

Predicting Mortality in Patients With Rheumatoid Arthritis

Frederick Wolfe,¹ Kaleb Michaud,² Olaf Gefeller,³ and Hyon K. Choi⁴

Objective. A number of different variables have been proposed as risk factors for mortality in patients with rheumatoid arthritis (RA), but limited prospective information on the magnitude of their effects is available. This study was undertaken to evaluate the relative predictive strength and usefulness of a wide range of variables on the risk of mortality in a large, long-term, prospectively studied cohort of patients with RA.

Methods. Over a 20-year period of followup beginning in 1981, 1,387 consecutive RA patients were seen in a single clinic. A wide range of clinical and demographic assessments were recorded and entered into a computer database at the time of each clinical assessment. Assessment of predictive strength included determination of standardized and fourth-versus-first-quartile odds ratios (ORs), goodness-of-fit measures, and contributing fraction.

Results. The Health Assessment Questionnaire (HAQ) disability index was the strongest clinical predictor of mortality. A 1-SD change in the HAQ resulted in a much larger increase in the odds ratio for mortality compared with a 1-SD change in global disease severity, the next most powerful predictor of mortality (OR 2.31 versus 1.83). Considering the contributing fraction, mortality would be reduced by 50% for the HAQ and by 33% for global disease severity if patients in the fourth quartile for these variables could be switched to the first quartile. Global disease severity, pain, depression, anxiety, and laboratory and radiographic features were significantly weaker predictors. Disease duration, nodules, and tender joint count were clinical variables that

provided very little predictive information. In multivariable analyses, HAQ and other patient self-report measures were significantly better predictors than were radiographic and laboratory variables. A single baseline observation provided the least information, with substantially increasing predictive ability associated with 1-year, 2-year, and all-time point followup observations (time-varying covariates).

Conclusion. In this large 20-year study from routine clinical practice, the HAQ was the most powerful predictor of mortality, followed by other patient self-report variables. Laboratory, radiographic, and physical examination data were substantially weaker in predicting mortality. We recommend that clinicians collect patient self-report data, since they produce more useful clinical outcome information than other available clinical measures.

Rheumatoid arthritis (RA) is a chronic progressive disease associated with systemic inflammation. The disease directly affects physical function and mobility and causes substantial short-term and long-term morbidity. Furthermore, in most studies, patients with RA have a significantly shorter life expectancy compared with the general population (1–13).

While many clinical trials and short-term observational studies evaluate predictors of RA morbidity outcomes, there has been only limited information available on predictors of mortality (a long-term outcome) (12,14–18). Accurate understanding of mortality predictors is important from a public health standpoint as well as for clinical management of RA. Furthermore, predictive information is vital in assessing the potential impact of changes observed in short-term studies, including many recent clinical trials, on the life expectancy of RA patients. Similarly, the same knowledge will greatly improve the validity of long-term cost-effectiveness analyses.

Modern interest in the predictors of mortality in RA was stimulated by studies by Pincus and colleagues (19,20). They reported that baseline functional status

¹Frederick Wolfe, MD: Arthritis Research Center Foundation and University of Kansas School of Medicine, Wichita, Kansas; ²Kaleb Michaud, MS: Arthritis Research Center Foundation, Wichita, Kansas; ³Olaf Gefeller, PhD: University of Erlangen-Nuremberg, Erlangen, Germany; ⁴Hyon K. Choi, MD, MPH: Massachusetts General Hospital, Boston.

Address correspondence and reprint requests to Frederick Wolfe, MD, National Data Bank for Rheumatic Diseases, Arthritis Research Center Foundation, 1035 North Emporia, Suite 230, Wichita, KS 67214. E-mail: fwolfe@arthritis-research.org.

Submitted for publication August 28, 2002; accepted in revised form February 19, 2003.

and severe disease activity were strong risk factors for mortality, based on a cohort of 75 RA patients followed up over a 15-year period. In the current large prospective study, we comprehensively evaluated the relative predictive strength and usefulness of a wide range of variables, including functional status, on the risk of mortality in a cohort of patients with RA studied over a long period. The first objective of the study was to quantify in various ways the predictive strength of clinical and demographic variables that are commonly used in clinical trials and easily measurable in clinical practice. The second objective was to evaluate the relative predictive strength of variables as a function of observation time. We describe relative predictive strength of variables in 4 ways: 1) at baseline, 2) as a mean over the first year of observation, 3) as a mean over the first 2 years, and 4) as time-varying covariates at all followup time points.

PATIENTS AND METHODS

Patients. Beginning in 1974 and continuing through 1999, 1,922 consecutive RA patients seen at the Wichita (Kansas) Arthritis Center, an outpatient rheumatology clinic, were enrolled into a computerized databank at the time of their first clinic visit. Demographic, clinical, laboratory, and self-report data were obtained at each followup clinic visit and entered contemporaneously into the computer databank. The details of this data set in regard to mortality have been reported previously (21). Clinical variables included blood pressure, body mass index, tender joint count (22), grip strength (23), morning stiffness (23), Health Assessment Questionnaire (HAQ) disability index score (24,25), visual analog scale (VAS) for pain, VAS for global severity, Arthritis Impact Measurement Scales (AIMS) anxiety and depression scales (26,27), erythrocyte sedimentation rate (ESR) (28), hemoglobin level, and rheumatoid factor (RF) status (29), among other assessments, as shown in Tables 1 and 2. Radiographs of the hands were generally obtained at 2-year intervals and were read using the Larsen method (30) by Dr. Arvi Larsen. Radiographic data were available for only 1,154 patients. Radiographic progression was defined as the rate of increase in the Larsen score per year, using the last radiographic assessment.

The patients had RA fulfilling the 1958 (31) or 1987 (32) American College of Rheumatology (formerly, the American Rheumatism Association) criteria. Data analysis was restricted to patients who attended the Wichita Arthritis Center at least twice between January 1, 1981 (when the HAQ and AIMS anxiety and depression scores first became available) and December 31, 1999. Of the 1,922 potentially eligible patients, 187 had all of their visits prior to 1981 and 58 patients had all of their visits after December 31, 1999, leaving 1,677 patients. Of those, 290 had only 1 clinic visit, leaving 1,387. In addition, in order to best assess the predictive ability of study variables, we considered data on patients to be "censored" if they were not seen in clinic within 2 years of their death.

Table 1. Demographic and clinical variables in 1,387 rheumatoid arthritis patients at study entry*

RA factors	
HAQ, 0-3	1.20 ± 0.76
Global severity, 0-10	4.98 ± 2.54
Pain, 0-10	4.88 ± 2.63
Depression, 0-10	2.95 ± 1.84
ESR, mm/hour	40.69 ± 26.84
Anxiety, 0-10	4.10 ± 2.07
Grip strength, mm Hg	116.81 ± 56.79
RF, latex titer	4.36 ± 2.81
Larsen radiographic score, slope†	2.80 ± 5.24
Disease duration, years	7.06 ± 8.52
Hematocrit, %	38.00 ± 4.57
RF, % positive	84.8
Nodules ever, %‡	31.1
Joint count, 0-18	6.05 ± 3.82
Non-RA factors	
Age, years	55.45 ± 14.40
Lifetime comorbidities, 0-11‡	2.58 ± 1.80
Sex, % female	73.3
High school graduates, %	81.1
Married, %	72.2
White, %	94.2

* Except where indicated otherwise, values are the mean ± SD. RA = rheumatoid arthritis; HAQ = Health Assessment Questionnaire; ESR = erythrocyte sedimentation rate; RF = rheumatoid factor.

† Available for 1,154 patients.

‡ Last or maximum value.

Specifically, if patients died within the 2-year period after their last visit they were counted as dead. If they were still alive more than 2 years after their last visit, or if they had died but their death occurred more than 2 years after their last visit, they were considered to be censored. By this method we were assured that the last recorded clinical assessments would be meaningfully related to death (21). There were no other exclusions. This resulted in a study population of 1,387 patients and 88,063 patient-months of observation. Within the period of study observation there were 212 deaths.

The primary outcome measure was all-cause mortality. Death was confirmed by review of medical records, death certificates, and the National Death Index (National Center for Health Statistics, US Department of Health and Human Services, Hyattsville MD). We obtained all available hospital records and all official death certificates from states in which there were decedents from our cohort, and coded specific cause of death according to the International Classification of Diseases, Ninth Revision.

Statistical methods. Data were analyzed using pooled logistic regression, a method that yields results equivalent to those obtained by Cox proportional hazards regression (33), as previously reported (21). Using this method, results are reported as odds ratios (ORs), in contrast to the hazard ratios reported with Cox regression analysis.

To understand the effect of observation time on mortality prediction, 4 methods of categorizing predictor variables were used. The first method used 1) all covariate observations. This approach of using time-varying covariates was the primary method of the study. In addition, to understand how well less-complete methods of covariate categorization might work,

Table 2. Univariate time-varying predictors of mortality in rheumatoid arthritis*

Variable	OR	Z score	P	95% CI	OR change per SD	BIC	OR fourth vs. first quartile (95% CI)	Contributing fraction (95% CI)†
RA factors								
HAQ, 0–3	2.93	11.12	0.000	2.43, 3.54	2.313	–283.34	5.98 (3.81, 9.36)	0.50 (0.40, 0.58)
Global severity, 0–10	1.28	8.48	0.000	1.21, 1.36	1.833	–222.33	3.85 (2.59, 5.95)	0.33 (0.24, 0.42)
Pain, 0–10	1.25	8.27	0.000	1.19, 1.32	1.800	–218.86	4.58 (2.86, 7.31)	0.32 (0.23, 0.40)
Depression, 0–10	1.34	8.82	0.000	1.26, 1.43	1.648	–217.14	4.46 (2.96, 6.71)	0.36 (0.27, 0.45)
Anxiety, 0–10	1.28	7.15	0.000	1.20, 1.38	1.620	–197.98	3.81 (2.54, 5.73)	0.29 (0.21, 0.37)
Grip strength, mm Hg	1.01	6.28	0.000	1.01, 1.01	1.918	–195.32	4.40 (2.80, 6.92)	0.35 (0.26, 0.43)
ESR, mm/hour	1.01	5.70	0.000	1.01, 1.02	1.375	–176.81	1.89 (1.26, 2.83)	0.19 (0.07, 0.30)
RF, latex titer	1.13	4.62	0.000	1.07, 1.19	1.423	–170.38	2.94 (1.85, 4.67)	0.17 (0.09, 0.24)
Hematocrit, %	1.06	3.80	0.002	1.03, 1.10	1.307	–161.93	1.71 (1.18, 2.46)	0.15 (0.04, 0.24)
Larsen radiographic score, 0–86‡	1.04	4.71	0.000	1.02, 1.06	1.112		1.51 (1.00, 2.27)	0.13 (NA)
Disease duration, years	1.01	2.10	0.036	1.00, 1.03	1.137	–151.80	1.41 (0.89, 2.22)	0.09 (–0.13, 0.20)
Nodules, yes/no	1.24	1.37	0.170	0.91, 1.68	NA	–149.56	NA	0.06 (–0.03, 0.14)
RF positive, yes/no	1.28	1.01	0.314	0.79, 2.07	NA	–148.81	NA	0.20 (–0.24, 0.48)
Joint count, 0–18	1.01	0.76	0.445	0.98, 1.05	1.055	–148.31	1.14 (0.99, 1.00)	0.03 (–0.06, 0.11)
Non-RA factors								
Age, years	1.09	11.89	0.000	1.07, 1.10	3.156	–159.21	23.59 (9.62, 57.89)	0.55 (0.47, 0.61)
Lifetime comorbidities, 0–11§	1.19	4.69	0.000	1.11, 1.29	1.329	–168.06	3.59 (2.16, 5.98)	0.19 (0.12, 0.26)
Male sex, yes/no	2.10	5.28	0.000	1.59, 2.76	NA	–159.21	NA	0.22 (0.12, 0.30)
Non-HS graduate, yes/no	1.26	1.40	0.163	0.91, 1.75	NA	–150.01	NA	–0.20 (NA)
Systolic BP, mm Hg	1.00	–0.13	0.894	0.99, 1.01	0.991	–147.75	0.85 (0.57, 1.26)	–0.06 (NA)
Diastolic BP, mm Hg	0.97	–0.42	0.673	0.98, 1.01	0.971	–147.91	0.95 (0.65, 1.41)	–0.01 (NA)

* Within rheumatoid arthritis (RA) factor and non-RA factor groups, variables are ranked by the Bayesian information criterion (BIC). Univariate predictor variables are adjusted for age and sex, except for age and sex, which are adjusted for sex and age, respectively. OR = odds ratio; 95% CI = 95% confidence interval; HAQ = Health Assessment Questionnaire; ESR = erythrocyte sedimentation rate; NA = not available (could not be calculated); HS = high school; BP = blood pressure.

† Reported values for contributing fraction for continuous variables represent mortality reduction for those patients in the fourth quartile, assuming patients were switched to the first quartile. For dichotomous variables (e.g., nodules, rheumatoid factor [RF] positivity), contributing fraction is the proportion of mortality attributed to having the characteristic.

‡ Available for 1,154 patients. BIC was not calculated for Larsen score.

§ Comorbidity data were regularly available beginning in 1990.

we also used 2) predictors that were present at the first observation (baseline predictors), 3) predictors that represented the mean of covariate observations over the first year of followup, and 4) predictors that represented the mean of covariate observations over the first 2 years of followup.

To compare the relative predictive ability of the study variables, 6 statistics were determined, as described below and shown in Tables 2 and 3. Although there is redundancy in these measures, they are often reported separately or in combination in clinical studies. It is beneficial to see all of the results together to understand their similarities and to understand which measures are most useful.

Statistics calculated. The odds ratio and its 95% confidence interval (95% CI) represent the increased or decreased risk of dying associated with a 1-unit change in the predictor variable for continuous variables. For dichotomous variables, the odds ratio represents the risk associated with having the characteristic compared with not having it.

It may be difficult to understand the actual differences among variables because variables can be scaled in different units; for example, a 1-unit change in age is not comparable with a 1-unit change in HAQ. The standardized odds ratio, or the odds ratio per 1-SD change, allows comparison among predictor variables using common units. However, such mea-

asures are not meaningful for categorical variables or those that are distinctly non-normal in their distribution.

The odds ratio of fourth-versus-first-quartile values of the predictor variable provides a relative measure of the maximum effect of the predictor, standardized by converting values to quartiles. It is not a meaningful approach for categorical and dichotomous variables, and it can be affected by the distribution of the variable.

The contributing fraction is a method that is allied to the population attributable fraction (34) or attributable risk (35,36). The population attributable fraction is an epidemiologic measure that attempts to quantify the proportion of disease incidence that is due to a particular factor or exposure. The contributing fraction is a hypothetical measure that describes the effect on disease incidence (or death), assuming that the predictor variables are causally related to the outcome. It is a hypothetical measure because the predictors may not be causally related to the outcome, but by treating them in that way, it is possible to gain insight into the quantitative effect of the predictor variables on mortality. The contributing fraction increases *both* as the exposure to abnormal levels of the predictor variables becomes more common *and* as the relative risk (odds ratio) becomes larger. The calculations of attributable risk and the contributing fraction are similar

Table 3. Univariate time-varying predictors of mortality in rheumatoid arthritis: analyses restricted to 325 patients with RA duration of ≤1 year at first observation*

Variable	OR	Z score	P	95% CI	OR change per SD	BIC	OR fourth vs. first quartile (95% CI)	Contributing fraction (95% CI)
RA factors								
HAQ, 0–3	1.85	2.44	0.015	1.13, 3.02	1.560	2.88	3.00 (1.14, 7.85)	0.19 (–0.19, 0.45)
Global severity, 0–10	1.17	2.13	0.033	1.01, 1.35	1.474	4.28	1.30 (0.88, 6.81)	0.30 (0.02, 0.51)
Pain, 0–10	1.26	3.37	0.001	1.10, 1.44	1.834	–2.85	3.75 (1.33, 10.62)	0.40 (0.11, 0.61)
Depression, 0–10	1.31	2.96	0.003	1.10, 1.57	1.562	0.84	2.79 (1.06, 7.34)	0.31 (–0.03, 0.5)
Anxiety, 0–10	1.15	1.53	0.127	0.96, 1.39	1.321	6.48	3.08 (0.94, 10.03)	0.23 (–0.04, 0.43)
Grip strength, mm Hg	1.01	2.31	0.021	1.00, 1.02	1.887	2.56	4.24 (1.37, 13.31)	0.28 (–0.04, 0.49)
ESR, mm/hour	1.01	0.81	0.419	0.99, 1.02	1.137	8.15	1.51 (0.50, 4.59)	0.12 (–0.25, 0.38)
RF, latex titer	1.30	3.40	0.001	1.12, 1.52	2.143	–5.36	10.02 (0.78, 12.03)	0.40 (0.15, 0.57)
Hematocrit, %	1.04	0.88	0.380	0.95, 1.15	1.194	8.00	1.61 (0.55, 4.68)	0.13 (–0.19, 0.37)
Larsen radiographic score, 0–86†	1.04	3.45	0.001	1.02, 1.06	1.177	5.22	1.72 (0.62, 4.77)	0.16 (–0.41, 0.49)
Disease duration, years	0.40	1.11	0.268	0.08, 2.03	0.279	7.49	0.46 (0.00, 2.74)	0.00 (NA)
Nodules, yes/no	1.28	0.53	0.597	0.51, 3.20	NA	8.50	NA	0.05 (–0.15, 0.21)
RF positive, yes/no	6.57	1.84	0.066	0.88, 49.06	NA	2.67	NA	0.82 (–0.26, 0.97)
Joint count, 0–18	1.00	0.05	0.963	0.94, 1.07	1.001	8.76	0.55 (0.20, 1.53)	–0.22 (NA)
Non-RA factors								
Age, years	1.11	5.34	0.000	1.07, 1.15	4.229	–1.22	0.16 (0.01, 3.38)	NA
Lifetime comorbidities, 0–11	1.08	0.75	0.451	0.88, 1.32	1.139	8.21	1.10 (0.30, 3.97)	0.59 (0.30, 0.76)
Male sex, yes/no	2.14	1.97	0.049	1.00, 4.60	NA	–1.22	NA	0.20 (–0.06, 0.40)
Non-HS graduate, yes/no	0.54	0.99	0.323	0.16, 1.83	NA	10.53	NA	–0.09 (NA)
Systolic BP, mm Hg	0.99	0.88	0.381	0.97, 1.01	0.844	7.99	0.88 (0.26, 3.06)	–0.00 (–0.37, 0.28)
Diastolic BP, mm Hg	0.98	1.32	0.188	0.95, 1.01	0.781	7.03	0.49 (0.15, 1.58)	–0.14 (NA)

* Within RA factor and non-RA factor groups, variables are ranked by the order of the BIC in Table 2. See Table 2 for explanations and definitions.
 † Available for 205 patients.

(37,38), but the methods differ in their interpretation. In the analyses of the data in Tables 2 and 3, the contributing fraction is interpreted as the fraction of mortality that would be prevented if all persons in the fourth quartile of the variable under study were switched to the lowest (first) quartile. The variables, therefore, can be compared with regard to their strength in detecting mortality effects on a population level.

The Z score and the allied P value are commonly reported. The Z score represents the logistic regression coefficient divided by its standard error, and provides an indication of variable statistical significance. For example, a critical value of 2 for the Z score leads to an approximate level of significance of 0.05. Although the Z score and the P value provide essentially the same information, we present both statistics here, since it is not possible to distinguish differences between variables when both are associated with P values of <0.001.

The Bayesian information criterion (39,40) is a goodness-of-fit measure of overall model fit and is a means to compare nested and non-nested models (41). Based on the log-likelihood of the logistic regression, Bayesian information criterion values are useful to compare different models, but values have no directly interpretable meaning. According to Raferty (40), differences between models of 0–2, 2–6, 6–10, and >10 for the Bayesian information criterion provide weak, positive, strong, and very strong evidence, respectively, for the superiority of one model compared with another. Bayesian information criterion variable statistics can be compared only when sample sizes are the same (40).

Graphic representations. The relationship between the probability of dying and various levels of predictor variables

are shown in Figures 1–5. The horizontal line in each figure represents the baseline risk of dying in the absence of covariate information, or the line of no information. Risks are expressed as the monthly risk of dying. To calculate the risk of dying adjusted for the covariate, each covariate is divided into 20 groups and the risk for each group is obtained by logistic regression. Lines are connected by cubic splines.

Overall, 4.2% of data points for the variables in Table 1 were missing. Variables for which >2.5% of data points were missing were patient global assessment (3.6%), depression (13.5%), anxiety (14.1%), ESR (13.6%), and hematocrit level (11.5%). A single missing value in 1 of the 10 anxiety and depression questions invalidates the questionnaire, explaining the higher rate of missing data for these variables. For 70.0% of all observations, there were no missing data. Missing variables were replaced using the last observation carried forward method. Data for comorbidity were collected regularly beginning in 1990. In these analyses, missing comorbidity data prior to 1990 were imputed using the first available comorbidity data. For patients seen *only* prior to 1990, comorbidity values were imputed using the mean comorbidity values for the cohort by sex. To describe potential biases that might arise from using this imputation, separate analyses were conducted using only data obtained after 1990, when comorbidity data were complete.

Data from univariate analyses were adjusted for age and sex. The “best” multivariable models were fit by starting with models that included age, sex, HAQ score, and methotrexate treatment (time-varying covariates), and then adding additional variables while considering interactions, mediating

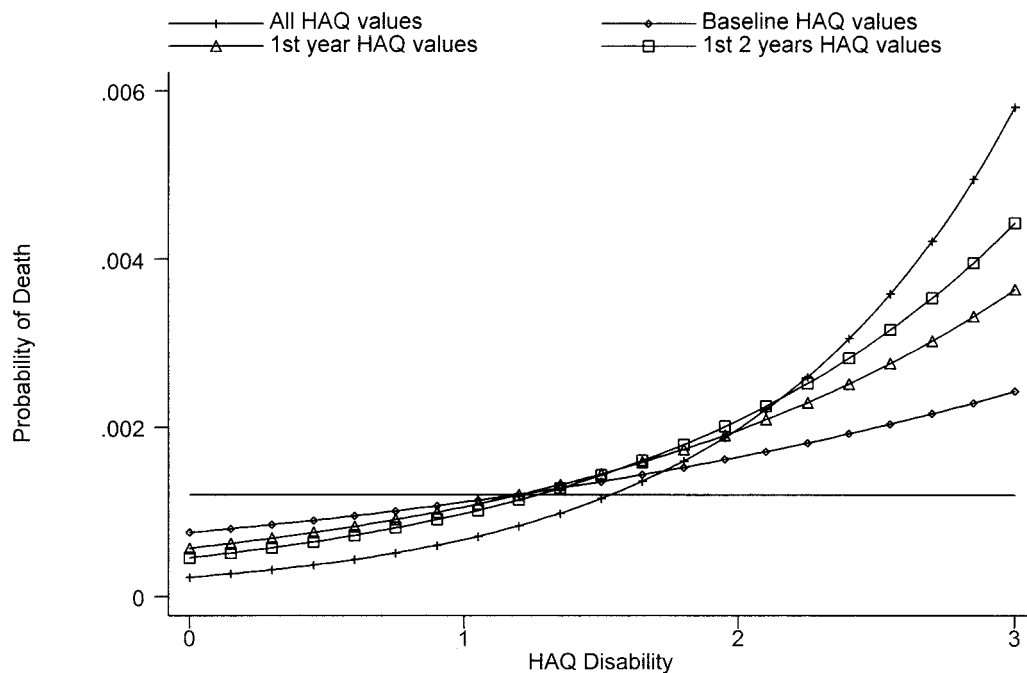


Figure 1. Monthly probability of dying by score on the Health Assessment Questionnaire (HAQ) disability scale, according to whether data used the first observation, the average of observations during the first study year, the average of observations during the first 2 study years, or all observations during the course of the patient's rheumatoid arthritis. Horizontal line represents the line of no information (baseline probability of dying). Increasing the period of observation results in clinically important increased information relative to predicting death. Note that the HAQ lines cross the line of no information in the approximate center of the x-axis, demonstrating that both low and high values of HAQ contribute information about mortality. The crossing point is near the median HAQ value at the time of death.

factors, clinical relevance, and causality. Among disease-modifying antirheumatic drugs (DMARDs), we adjusted for methotrexate use in our model since, in our previous work, utilizing appropriate methods to assess time-dependent confounding issues (21), it was the only DMARD that affected mortality. In addition, our estimates did not change materially when we further adjusted for other DMARDs. The significance level of all analyses was set at 0.05, and all tests were 2-tailed. Statistical computations were performed using Stata, version 7.0.

RESULTS

Baseline demographic and clinical characteristics. The baseline demographic and clinical variables in the 1,387 RA patients are presented in Table 1. At the time of the first study observation, the mean disease duration was 7 years (median 4 years), and 23% and 36% of the patients had a disease duration of <1 year and <2 years, respectively. Approximately 73% were female and 85% were RF positive.

Univariate predictors of mortality. Table 2 ranks the study variables according to their ability to predict

mortality, using the Bayesian information criterion as the primary ranking statistic. The values in the table are adjusted for age and sex (except when age and sex are studied separately) and include a time variable appropriate for the pooled logistic regression. Six measures of predictive strength were available, and, as might be expected, there was general agreement among the ranking statistics. In addition, the predictive effects of HAQ, ESR, depression, Larsen radiographic score, and RF titer are shown in Figures 1–5.

HAQ disability was by far the most important univariate predictor of mortality. Quantitatively, a 1-SD change in the HAQ resulted in a 26.2% greater increase in the odds ratio for mortality compared with a 1-SD change in global disease severity, the next most powerful predictor of mortality. Considering the contributing fraction, mortality would be reduced by 50% for the HAQ and by 33% for global disease severity if patients in the fourth quartile for these variables were switched to the first quartile. Determined roughly, the HAQ was a better predictor by one-quarter to one-third than the

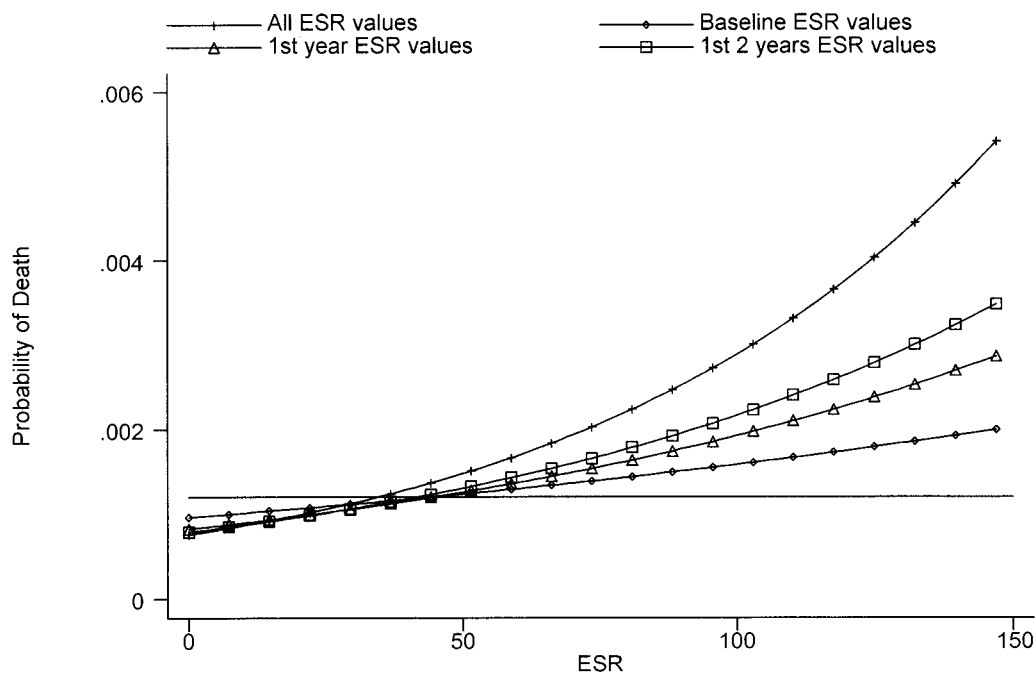


Figure 2. Monthly probability of dying by erythrocyte sedimentation rate (ESR), according to whether data used the first observation, the average of observations during the first study year, the average of observations during the first 2 study years, or all observations during the course of the patient’s rheumatoid arthritis. Horizontal line represents the line of no information (baseline probability of dying). Increasing the period of observation results in clinically important increased information relative to predicting death. Note that the ESR lines cross the line of no information on the left of the x-axis, demonstrating that high values of ESR contribute the most information. However, relatively few patients had such high values. The crossing point is near the median ESR value at the time of death.

next ranked variable when standardized odds ratios or fourth-versus-first-quartile values of the contributory fraction were used.

Among the clinical and laboratory variables a general pattern could be seen, in which the HAQ stood alone as the best predictor, followed by global disease severity, pain, depression, anxiety, and grip strength—variables that may be considered to be patient derived. For this second set of variables the contributory fraction ranged from 0.36 to 0.29. Laboratory variables were less important, with contributory fractions ranging from 0.19 to 0.15.

Dichotomous variables cannot be compared directly with continuous variables using the standardized odds ratio or the contributing fraction. However, it is possible to estimate their relative position by examining the Bayesian information criterion. Thus, nodules and RF positivity were positioned toward the lesser-effect end of the spectrum. Because the number of observations differed for the Larsen score slope compared with the other variables, the Bayesian information criterion is

not a good ranking statistic, and the Larsen score was ranked according to the contributing fraction. Radiographic progression rates represented weak predictors of mortality as evidenced by the standardized odds ratio and the fourth-versus-first-quartile odds ratio.

The strongest predictor of mortality was age, which exhibited the greatest contributing fraction (0.55). The Bayesian information criterion for age and sex is not valid as a comparison statistic for clinical variables because the model is smaller (lacks the clinical variables). The comorbidity score was the equivalent of mid-ranked clinical variables in its predictive ability concerning mortality. Because complete comorbidity data were available only after 1990, we also conducted analyses of the predictive ability of comorbidity with a data set restricted to data from 1990 or later. The standardized odds ratio was 1.306, the contributing fraction was 0.258, and the fourth-versus-first-quartile OR was 2.93 (95% CI 1.59, 5.41). Overall, these values were only slightly different from the values in Table 2.

To determine if study results differed in patients

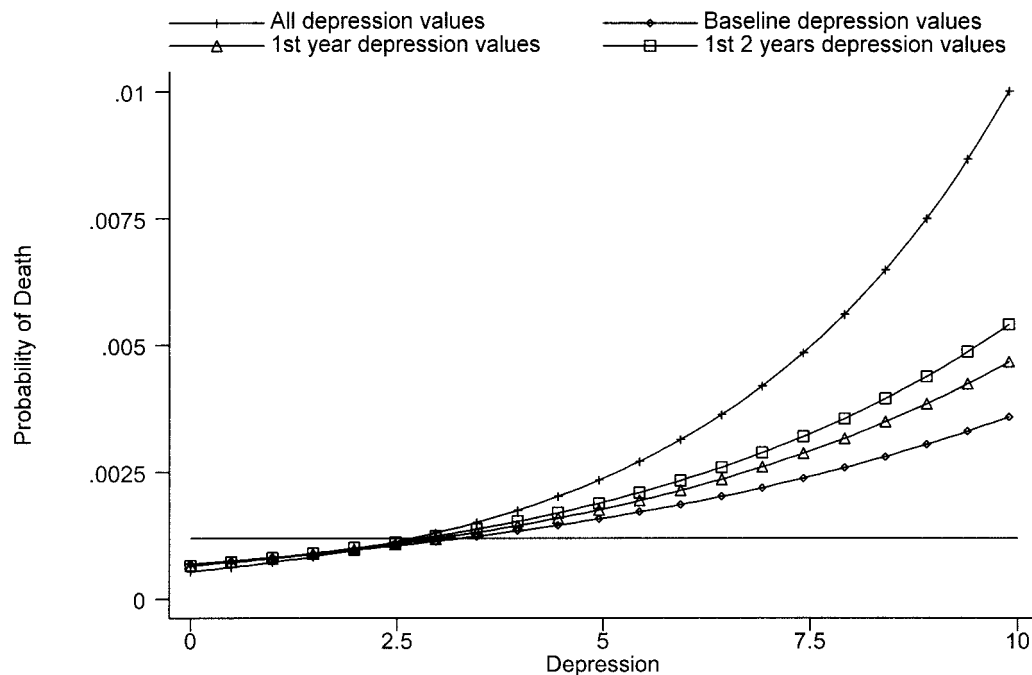


Figure 3. Monthly probability of dying by level of depression, according to whether data used the first observation, the average of observations during the first study year, the average of observations during the first 2 study years, or all observations during the course of the patient's rheumatoid arthritis. Horizontal line represents the line of no information (baseline probability of dying). Increasing the period of observation results in clinically important increased information relative to predicting death. Note that the depression lines cross the line of no information on the left of the x-axis, demonstrating that high values of depression contribute the most information. However, relatively few patients had such high values. The crossing point is near the median depression value at the time of death.

with recent-onset RA compared with those who had a longer duration of disease, we compared the odds ratios of the clinical variables (as logit coefficients) in the 325 patients who had RA of ≤ 1 year's duration at their first observation versus those who had >1 year's duration of RA at their first observation. For each clinical variable, there was no statistically significant difference in the predictor coefficient between disease duration groups. Because the actual data on this set of patients with early RA may be of interest to some readers, they are presented separately in Table 3.

We tested whether the logistic regression coefficients for the variables in Table 2 differed according to age or sex. There were no significant differences, indicating no effect of age and sex on the association between the study variables in Table 2 and mortality. Of note, the OR for the effect of HAQ on mortality was 3.37 for women and 2.49 for men ($P = 0.089$). This suggests the possibility that HAQ is more predictive of mortality in women than men.

Multivariable models of mortality. Table 4 displays the "best" multivariable models of mortality, along with a separate analysis based on the smaller number of patients ($n = 1,154$) for whom radiographs were available. Although global disease severity and pain were strong univariate predictors, they were not significant in the model as long as depression was present. In addition, the inclusion of depression rather than global disease severity or pain led to a slightly better fit. Even so, depression, global disease severity, or pain all worked almost equivalently as well as one another. Except for radiographic progression, none of the predictors that were significant in the univariate model were significant in the multivariable model. The relative predictive strength of the clinical variables is shown in Figure 6.

Rates of radiographic progression are highly dependent on estimated duration of RA, and apparently high rates can occur early in RA if disease duration is underestimated. In addition, rates are unstable as disease duration becomes smaller and approaches 0. To

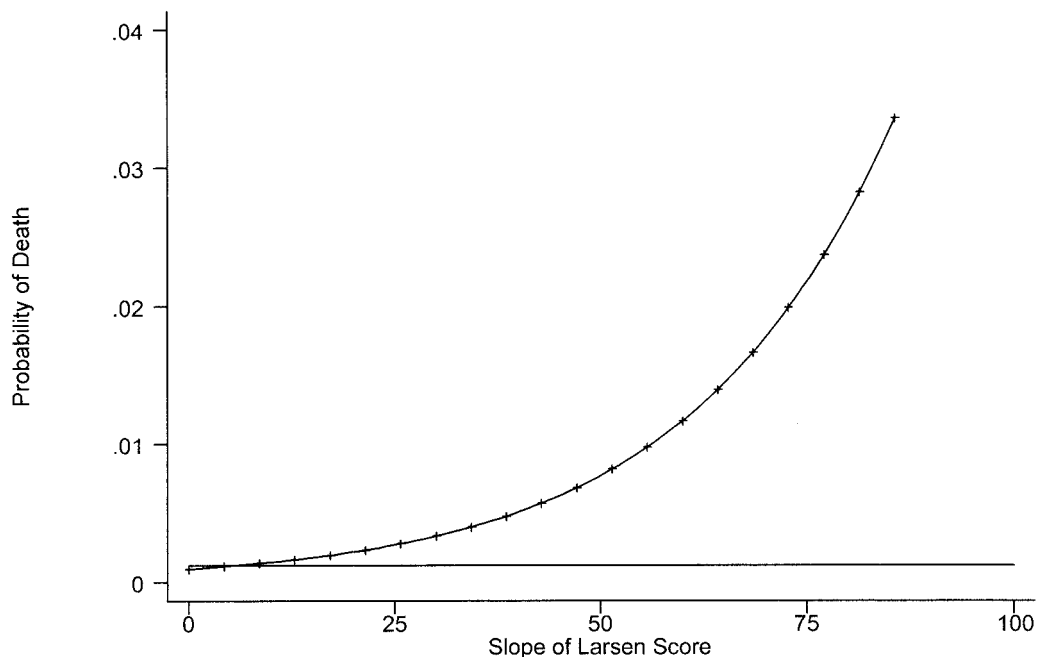


Figure 4. Monthly probability of dying by Larsen radiographic score. Horizontal line represents the line of no information (the baseline probability of dying). Note that the Larsen slope score has little effect at the commonly achieved values; relatively few patients had high values. The crossing point of the Larsen slope score and the line of no information is near the median Larsen slope value at the time of death.

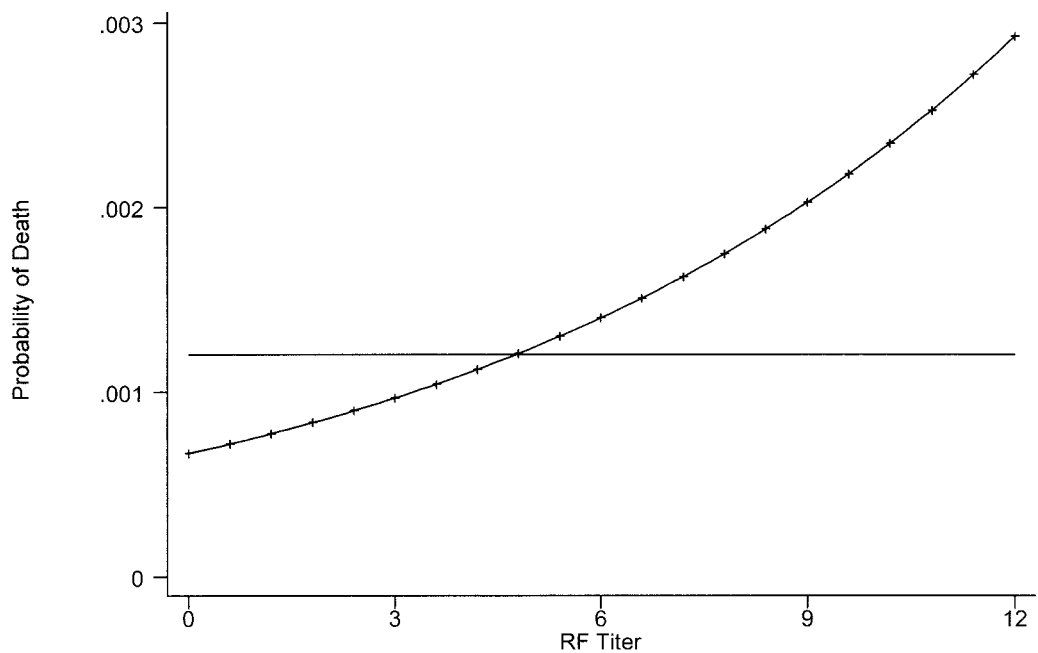


Figure 5. Monthly probability of dying by rheumatoid factor (RF) titer. Horizontal line represents the line of no information (baseline probability of dying). The crossing point of the RF titer and the line of no information is near the median RF titer for the study patients.

Table 4. Final multivariable models of mortality in rheumatoid arthritis with radiographic data not included (n = 1,387) and included (n = 1,154)*

	OR	SE	Z score	P	95% CI
Without radiographic data					
RA and disease factors					
HAQ disability, 0-3	2.22	0.24	7.31	0.000	1.79, 2.75
Depression, 0-10	1.12	0.04	3.05	0.002	1.04, 1.21
Lifetime comorbidities, 0-11	1.13	0.04	3.10	0.002	1.05, 1.22
RF, latex titer	1.08	0.03	2.97	0.003	1.03, 1.14
ESR mm/hour	1.01	0.00	2.17	0.030	1.00, 1.01
Methodretrate treatment, 0/1	0.51	0.01	-3.84	0.000	0.37, 0.72
Demographic factors					
Age, years	1.07	0.01	9.56	0.000	1.06, 1.09
Sex, male	2.76	0.42	6.60	0.000	2.04, 3.37
With radiographic data					
RA and disease factors					
HAQ disability, 0-3	1.97	0.25	5.40	0.000	1.54, 2.52
Depression, 0-10	1.17	0.05	3.55	0.000	1.07, 1.27
Lifetime comorbidities, 0-11	1.14	0.05	3.12	0.002	1.05, 1.24
RF, latex titer	1.10	0.03	3.13	0.002	1.04, 1.17
ESR, mm/hour	1.01	0.00	2.65	0.008	1.00, 1.01
Radiographic score, slope of Larsen score	1.03	0.01	2.93	0.003	1.01, 1.05
Methodretrate treatment, 0/1	0.56	0.10	-3.21	0.001	0.39, 1.09
Demographic factors					
Age, years	1.07	0.01	8.06	0.000	1.05, 1.09
Sex, male	2.73	0.47	5.86	0.000	1.95, 3.83

* Table also includes terms for time (months and months squared) and disease duration at first study assessment. See Table 2 for definitions.

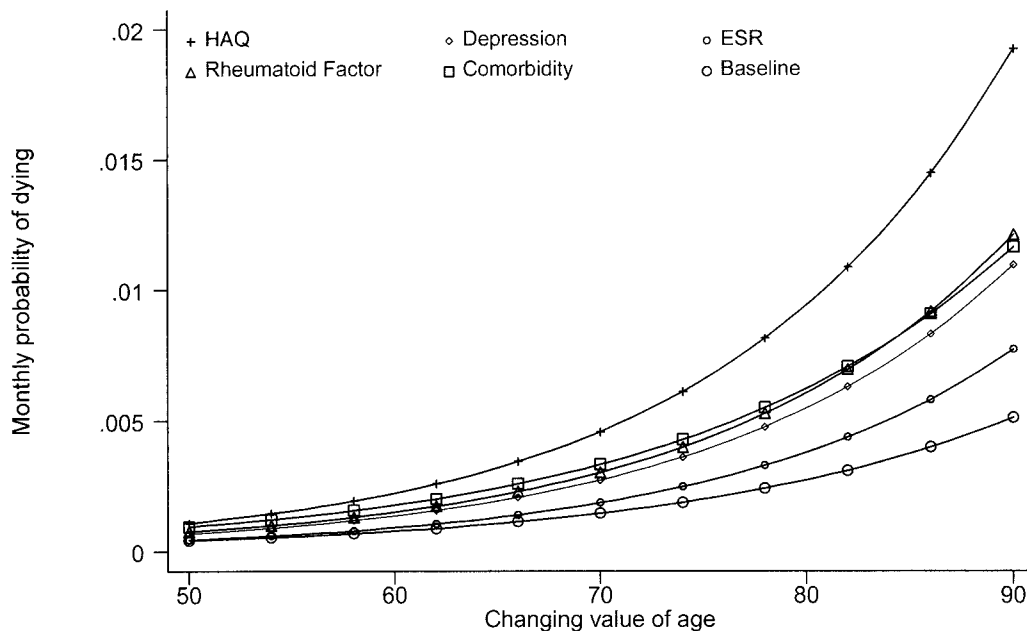


Figure 6. Relative strength of the clinical variables that were significant in the multivariable model as mortality predictors in rheumatoid arthritis. Shown is the monthly probability of dying among patients with values in the fourth quartile of each predictor, with the other variables set to their mean values. The baseline line represents the line of no information, or the probability of dying without covariate information. The Health Assessment Questionnaire (HAQ) was the best predictor, the erythrocyte sedimentation rate (ESR) was the weakest predictor, and the other variables were in intermediate positions.

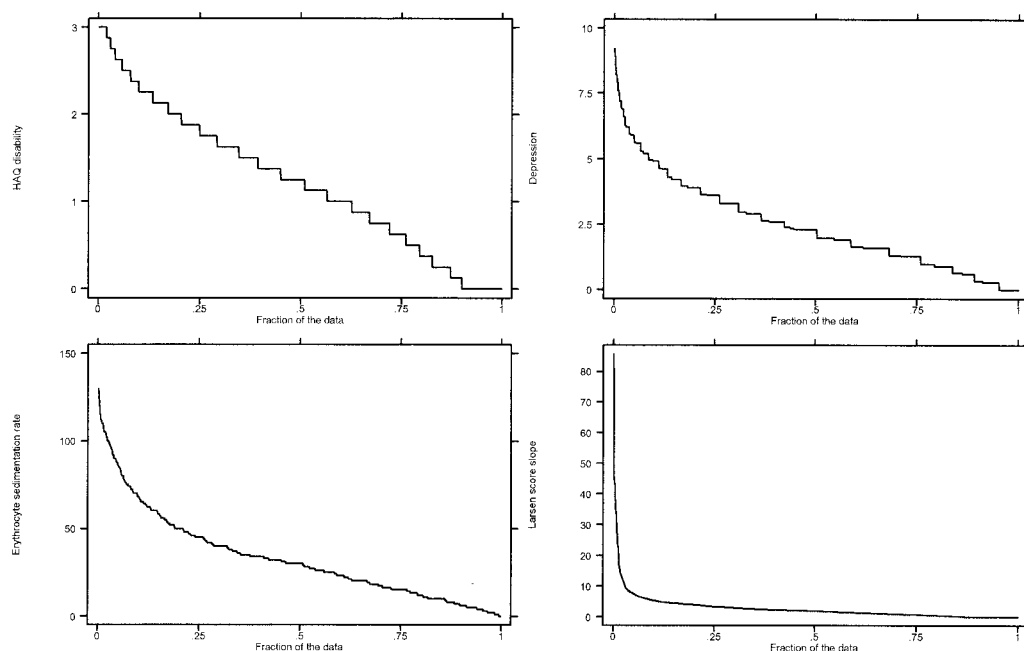


Figure 7. Quantile plots illustrating the distribution of 4 predictor variables. A variable that is normally distributed will appear as a relatively straight diagonal line, such as is demonstrated for the Health Assessment Questionnaire (HAQ) disability scale. Distributions in which most of the patients have low scores will have a curve bending toward the bottom left. The curve for erythrocyte sedimentation rate shows that high scores are experienced by few patients, as does the curve for depression. The extreme distribution for the Larsen radiographic score slope shows that high Larsen scores are very rare. These graphs emphasize that, except in the case of the HAQ, the increased risk of mortality accorded by high scores is experienced by relatively few patients.

account for these possible effects, we performed separate analyses of radiographic progression after excluding patients with RA of <5 years' duration. The effect of radiographic progression increased in the univariate analysis (OR 1.22 [95% CI 1.13, 1.32]) but was weaker in the multivariable analysis (OR 1.12 [95% CI 1.02, 1.24]).

Prediction of mortality and the length of observation. Figures 1–3 demonstrate that single-observation prediction, in this case the use of baseline values, results in impaired predictive ability. For time-varying covariate data versus 2-year covariate data versus 1-year covariate data versus baseline covariate data (baseline), examination of the Bayesian information criterion indicated an increasingly better fit for the models with longer periods of covariate observation. For the HAQ, the Bayesian information criterion results for the respective 4 observation groups were –283, –200, –181, and –164. For the ESR, the values were –176, –163, –159, and –152, and for depression, they were –217, –176, –173, and –171. Although, as shown in Figures 1–3, use of average values from the first year or the first 2 years increases predictive ability and is to be preferred strongly over a

baseline measurement, longitudinal observation adds much more information. Of additional interest, only the HAQ differed in predictive ability at values below the median (see Figure 7). There appeared to be little difference in predictive ability of the 4 methods of covariate observation for ESR values <40 mm/hour and for depression scores <3; this is another reason for the overall superiority of the HAQ.

DISCUSSION

The data from this large prospective study demonstrate the extent to which clinical and demographic variables differ in their ability to predict mortality in patients with rheumatoid arthritis. There appears to be a clear hierarchy. By far, the HAQ is the best predictor among all variables we studied. The HAQ is predictive across its full range, compared with variables like the ESR, which are predictive only at higher levels and, therefore, in fewer patients. The second tier of variables is also patient based: global disease severity, pain, depression, and anxiety. Grip strength, a variable that

has fallen out of favor, is the approximate equal of these second-tier variables in univariate analyses. The third tier of clinical variables are those that represent laboratory and radiographic tests. These variables provide much less predictive information about mortality than the HAQ and second-tier variables. Finally, there are clinical variables that produced very little predictive information, including disease duration, nodules, and tender joint count.

Even in the multivariable model, the HAQ and other patient self-report measures were significantly better than the rest. When these variables were analyzed together with demographic variables in a multivariable model, a core grouping of variables emerged that represented different aspects of RA in predicting mortality. These variables might be said to represent function (HAQ), severity or personal consequences (depression, pain, and global severity), inflammatory activity (ESR), seropositivity (RF titer), and damage (radiographic erosions). While our univariate and multivariable analysis results have obvious implications for research, they also should be of signal importance in the clinic, where laboratory and radiographic measures currently dominate clinical observation and self-report data are uncommonly acquired.

These results have immediate practical value. For example, the OR per unit change of the HAQ (2.93) can be used to provide insight into disease severity in patients in the clinic and in clinical trials. For example, the mean HAQ score of 1.80 for patients entering the ATTRACT trial of infliximab and methotrexate (42) indicates that such patients have a risk of mortality that is 5.3 times greater than that for persons without disability (HAQ score 0) or, in practical terms, without RA.

Although the HAQ is now in common use in almost all research studies, relatively few mortality studies have examined its comparative predictive value. Using baseline modified HAQ (M-HAQ) values (43), Callahan et al described an odds ratio for mortality of 2.14 in 1,416 RA patients from 15 private practices in the US (14). Except for disease duration, none of the other clinical variables in Table 2 were reported in their study. Callahan and colleagues also reported the results of baseline Cox regression analyses for 210 RA patients seen in Nashville, Tennessee and followed up 5 years later (15). In univariate analyses, the OR for the M-HAQ was 2.0, and comorbidity score, disease duration, education, and ESR were significantly associated with mortality. Joint count, radiographic changes, RF status, pain, and global scores were not associated with mortality in their analyses (15). An earlier study from

that group used a precursor of the HAQ/M-HAQ and found it to be the best mortality predictor, but because of the use of different study questionnaires, results are not comparable (16). Soderlin et al reported 5-year mortality in 102 RA patients from northern Finland (17). The HAQ was a significant predictor of mortality in Cox analyses, but data on the effect of the HAQ as a continuous variable are not provided. Using a Weibull baseline model with only survey data available ($n = 263$), Leigh and Fries found the HAQ, but not pain, to be associated with mortality (18). Wolfe et al reported mortality results estimated from baseline values from Stanford, California, Saskatoon, Canada, and Wichita, Kansas, from studies using Cox regression analysis (12). Depending on the center, various baseline questionnaire and laboratory data were available, and HAQ, pain, ESR, grip strength, and joint count were significantly associated with mortality in most centers where the data on the variables were available.

Our analysis applied the epidemiologic concept of attributable risk to these identified mortality risk factors in RA patients and demonstrated a high contributing fraction from the HAQ (i.e., 50% mortality reduction with the HAQ in the fourth quartile switched to that in the first quartile). Why does the HAQ have such a high contributing fraction statistic compared with other clinical measures? In addition to the biologically important effect of loss of function, Figures 1–5 and Figure 7 add further insight. As shown in the quantile plots of Figure 7, the HAQ is quite uniformly distributed over its range, and from Figure 1 it can be seen that the values of HAQ that are associated with substantial increased risk occur frequently in the study population. In contrast, the effect and distribution of ESR (Figure 2), depression (Figure 3), and Larsen slope score (Figure 4) are increasingly skewed, such that the values associated with high risks are found in only a small number of patients. For example, 95% of patients have scores of ≤ 5 for the Larsen slope. Although the RF titer effect is evenly distributed throughout the population (Figure 5), it is not a strong predictor. At its maximum value in Figure 5, the monthly probability of death (0.003) is less than half that for the HAQ at its maximum (0.006), as shown in Figure 1. In addition, the RF titer contributes only moderately according to the contributing fraction statistic.

Why do other variables that are explanatory in the univariate models not appear in the multivariable model? As indicated above, global disease severity and pain have value approximately equal to that of depression in predicting mortality, but they do not appear in

the multivariable model because they are collinear and are slightly “worse” predictors. In clinical practice or in research, any of these 3 variables could be substituted for one another. Although education level, in the form of high school graduate/not high school graduate, separate category levels, or a continuous variable, is a univariate predictor, its effect is distributed over measures like HAQ, depression, and comorbidity. Thus, when data on these variables are available, the additional information about education level adds nothing to the ability to predict mortality.

Another objective of this study was to provide information regarding how many observation points are necessary for clinical prediction, by evaluating the values of repeated measures. As shown in Figures 1–3, a single baseline observation provides the least information, while increasing duration of observation is associated with substantially more accurate prediction.

We present an often-duplicative series of statistics in an attempt to understand quantitative prediction in RA. Because the sample size was the same for all of the clinical variables except radiographic progression, presenting many statistics allowed for their comparison. From Table 2 it can be seen that it is difficult to judge predictive ability or strength on the basis of odds ratios, since the differing units make interpretation extremely difficult. Similarly, *P* values cannot distinguish ranking. Both the *Z* score and the Bayesian information criterion are dependent on sample size. The fourth-versus-first-quartile odds ratio provides information regarding prediction, but is dependent on the distribution of scores. Perhaps most useful, though not for dichotomous data, is the standardized odds ratio, because it is unaffected by sample size. We also found the contributing fraction to be helpful in gauging relative usefulness of the study variables, and it is a method that deserves additional study. Although we have not presented our analyses that determined the attributable fraction by comparing all patients with the mean of the variable under study, that method provided ranked results similar to those presented for the contributing fraction in this study. Such a method, as suggested by Brady (37), is conceptually attractive and merits further study.

The difference between odds ratios for the effect of HAQ on mortality by sex (OR 3.37 for women and 2.49 for men; *P* = 0.089) might possibly be explained by differential item response and, therefore, might be more apparent than real when HAQ scores rather than functional ability are considered. Preliminary work from our group using Rasch analysis suggests that women and men perceive certain HAQ items to have different levels

of difficulty. Such perceptive differences might also be the explanation for the difference in HAQ scores between men and women that has been noted in all studies. This important measurement issue should be explored further.

For clinicians, an important message of this study is that self-report measures, particularly the HAQ, enhance understanding and prediction of long-term outcomes of RA. Such data are easy to acquire and provide considerably more information about these outcomes than do the traditional clinical and laboratory examinations. In addition, our findings underscore differences in the importance of individual variables in randomized clinical trials, where joint examination data, laboratory data, and radiographic data are essential, and in clinical practice, where they are of considerably less use. The data in this report also emphasize that observation of RA patients over 1–2 years provides considerable insight into the long-term outcome in each patient, and the longer one follows up patients the more reliable are the clinical measures. In contrast, single observation data are of little use.

The results also suggest that dichotomizing patients into severity groups according to HAQ score (e.g., HAQ ≥ 1) or other clinical variables underestimates the contribution to mortality of even slight increases in the HAQ above 0, as shown in Figure 1. In this respect, HAQ scores might be considered analogous to blood pressure measures, where lower is better and there is no strict cutoff between normal and not normal or between desirable and undesirable. Finally, these data should provide some insight into the issue of the definition of DMARD failure and DMARD/biologic agent success, since even small increases above 0 for variables such as the HAQ and patient global assessment presage loss of years of life. Although important treatment advances have occurred, much remains to be done if the long-term outcomes of RA are to be improved. Self-report data such as the measures described here can provide clinicians with additional insight into the status of each individual patient.

In summary, the HAQ was the most powerful predictor of mortality, followed by other patient self-report variables. Laboratory, radiographic, and physical examination data were substantially weaker in predicting mortality. We recommend that clinicians collect patient self-report data, since these data produce more useful clinical outcome information than other available clinical measures.

REFERENCES

1. Chehata JC, Hassell AB, Clarke SA, Matthey DL, Jones MA, Jones PW, et al. Mortality in rheumatoid arthritis: relationship to single and composite measures of disease activity. *Rheumatology (Oxford)* 2001;40:447–52.
2. Gabriel SE. The epidemiology of rheumatoid arthritis. *Rheum Dis Clin North Am* 2001;27:269–81.
3. Gordon P, West J, Jones H, Gibson T. A 10 year prospective followup of patients with rheumatoid arthritis 1986–96. *J Rheumatol* 2001;28:2409–15.
4. Pincus T, Sokka T, Wolfe F. Premature mortality in patients with rheumatoid arthritis: evolving concepts [editorial]. *Arthritis Rheum* 2001;44:1234–6.
5. Riise T, Jacobsen BK, Gran JT, Haga HJ, Arnesen E. Total mortality is increased in rheumatoid arthritis: a 17-year prospective study. *Clin Rheumatol* 2001;20:123–7.
6. Gabriel SE, Crowson CS, OFallon WM. Mortality in rheumatoid arthritis: have we made an impact in 4 decades? *J Rheumatol* 1999;26:2529–33.
7. Lindqvist E, Eberhardt K. Mortality in rheumatoid arthritis patients with disease onset in the 1980s. *Ann Rheum Dis* 1999;58:11–4.
8. Myllykangasluosujarvi R, Aho K, Isomäki HA. Mortality in rheumatoid arthritis. *Semin Arthritis Rheum* 1995;25:193–202.
9. Symmons D. Excess mortality in rheumatoid arthritis: is it the disease or the drugs? *J Rheumatol* 1995;22:2200–2.
10. Coste J, Jouglu E. Mortality from rheumatoid arthritis in France, 1970–1990. *Int J Epidemiol* 1994;23:545–52.
11. Isomaki HA. Mortality in patients with rheumatoid arthritis. In: Wolfe F, Pincus T, editors. *Rheumatoid arthritis: pathogenesis, assessment, outcome, and treatment*. New York: Marcel Dekker; 1994. p. 235–46.
12. Wolfe F, Mitchell DM, Sibley JT, Fries JF, Bloch DA, Williams CA, et al. The mortality of rheumatoid arthritis. *Arthritis Rheum* 1994;37:481–94.
13. Pincus T. Is mortality increased in rheumatoid arthritis? *J Musculoskel Med* 1988;5:27–46.
14. Callahan LF, Cordray DS, Wells G, Pincus T. Formal education and five-year mortality in rheumatoid arthritis: mediation by helplessness scale score. *Arthritis Care Res* 1996;9:463–72.
15. Callahan LF, Pincus T, Huston JW III, Brooks RH, Nance EP Jr, Kaye JJ. Measures of activity and damage in rheumatoid arthritis: depiction of changes and prediction of mortality over five years. *Arthritis Care Res* 1997;10:381–94.
16. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* 1994;120:26–34.
17. Soderlin MK, Nieminen P, Hakala M. Functional status predicts mortality in a community based rheumatoid arthritis population. *J Rheumatol* 1998;25:1895–9.
18. Leigh JP, Fries JF. Mortality predictors among 263 patients with rheumatoid arthritis. *J Rheumatol* 1991;18:1307–12.
19. Pincus T, Callahan LF, Vaughn WK. Questionnaire, walking time and button test measures of functional capacity as predictive markers for mortality in rheumatoid arthritis. *J Rheumatol* 1987;14:240–51.
20. Pincus T, Callahan LF, Sale WG, Brooks AL, Payne LE, Vaughn WK. Severe functional declines, work disability, and increased mortality in seventy-five rheumatoid arthritis patients studied over nine years. *Arthritis Rheum* 1984;27:864–72.
21. Choi HK, Hernan MA, Seeger JD, Robins JM, Wolfe F. Methotrexate therapy and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet* 2002;359:1173–7.
22. Thompson PW, Hart LE, Goldsmith CH, Spector TD, Bell MJ, Ramsden MF. Comparison of four articular indices for use in clinical trials in rheumatoid arthritis: patient, order and observer variation. *J Rheumatol* 1991;18:661–5.
23. American Rheumatism Association Glossary Committee. *Dictionary of the rheumatic diseases*. Vol. 1: Signs and symptoms. New York: Contact Associates; 1982.
24. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
25. Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. *Arthritis Rheum* 2000;43:2751–61.
26. Meenan RF, Gertman PM, Mason JH, Dunaif R. The Arthritis Impact Measurement Scales: further investigations of a health status measure. *Arthritis Rheum* 1982;25:1048–53.
27. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. *Arthritis Rheum* 1980;23:146–52.
28. Wolfe F. Comparative usefulness of C-reactive protein and erythrocyte sedimentation rate in patients with rheumatoid arthritis. *J Rheumatol* 1997;24:1477–85.
29. Wolfe F, Cathey MA, Roberts FK. The latex test revisited: rheumatoid factor testing in 8,287 rheumatic disease patients. *Arthritis Rheum* 1991;34:951–60.
30. Larsen A. How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in longterm studies? *J Rheumatol* 1995;22:1974–5.
31. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA. 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull Rheum Dis* 1958;9:175–6.
32. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
33. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med* 1990;9:1501–5.
34. Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum* 1953;9:531–41.
35. Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res* 2001;10:231–8.
36. Gefeller O. Definitions of attributable risk—revisited. *Public Health Rev* 1995;23:343–55.
37. Brady AR. Adjusted population attributable fractions from logistic regression. *Stata Technical Bulletin* 1998;42:137–43.
38. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics* 1993;49:865–72.
39. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461–4.
40. Raftery AE. Bayesian model selection in social research. *Soc Methodol* 1995;25:111–63.
41. Long JS, Freeze J. *Regression models for categorical dependent variables using Stata*. College Station (TX): Stata Press; 2001.
42. Lipsky PE, van der Heijde DMFM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, et al, for the Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis With Concomitant Therapy Study Group. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *N Engl J Med* 2000;343:1594–602.
43. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. *Arthritis Rheum* 1983;26:1346–53.