEF) versus MTX effectiveness in clinical practice; (2) to obtain an estimate of overall clinical effectiveness; and (3) to identify factors associated with treatment successes and failure.

We used the National Data Bank for Rheumatic Diseases (NDB), a large rheumatic disease data bank, to investigate survival time on LEF and MTX for patients who had received these drugs as part of their routine medical (not by random allocation). To insure comparability of treatment groups, we restricted the study population to patients who had never been on either MTX or LEF previously, and we also restricted the analysis period to the same years, 1998 or later, the time LEF first became available. To ascertain the effect of illness severity on survival time, we collected key covariate variables prior to the start of either of the 2 therapies. Treatment failure was defined as discontinuation of therapy or the addition of a second DMARD.

MATERIALS AND METHODS

Persons in this study are patients with RA who participated in NDB survey research. The NDB began patient assessment in January 1999 for the 6-month period ending December 1998. Assessments are conducted at 6 month intervals every January and July. The NDB is now in it 8th biannual assessment, and 14,091 patients with RA had completed at least one detailed questionnaire assessment by December 31, 2001, corresponding to the study closing date of June 30, 2001. NDB methodology has been described previously[14,16].

NDB participants are asked to complete semiannual, detailed, 28 page questionnaires about all aspects of their illness. At each questionnaire assessment, demographic variables were recorded including sex, age, ethnic origin, education level, and current marital status. Study variables included the Stanford Health Assessment Questionnaire functional disability index (HAQ disability)[17,18], visual analog scale (VAS) for pain, global disease severity and fatigue scales[19], the Arthritis Impact Measurement Scales (AIMS)[20,21] anxiety and depression scales[22,23], and Likert scales that assessed current satisfaction with health. Patients also report start and stop date of all medication, as well as side effects (adverse events) that they associate with the various RA treatments. Coders translate patient reported adverse events to standard codes using an NDB coding manual. Adverse events that resulted in hospitalization or related to liver or ulcer disease were validated by obtaining medical records. All study participants completed at least one detailed questionnaire.

At the time of enrollment, prior to the first detailed research questionnaire, we obtained lifetime drug histories from each patient and obtained HAQ, pain, global severity, sleep, fatigue, and satisfaction values.

There were 3 major groupings of patients with RA who were enrolled from the practices of 834 US rheumatologists. Patients in the first group (n = 9603) were encouraged to enroll in the NDB by their rheumatologists because of their physicians' general support of NDB research. Physicians aiding in this enrollment received no compensation from any source. A second group of 3182 patients agreed to participate in this research at the time they were started on LEF by their rheumatologist. The rheumatologists aiding in the enrollment process received compensation of $50 per patient from Aventis Pharmaceuticals, Inc. Enrollment of this group was completed by the end of 1999. A third patient group of 1397 patients were enrolled as part of an ongoing safety assessment registry for infliximab. Physicians aiding in obtaining safety outcome data on these patients also received compensation for their work in obtaining disease activity and followup safety data. The mean age and disease duration of the 14,091 patients in this data set at the time of last observation were: age 60.7 (SD 13.4) years, disease duration 14.0 (SD 10.7) years.

To evaluate the treatment effect on patients new to both MTX and LEF, we started with a sample size of 8504 and 4551 patients who had ever taken MTX or LEF, respectively. Patients who had received either of these 2 drugs prior or simultaneously to receiving the alternative drug (MTX = 2833, LEF = 2559) or who had started the DMARD prior to observation in the data bank (MTX = 4697, LEF = 778) were excluded. In addition, 299 patients in the MTX group and 458 in the LEF group had incomplete preassessment disease activity data or prior DMARD use data and were also excluded. Thus the final study sample consisted of 675 patients first starting MTX and 756 patents first starting LEF, or 11.0% of all patients who received MTX or LEF in the data set. In addition, no patient in the study sample had received infliximab or etanercept prior to starting MTX or LEF. At enrollment into the data bank the mean (standard deviation) and 5th, 25th, and 50th percentiles of disease duration for study participants were 11.2 (10.6), 0.3, 2.4, and 8.7 years, respectively. For the remaining data bank patients the values were 11.8 (10.8), 0.50, 3.3, 8.8 years. These

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

values are expected for community practices in which patients with RA are cared for over their lifetimes.

Patients received no payments for study participation or for drug costs. Treatment decisions were made by patients and physicians, and there were no inducements for patients to continue or discontinue therapy.

Statistical methods. Time on drug is usually assessed by Kaplan-Meier survivor function methods, including Cox regression. The endpoint of such analyses is drug discontinuation. Kaplan-Meier methods properly account for patients who have not yet discontinued therapy and are being assessed at differing followup times. Because Kaplan-Meier methods were developed in the setting of mortality assessment, it is common to speak of "survival time" for time on drug even though mortality is not the issue. Other phrases that describe duration of therapy include "time to discontinuation," "failure," "treatment failure," and "failure time."

Treatment failure was defined as treatment discontinuation or the use of an additional DMARD (not prednisone) at a time after the start of LEF or MTX. Adverse events or side effects were those events occurring at or prior to the time of drug discontinuation or the addition of other DMARD therapy. Adverse events were self-reported by patients as being caused by either LEF or MTX.

The primary analytic methods used in this report were Kaplan-Meier estimates and Cox regression. Data were analyzed using Stata 7.0[24]. To estimate longterm survival, the extended mean survival was also calculated. Extended mean computes the mean survival by exponentially extending the survival curve to zero[25]. T tests and chi-square tests (Table 1) and multinomial logistic regression were used as described. Statistical significance was set at the 0.05 level.

RESULTS

Demographic and severity variables. Table 1 describes the demographic, treatment, and clinical severity variables for the study participants. Although patients did not differ in age, sex, or other demographic characteristics, patients beginning LEF had slightly more severe RA compared to those taking MTX. Disease duration was greater among LEF patients (13.3 vs 12.2 yrs), as was prednisone use (68.8% vs 54.2%) and lifetime DMARD use (76.2% vs 72.9%). In addition, LEF patients had slightly more abnormal values for HAQ (1.30 vs 1.22) and fatigue (5.2 vs. 4.8), but treatment groups did not differ in pain, patient global, sleep disturbance, or satisfaction with health.

Rate of treatment failure among RA patients receiving LEF or MTX. The survival curves and median time to treatment failure were similar for patients beginning MTX and for those starting LEF (log-rank test, chi-square = 0.30, p = 0.584). As shown in Table 2 and Figure 1a, the median time to treatment failure was 14 (95% CI 12, 18) months for MTX and 15 (95% CI 13, 17) months for LEF. The treatment failure rate was 57.3 and 55.5 per 100 patient-years for MTX and LEF, respectively. The survival curves were not significantly different after adjusting for the following major covariates of Table 1: age, sex, disease duration, total income, prednisone use, previous DMARD, and pretreatment values of HAQ and fatigue (log-rank test, chi-square = 0, p = 1.000).

We explored the 2 components of treatment failure, discontinuation and addition of other DMARD therapy. As shown in Figure 1b and Table 2, patients receiving LEF had higher rates of discontinuation than patients taking MTX, 42.4 versus 33.2 per 100 patient-years (chi-square = 13.13, p = 0.000). The median time to discontinuation was 22 (17, 28) months for LEF, but the 50th percentile had not been reached by MTX patients by study closure. An extended mean estimate of survival times was 52.8 months for MTX compared with 37.3 months for LEF.

Table 1. Demographic and clinical data at treatment start from 1431 Patients with RA treated with MTX or LEF.

Variable	MTX, n = 675 Mean (SD) or %	Leflunomide, n = 756 Mean (SD) or %	t (chi-square)	p
Age, yrs	59.87 (14.14)	58.96 (12.31)	1.293	0.196
Sex, % male	20.59	21.43	(0.1500)	0.698
White, %	90.22	91.53	(0.7434)	0.389
High school graduates, %	87.70	89.95	(1.817)	0.178
College graduates, %	24.33	26.23	(0.6745)	0.411
Total family income, US$	41,896 (27,714)	43,730 (28,284)	–1.236	0.217
Comorbidity, 0–11	2.20 (1.84)	2.18 (1.72)	0.1678	0.867
Disease duration, yrs	12.21 (10.82)	13.28 (10.38)	–1.908	0.057
Prior prednisone, %	54.22	68.78	(32.06)	0.000
No. of previous DMARDs, %			(29.89)	0.000
0	27.11	23.81		
1	49.48	43.92		
2	18.37	21.03		
3	4.89	7.94		
4	0.15	2.78		
5	0.00	0.53		
HAQ disability, 0–3	1.22 (0.71)	1.30 (0.68)	–2.328	0.020
Pain, 0–10	4.53 (2.79)	4.61 (2.66)	–0.0523	0.601
Global severity, 0–1	3.99 (2.61)	3.90 (2.35)	0.681	0.496
Fatigue, 0–10	4.79 (3.03)	5.18 (2.80)	–2.544	0.011
Sleep disturbance, 0–10	4.06 (3.20)	4.06 (3.13)	0.004	0.997
Satisfaction with health, –2 to 2	0.09 (1.27)	0.12 (1.19)	–0.569	0.570

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

Table 2. Survival times and failure rates for patients starting MTX or LEF.

Therapy Group	Patients	Time at Risk (patient-yrs)	Failure Rate per 100 Patient-yrs	25% (95% CI), mo	50% (95% CI), mo	75% (95% CI), mo
Treatment failure: discontinuation or addition of DMARD						
MTX	675	486.7	57.3	6 (4, 6)	14 (12, 18)	—
LEF	756	740.3	55.5	5 (4, 5)	15 (13, 17)	—
Treatment discontinuation: discontinuation of DMARD among all patients						
MTX	675	559.8	33.2	9 (6, 12)	—	—
LEF	756	809.8	42.4	6 (5, 6)	22 (17, 28)	—
Treatment discontinuation subanalysis: discontinuations of DMARD only						
MTX	165	80.4	205	2 (1, 2)	4 (3, 6)	7 (6, 11)
LEF	317	182.8	173	2 (2, 3)	5 (4, 6)	10 (8, 11)
Treatment addition subanalysis: addition of DMARD only						
MTX	114	62	184	2 (1, 2)	5 (3, 6)	10 (7, 12)
LEF	94	82	115	4 (2, 5)	10 (6, 12)	16 (13, 19)

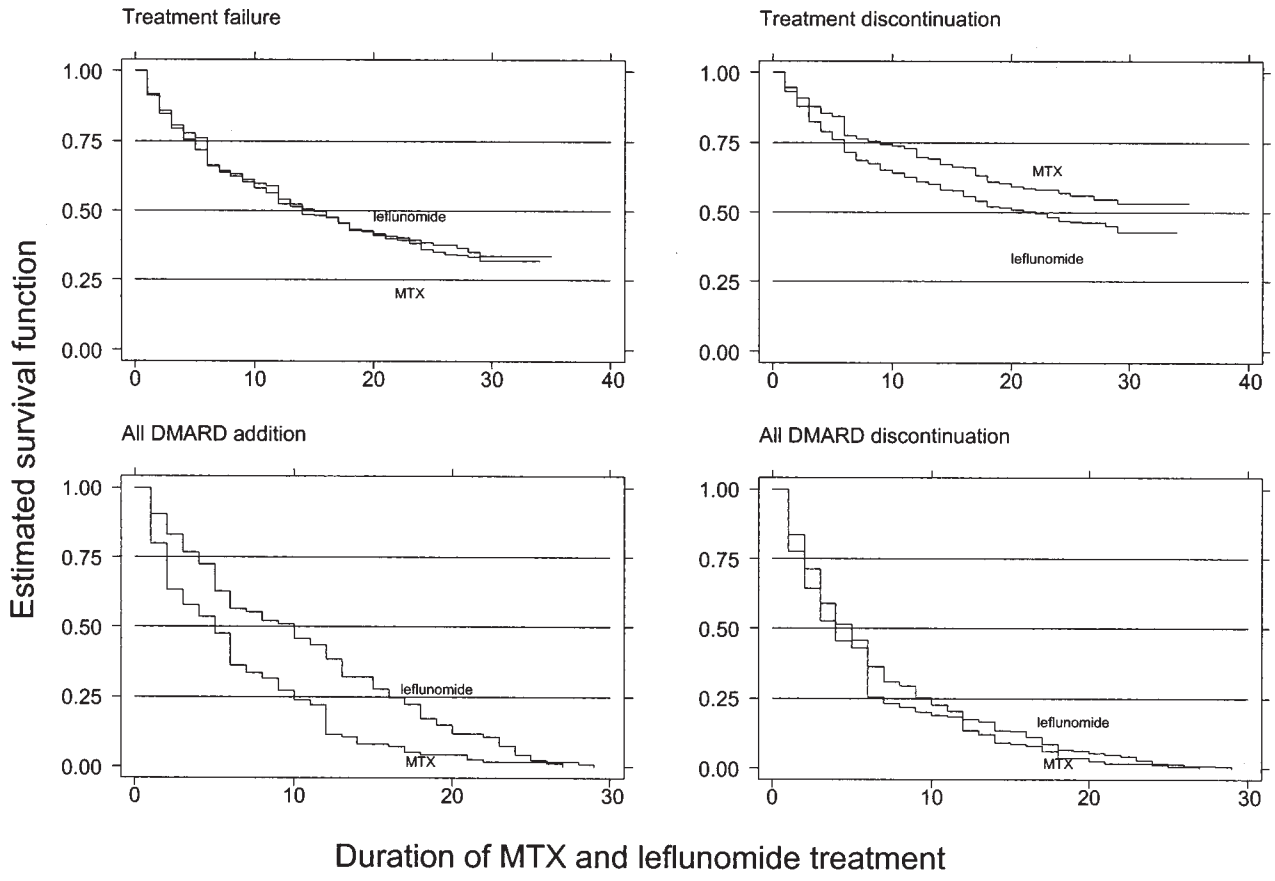

Figure 1. A. Kaplan-Meier survival estimates for 675 and 756 patients treated with MTX and LEF, respectively. Failure is treatment discontinuation or use of additional DMARD therapy. B. Kaplan-Meier survival estimates for 675 and 756 patients treated with MTX and LEF, respectively. Failure is treatment discontinuation. C. Kaplan-Meier survival estimates for 114 and 94 patients treated with MTX and LEF, respectively. Only patients adding additional therapy are analyzed. Failure is the use of additional DMARD therapy. D. Kaplan-Meier survival estimates for 165 and 317 patients treated with MTX and LEF, respectively. Only patients discontinuing therapy are analyzed. Failure is treatment discontinuation.

Tables 2 and 3 and Figures 1c and 1d provide additional insight into the discontinuation/addition process. As shown in Table 3, for patients who were treatment failures, more patients failed MTX than LEF by adding additional DMARD (40.9% vs 32.9%), but more discontinued LEF than MTX (77.1% vs 59.1%). Additions were not only more frequent among MTX patients, but they also occurred much earlier (Figure 1c) (log-rank test, chi-square = 13.83, p =

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

Table 3. Classification of MTX and LEF treatment failures.

DMARD	Treatment Failure by DMARD Addition, %	Treatment Failure by Discontinuation, %	All Treatment Failures, % (n)
MTX	40.9	59.1	100 (279)
LEF	32.9	77.1	100 (411)

0.000). Conversely, treatment differences related to discontinuation (Figure 1d) were not statistically significant (log-rank test, chi-square = 3.25, p = 0.072).

Influence of adverse events on the treatment failure rate. Persons with self-reported adverse events to specific treatments had higher rates of treatment failure (adverse events reported as occurring after treatment failure are not included in these analyses as they are unrelated to risk of treatment failure). The univariate hazard ratio for adverse events was 1.76 (95% CI 1.51, 2.04). To clarify the interaction of adverse events with specific treatments, we compared treatment and adverse event categories in Figure 2. Although the treatment failure rates for patients with adverse events and those without adverse events differed significantly (chi-square = 38.93, p = 0.000), there was no significant difference between therapy groups with adverse events (chi-square = 0.08, p = 0.781) and groups without adverse events (chi-square = 1.18, p = 0.277). Adverse events are defined as self-reported side effects to specific treatments. These data indicate that although adverse events are important determinants of treatment failure, LEF and MTX do not differ in regard to failure as a function of adverse events.

Why is the discontinuation/addition pattern for treatment failure different among LEF and MTX treated patients? As noted above, patients taking LEF were more likely to be treatment failures because of discontinuation, while for MTX patients it is because of the use of additional DMARD therapy. Table 4 shows that the major additions in therapy were with anti-TNF agents, and that MTX patients were more likely to add infliximab, but LEF patients were slightly more likely to add etanercept. Multinomial logistic regression using infliximab, etanercept, and other treatments as dependent variables and age, sex, disease duration, education, income, and ethnic origin as explanatory variables

Table 4. DMARD therapies added by patients who were treatment failures by the addition of another DMARD.

DMARD Therapy	MTX (n = 114), %	Leflunomide (n = 94), %
Etanercept	21.1	44.7
Infliximab	56.1	38.3
Hydroxychloroquine	9.6	7.4
Sulfasalazine	8.8	6.4
Minocycline	2.6	4.3
Azathioprine	1.8	3.2
MTX	—	0.0
LEF	0.9	—

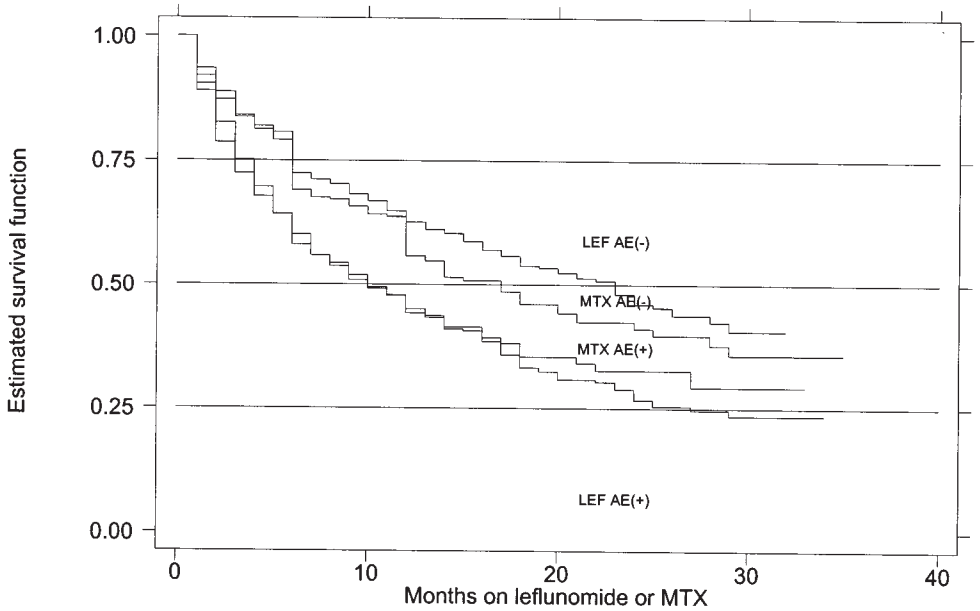

Figure 2. Kaplan-Meier survival estimates for LEF patients without adverse events [LEF AE(–); n = 394], MTX patients without adverse events [MTX AE(–); n = 486], MTX patients with adverse events [MTX AE(+); n = 189], and LEF patients with adverse events [LEF AE(+); n = 362]. LEF and MTX groups without adverse events are not significantly different (chi-square = 1.18, p = 0.277). LEF and MTX groups with adverse events are not significantly different (chi-square = 0.08, p = 0.781), but groups with adverse events and those without adverse events do differ (log-rank test, chi-square = 38.93, p = 0.000). Adverse events are defined as self-reported side effects to specific treatments. Adverse events reported as occurring after treatment failure are not included.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

suggested a possible role of total family income per US$10,000 [odds ratio 1.11 (95% CI 0.989, 1.249)]. In univariate analysis the income odds ratio for additional therapy was 1.12 (95% CI 1.005, 1.246).

Predicting treatment failure. To study treatment failure, we used demographic, treatment, and severity variables obtained just prior to treatment start, and adverse event data occurring prior to treatment failure. In univariate analyses, adverse event was the strongest predictor of treatment failure (Table 5). Health satisfaction was the highest ranked clinical variable as a predictor of drug status, followed by global severity, fatigue, and pain.

In a multivariate model of treatment failure (Table 6), only adverse events, comorbidity, pain, satisfaction with health, and global severity were significant at a level of 0.1 or less. Addition of treatment group had no appreciable effect on regression hazard ratios, nor was it significant when added to the model in Table 6 [hazard ratio 0.96 (95% CI 0.820, 1.119)].

DISCUSSION

There are a number of results of interest. The broader definition of treatment failure (treatment discontinuation or subsequent addition of a second DMARD) shows that time to discontinuation overestimates DMARD efficacy. Compared to median treatment failure times of 15 and 14 months for LEF and MTX, respectively, failure as time to discontinuation was more than 22 months for LEF and in excess of that for MTX. Treatment failure rates per 100 patient-years were 55 for LEF and 57 for MTX, and 42 and 33 for treatment discontinuation.

These data show that treatment failure and its converse, treatment success, are similar for LEF and MTX. This relatively large clinical sample (n = 756 for LEF, 675 for MTX) provides the first evidence of equivalency in clinical practice, in support of clinical trial results[26-29]. In addition, in clinical practice there are many reasons to discontinue therapy or to add additional DMARD. As such, effectiveness in the "real world" setting of the clinic is a rigorous clinical test.

While the results showed that LEF was equivalent to the mainstay of RA treatment, MTX, they also provided evidence of the limitations of current therapy, for half the patients started on these therapies were treatment failures in 14–15 months. In 1995, Wolfe analyzed length of time taking DMARD in 16 studies involving 5809 patients[2]. The

Table 5. Univariate Cox regression analyses of predictors of MTX anf LEF treatment failure among 1431 RA patients. HAQ, pain, global, fatigue, sleep, and health satisfaction data are for assessments prior to LEF treatment start. Side effects refer to patient reported side effects specific to LEF or MTX not occurring after treatment failure.

Variable	Hazard Ratio	SE	Z	p	95% CI
Treatment side effects, yes/no	1.76	0.13	7.37	0.000	1.51, 2.04
Global severity, 0–10	1.10	0.02	5.93	0.000	1.06, 1.13
Health satisfaction, –2 to 2	1.20	0.04	5.65	0.000	1.12, 1.27
Pain, 0–10	1.08	0.02	5.60	0.000	1.05, 1.11
Fatigue, 0–10	1.06	0.01	4.56	0.000	1.04, 1.09
Health satisfaction, no/yes	1.43	0.11	4.52	0.000	1.23, 1.68
HAQ, 0–3	1.26	0.07	4.20	0.000	1.13, 1.41
Sleep disturbance, 0–10	1.05	0.01	4.05	0.000	1.03, 1.07
No. of comorbid conditions	1.08	0.02	3.99	0.000	1.04, 1.13
High school graduate, yes/no	0.87	0.10	–1.29	0.197	0.68, 1.08
Age, years	1.00	0.00	–1.19	0.233	0.99, 1.00
Prior prednisone, yes/no	1.10	0.09	1.17	0.243	0.94, 1.28
Total income, US$	1.00	0.00	–0.84	0.398	1.00, 1.00
Male sex, yes/no	0.95	0.09	–0.54	0.589	0.79, 1.14
Disease duration, years	1.00	0.00	–0.29	0.770	0.99, 1.01
No. of prior DMARD	1.00	0.04	–0.05	0.961	0.92, 1.08

Table 6. Multivariate model of therapy discontinuation in 1431 RA patients receiving MTX or LEF.

Variable	Hazard Ratio	SE	Z	p	95% CI
Had side effect before failure, yes/no	1.56	0.12	5.87	0.000	1.35, 1.82
Number of comorbid conditions, 0–11	1.06	0.02	2.66	0.008	1.01, 1.10
Pain, 0–1	1.23	0.11	2.30	0.021	1.03, 1.47
Satisfaction with health, -2 to 2	1.09	0.04	2.25	0.025	1.01, 1.17
Global severity, 0–1	1.22	0.11	2.17	0.030	1.02, 1.47

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

weighted average (median) survival time for MTX was 4.62 years. For other drugs, survival time was auranofin 1.2 years, intramuscular gold 1.4 years, sulfasalazine 1.1 years, penicillamine 1.5 years, azathioprine 2.3 years, and antimalarials 1.6 years. It would seem that the efficacy of MTX shown by previous studies is actually considerably less when alternative treatment options are available. This should be no surprise, for although both MTX and LEF are efficacious agents, most patients continue to have unacceptable levels of RA problems and would be expected to want to utilize newer therapies.

The most important predictor of treatment failure was adverse events, as might be expected (Figure 2). In addition, baseline measure of severity, particularly pain, global severity, and satisfaction with health, and comorbidity added predictive ability. Of interest, prednisone use and the number of previous DMARD were of less use to predict treatment outcome. These observations underscore the need to be quantitative in measuring the severity in clinical care, as they provide help in judging potential therapy response.

More patients added another DMARD to MTX than to LEF. We found that income might play some role in the decision, but it seems more likely that patterns of acceptable use played the stronger role. For example, MTX and infliximab are a combination approved by the US Food and Drug Administration, but LEF and infliximab are not.

The data from this study suggest the methods we used here may be helpful to judge actual effectiveness in clinical practice. Because we analyzed patients from many rheumatologists, the results are a reasonable reflection of contemporary practice; and using standardized questionnaires, the same level of information is obtained from each patient. In addition, by using patients who were unexposed to either drug previously, we were able to obtain appropriate groups for comparison. As noted in Table 1, the 2 treatment groups were generally similar. Given a similar exposure period and entry criteria, this methodology allows one to compare a new treatment with one of known efficacy. Thus, in many respects this observational study resembles a controlled trial. Although there is no randomization, there is also no exclusion. Patients have a greater representativeness than may be found in clinical trials, and they have few inducements to inflate response or remain on treatment. Among the potential limitations of this study are the exclusive use of self-report data and the absence of physician examination and laboratory data to predict discontinuation (Tables 1, 5, and 6). It is possible that data from these additional sources might have provided additional information about univariate and multivariate predictors.

We believe that the use of treatment failure rather than treatment discontinuation is an important advance in assessing treatment value. Besides being helpful in comparing drugs that may have different pathways for failure, the method also provides an insight into overall treatment effectiveness.

The definition of treatment failure used here is a conservative one. We did not distinguish between treatment failure and insufficient response because such differences are merely semantic, and many patients with insufficient response continue to take their DMARD. Elsewhere, based on patient evaluations, we have presented preliminary data showing that a HAQ score ≤ 0.5 represents successful treatment[30]. Treatments that result in HAQ scores above that level may be considered treatment failures, as all such responses are insufficient regardless of whether patients discontinue therapy or not.

The major limitation of this new methodology is the requirement for patients to be unexposed to either treatment. Because MTX is the current treatment standard, most patients will have been exposed to it (> 70% in the NDB); and when a new treatment such as LEF becomes available, most patients receiving the new treatment will have had exposure to MTX. Therefore, only RA databases that include large numbers of patients (e.g., 10,000) over short periods of time (3–4 years) are able to produce such studies.

We performed a nonrandomized "clinical trial" using a contemporary longitudinal data bank. Using time to treatment failure as the outcome, the results were similar for leflunomide and methotrexate. The effect of adverse events was equal on both treatment arms, and adverse events were the most important predictor of outcome. But nonpredicted failure was the most important effect. Treatment failure rather than treatment discontinuation more accurately describes treatment effect. Finally, it is possible to conduct a nonrandomized clinical trial in the setting of an observational study, with results that have applicability to community practice.

REFERENCES

1. 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.
2. Wolfe F. The epidemiology of drug treatment failure in rheumatoid arthritis. Baillieres Clin Rheumatol 1995;9:619-32.
3. Pincus T, Marcum SB, Callahan LF. Longterm drug therapy for rheumatoid arthritis in seven rheumatology private practices. 2. Second line drugs and prednisone. J Rheumatol 1992;19:1885-94.
4. Maetzel A, Wong A, Strand V, Tugwell P, Wells G, Bombardier C. Meta-analysis of treatment termination rates among rheumatoid arthritis patients receiving disease-modifying anti-rheumatic drugs. Rheumatology 2000;39:975-81.
5. Morand EF, McCloud PI, Littlejohn GO. Life table analysis of 879 treatment episodes with slow acting antirheumatic drugs in community rheumatology practice [published erratum appears in J Rheumatol 1992;19:1998]. J Rheumatol 1992;19:704-8.
6. Wijnands MJ, van't Hof MA, van Leeuwen MA, van Rijswijk MH, van de Putte LBA, van Riel PL. Long-term second-line treatment: a prospective drug survival study. Br J Rheumatol 1992;31:253-8.
7. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Read JL. Auranofin therapy and quality of life in patients with rheumatoid

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org

arthritis. Results of a multicenter trial. Am J Med 1986;81:565-78.
8. Hamdy H, McKendry RJ, Mierins E, Liver JA. Low-dose methotrexate compared with azathioprine in the treatment of rheumatoid arthritis. A twenty-four-week controlled clinical trial. Arthritis Rheum 1987;30:361-8.
9. Hochberg MC. Auranofin or D-penicillamine in the treatment of rheumatoid arthritis. Ann Intern Med 1986;105:528-35.
10. Rau R, Herborn G, Menninger H, Blechschmidt J. Comparison of intramuscular methotrexate and gold sodium thiomalate in the treatment of early erosive rheumatoid arthritis: 12 month data of a double-blind parallel study of 174 patients. Br J Rheumatol 1997;36:345-52.
11. 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.
12. Tugwell P, Pincus T, Yocum D, et al. Combination therapy with cyclosporine and methotrexate in severe rheumatoid arthritis. N Engl J Med 1995;333:137-41.
13. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.
14. Wolfe F, Flowers N, Burke T, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.
15. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. BMJ 1997;315:1151-4.
16. Wolfe F, Kong SX, Watson DJ. Gastrointestinal symptoms and health related quality of life in patients with arthritis. J Rheumatol 2000;27:1373-8.
17. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
18. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
19. Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. J Rheumatol 1996;23:1407-17.
20. Meenan RF, Gertman PM, Mason JH, Dunaif R. The Arthritis Impact Measurement Scales. Arthritis Rheum 1982;25:1048-53.
21. Meenan RF. The AIMS approach to health status measurement: conceptual background and measurement properties. J Rheumatol 1982;9:785-8.
22. Hawley DJ, Wolfe F. Depression is not more common in rheumatoid arthritis: a 10 year longitudinal study of 6,608 rheumatic disease patients. J Rheumatol 1993;20:2025-31.
23. Hawley DJ, Wolfe F. Anxiety and depression in patients with RA: A prospective study of 400 patients. J Rheumatol 1988;15:932-41.
24. Stata Corporation. Stata Statistical Software: Release 7.0. College Station, TX: Stata Corp.; 2001.
25. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer; 1997:91-2.
26. Cohen S, Cannon GW, Schiff M, et al. Two-year, blinded, randomized, controlled trial of treatment of active rheumatoid arthritis with leflunomide compared with methotrexate. Utilization of Leflunomide in the Treatment of Rheumatoid Arthritis Trial Investigator Group. Arthritis Rheum 2001;44:1984-92.
27. Emery P, Breedveld FC, Lemmel EM, et al. A comparison of the efficacy and safety of leflunomide and methotrexate for the treatment of rheumatoid arthritis. Rheumatology 2000;39:655-65.
28. Schiff MH. Leflunomide versus methotrexate: A comparison of the European and American experience. Scand J Rheumatol 1999;28:31-5.
29. Strand V, Cohen S, Schiff M, et al. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Leflunomide Rheumatoid Arthritis Investigators Group. Arch Intern Med 1999;159:2542-50.
30. Wolfe F, Michaud K. Toward an acceptable definition of DMARD failure and abnormal health status in RA [abstract]. Arthritis Rheum 2002;46 Suppl:S865.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on October 25, 2021 from www.jrheum.org