A Prognostic Model for Functional Outcome in Early Rheumatoid Arthritis

NICK BANSBACK, ADAM YOUNG, ALAN BRENNAN, and JOSH DIXEY

ABSTRACT. ***Objective***. To construct a prognostic algorithm to predict 5-year functional outcome in rheumatoid arthritis (RA), based on the Health Assessment Questionnaire (HAQ).

Methods. Data from all patients with 5-year followup (n = 985) were used from an inception cohort, the Early Rheumatoid Arthritis Study (ERAS). Possibly relevant prognostic factors considered in the initial stage of the model-building process were standard clinical, radiological, and laboratory features measured at baseline and at 1 year. Multivariate analysis was performed using logistic regression, and the predictive performance of the model was tested using measures of discrimination and calibration.

Results. Bootstrap resampling identified 6 variables that consistently predicted severe functional outcome. Functional grade III/IV (odds ratio 6.7) and HAQ at 1 year (odds ratio 2.4) were the most important. Other variables included socioeconomic status, hemoglobin, and radiographic and disease activity scores. Estimates of the regression coefficients and performance were corrected for over-fitting. Reasonably large values for the c-index (0.82) and the Nagelkerke R^2 (0.39) indicate that the set of prognostic factors explains the variation in outcome to a degree that implies good prediction for individual patients.

Conclusion. The algorithm identifies patients in the first year of RA who are likely to have poor function by 5 years and who could potentially benefit from aggressive drug therapy. A nomogram is produced for simple application of the model in clinical practice. While further external validation is necessary, this model could allow clinicians to target aggressive therapy earlier in a patient's disease course. (First Release July 1 2006; J Rheumatol 2006;33:1503–10)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
HEALTH ASSESSMENT QUESTIONNAIRE
FUNCTIONAL OUTCOME
PROGNOSIS

Rheumatoid arthritis (RA) is a chronic systemic inflammatory disorder characterized by a symmetrical polyarthritis of varying extent and severity, associated with synovitis of joint and tendon sheaths, articular cartilage loss, erosion of juxtaarticular bone, and in most patients, the presence of IgM rheumatoid factor (RF) in serum. Antirheumatic therapies are employed to ameliorate symptoms and prevent irreversible joint damage and deformity, and their use is based on various studies that have shown these beneficial effects, in the short term at least. The relative effectiveness in the long term of standard disease modifying drugs (DMARD) is based on longitudinal observational studies because it has proved difficult to maintain randomized clinical trials (RCT) for more than 2 to 3 years. Such studies have shown that differences between the main DMARD are mainly in side effect profiles rather than efficacy, and patients who are maintained on these drugs have better longterm outcomes[1].

It is common clinical practice now to use "step up" and combined regimens (e.g., triple therapy with or without steroids) that include methotrexate (MTX) in patients who are partial responders to DMARD monotherapy[2]. Some trials of combination DMARD therapies from first presentation have shown short-term benefits[3], but longterm results are still awaited. The new biologic agents, for example, tumor necrosis factor-α (TNF-α) inhibitors, have greater efficacy in the short term[4,5], but also potential longterm side effects, making them considerably more expensive. The use of TNF-α drugs depends on agreed criteria, which in the UK are failure of at least 2 DMARD (including MTX) and high disease activity measured using the Disease Activity Score 28 (DAS28)[6,7].

Establishing the prognosis of a patient with early RA is an important part of the management of the condition[8]. This has become increasingly vital with the advent of TNF-α inhibitors. These effective yet expensive treatments should be targeted towards recipients who can benefit most. Several factors such as erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), RF, disease duration, and the modified Health Assessment Questionnaire (HAQ)[9] have been reported as

From the Early Rheumatoid Arthritis Study Group (ERAS), Department of Rheumatology, City Hospital, St. Albans, UK.

Supported by an ARC Academic Secretary Grant and a BUPA Foundation Research Grant.

N. Bansback, BSc, Senior Operational Research Analyst; A. Brennan, PhD, Director, Operational Research Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK; A. Young, MD, Consultant Rheumatologist; J. Dixey, MD, Consultant Rheumatologist, ERAS, Department of Rheumatology, City Hospital.

Address reprint requests to Dr. A. Young, Department of Rheumatology, City Hospital, St. Albans, AL3 5PN, UK. E-mail: eras@whht.nhs.uk

Accepted for publication March 29, 2006.

valuable predictors of more severe RA. Their strength and reliability depend to a certain extent on which outcome measures are used. Functional disability has always proved more difficult to predict than radiological damage[10]. Attempts have been made to construct prognostic algorithms[11-13], but these efforts have been hampered either by the small sample size of available datasets or lack of good validation techniques to provide indicators of generalizability. Large databases of patients with early RA are required to reliably establish the effects of different prognostic factors on medium to longterm outcomes.

We analyze data from the Early RA Study (ERAS), based on an inception cohort. Data have been collected annually on over 1500 patients, with 5-year and 10-year followup of around 1000 and 700 patients, respectively.

MATERIALS AND METHODS

Study population. ERAS was established in 1986 with the primary aims to record longterm outcomes and examine predictive features in patients receiving conventional therapy. All consecutive RA patients seen within 2 years of initial symptoms, and prior to second-line drug use, were followed yearly using standardized assessments in 9 centers in different regions of England. Although ERAS started before agreements on minimum core datasets for disease activity, outcome measures, and response criteria in RA, all components except patient and physician global assessments have been included. The entry criteria, annual assessments, and followup details have been described in previous reports from this group[14]. Only patients who completed 5 years of followup are included in the present analysis.

Outcome of interest. Pain, disability, and loss of independence are the main concerns of patients who develop RA. Functional outcome has been shown to be correlated with patient perceived quality of life, premature mortality, and direct costs[15-18]. The HAQ is the primary measure for functional outcome in RA[9,19]. We have used a HAQ score ≥ 1.5 at 5 years as a definition of moderate to severe disease[20]. Figure 1 illustrates this criterion and describes a clear distinction in functional progression.

Prognostic factors. Previous reports based on inception cohort studies reviewed by Young and van der Heijde[10] have shown that functional measures at baseline (e.g., HAQ) are consistently associated with longterm disability, but in the main lack clinically useful predictive power. HAQ at early stages and over time is also associated with increased mortality[21]. ERAS has reported that HAQ stabilizes by the first year of treatment[14] and that the predictive power of HAQ, together with other clinical and laboratory features, only achieves clinical utility if variables at first-year followup are also included[22-24]. We have therefore used baseline and first-year disease variables in the model.

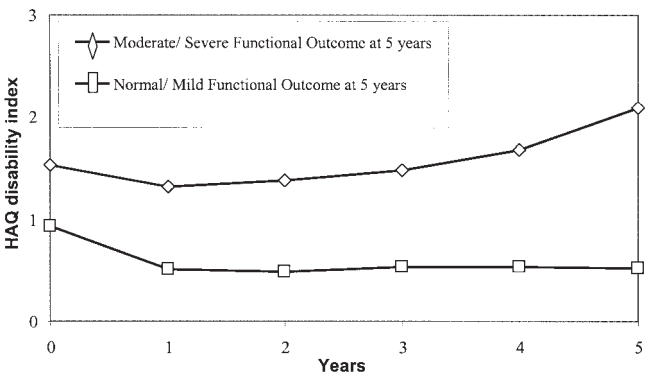

Figure 1. Trends in disability in the 2 identified groups.

The possibly relevant prognostic factors considered in the initial stage of the model-building process included gender, age at onset, months of RA symptoms prior to diagnosis, RF, presence of nodules, Carstairs deprivation index[25], American College of Rheumatology criteria[26], number of DMARD used in the first year, functional grades I–IV[27], morning stiffness (hours), swollen and tender joint counts, grip strength (0–300 mm), HAQ[9], pain score (VAS 0–100), hemoglobin, ESR, and DAS28[6]. We were particularly interested in the predictive power of radiographic changes on longterm functional outcome. Radiographs of hands and feet at baseline, 1, 2, 3, 5, 7, and 10 years[28] were scored using the Larsen damage and erosion scores[29].

Treatment profiles. The DMARD used were chosen according to physician's preference, employing standard practice of the late 1980s/1990s, which was sequential monotherapy, and "step up" combination therapies for more severe disease. Eight hundred one patients (81%) received at least one DMARD, started at a median 2 months (68% within 3 mo and 87% by 12 mo), consistent with the group's early treatment practice. The remaining patients (19%) were managed with nonsteroidal antiinflammatory drugs and/or low dose steroids. Overall, use of these drugs was sulfasalazine (54%), MTX (18%), intramuscular gold (13%), D-penicillamine (9%), antimalarials (4%), and various others (2%). Fifty-five percent of DMARD-treated patients required more than one drug. Steroids in doses of ≥ 7.5 mg daily for ≥ 12 months were used in 164 (17%) patients.

Statistical methods. The primary method of multivariate analysis was logistic regression. The model was formulated by systematically removing predictors from a full 35-predictor model. A more liberal criterion for selection was applied by including covariables with $p < 0.2$[30]. The final variable selection studied the internal validity of the modelling strategies by using bootstrap resampling[31].

We evaluated the predictive performance of the models by considering measures of discrimination and calibration. Discrimination refers to the ability to distinguish between high-risk and low-risk patients, and was quantified using the c-index and Nagelkerke R^2 [32]. The c-index or concordance probability is a generalization of the area under the receiver operating characteristic curve. While a $c = 0.5$ indicates random predictions, $c = 1$ indicates a perfectly discriminating model. Similarly $R^2 = 0$ indicates no predictive ability, and $R^2 = 1$ indicates perfect predictions. Calibration, or reliability, refers to the amount of agreement between predicted and observed outcomes. Calibration was quantified using estimates of slope shrinkage based on 200 bootstrap samples, where a value of 1 indicates perfect calibration[33]. All statistical analyses were carried out with S-PLUS 6.1 using the HMISC and Design software libraries[34].

RESULTS

Patient characteristics. Nine hundred eighty-five patients had a minimum of 5 years' followup. Patients were 17 to 93 years of age at disease onset (median 55 yrs) and 654 were female (66%). At baseline, 730 (74%) were seropositive and 241 (24%) patients had already developed erosions, similar to other early RA cohorts. The majority were in functional grade I or II (907, 92%). Over the first year of treatment, the median HAQ score at baseline for all patients decreased from 1.00 to 0.63. In the mild group (687, 70%), the mean change in HAQ was from 0.92 to 0.55, compared to 1.51 to 1.41 in the moderate/severe group (298, 30%). Patients in the latter group were on average 3 years older (57 vs 54). Tables 1 and 2 compare the demographic details and potential prognostic factors of the mild and moderate/severe functional groups, at baseline and at 1 year. These tables include the frequency of missing values for each prognostic factor. Patient records with missing values are omitted from the analysis.

Table 1. Potential prognostic factors at baseline in patients with none/mild and moderate/severe functional impairment at 5 years. Values are n (%) or median (quartiles).

Prognostic Variable	Mild Group, n = 687 (70%)	Missing	Severe Group, n = 298 (30%)	Missing
ACR criteria				
< 4	242 (35.2)	0	64 (21.5)	0
≥ 4	445 (64.8)		234 (78.5)	
Age at onset, yrs	54 (43–63)	0	57 (47–67)	0
Carstairs Index				
1	143 (20.8)	6 (0.9)	52 (17.4)	5 (1.7)
2	172 (25.0)		58 (19.5)	
3	142 (20.7)		78 (26.2)	
4	119 (17.3)		57 (19.1)	
5+	105 (15.3)		48 (16.1)	
DAS28	5.4 (3.5–7.9)	4 (0.6)	6.2 (3.6–8.1)	3 (1)
ESR	35 (18–59)	3 (0.4)	40 (21–65)	2 (0.7)
Functional grade				
I	279 (40.6)	0	53 (17.8)	0
II	377 (54.9)		198 (66.4)	
III + IV	30 (4.5)		47 (15.8)	
Male	259 (38.7)	0	97 (32.6)	0
Female	428 (62.3)		201 (67.4)	
Grip strength, mm	155 (115–230)	1 (0.1)	110 (75–155)	1 (0.3)
Hemoglobin	12.7 (11.6–13.6)	0	12.5 (11.4–13.7)	0
HAQ index	0.88 (0.4–1.4)	1 (0.1)	1.50 (1–2)	0
Joint score				
Swollen	9 (4–16)	1 (0.1)	15 (8–24)	1 (0.3)
Tender	12 (6–23)	3 (0.4)	18 (10–31)	0
Morning stiffness, h	1 (1–2)	0	2 (1–3)	0
Nodules				
Absent	621 (90.4)	0	268 (89.9)	0
Present	66 (9.6)		30 (10.1)	
Pain score, VAS	40 (20–59)	0	50 (33–74)	0
Radiography				
No erosions	514 (74.8)	9 (1.3)	215 (72.1)	6 (2)
Erosions	164 (23.9)		77 (25.8)	
RA symptoms, mo	6 (4–11)	0	6 (4–12)	0
Rheumatoid factor				
Negative	183 (26.6)	6 (0.9)	65 (21.8)	1 (0.3)
+	81 (11.8)		29 (9.7)	
++	160 (23.3)		79 (26.5)	
+++	257 (37.4)		124 (41.6)	

Larsen damage and erosion scores: median < 1 in both groups. ACR: American College of Rheumatology, DAS: Disease Activity Score, ESR: erythrocyte sedimentation rate, HAQ: Health Assessment Questionnaire, VAS: visual analog scale.

Patients not in the main analysis. Two hundred eighty patients were not included in this analysis because they dropped out of the study before the 5-year followup, for the following reasons: deceased 184 (66%), moved 20 (7%), declined 26 (9%), not known 42 (15%), and 8 (3%) defaulted at 5 years because of inpatient treatment for coexistent medical conditions or surgery, but remained in the study. A comparison was made with this group and the main cohort by taking the last HAQ as outcome (median followup time 3 yrs). In this group, 107 (38%) had a last HAQ ≥ 1.5, indicating worse functional outcome. The clinical features at baseline of this group were similar to the main cohort except for older age (median 59 and 67 yrs in mild and severe outcome groups, respectively) and more had erosive radiographs (121, 43%). The more severe outcome group had worse baseline function, as shown by 32 (30%) in functional group III and median HAQ 1.9. The deceased patients mainly accounted for these differences, which was not surprising given the known association between worse measures of RA and mortality[17].

Multivariate analysis. In the bootstrap resampling validation of the logistic regression, 6 variables consistently predicted moderate/severe functional outcome. Of these, patients with a functional grade of III or IV at 1 year [odds ratio (OR) = 6.7] and HAQ at 1 year (OR 2.4) were most important. Other vari-

Table 2. Potential prognostic factors at 1 year. Values are n (%) or median (quartiles).

Prognostic Variable	Mild Group, n = 687 (70%)	Missing	Severe Group, n = 298 (30%)	Missing
ACR criteria				
< 4	118 (17.2)	1 (0.1)	24 (8.1)	0
≥ 4	568 (82.8)		274 (91.9)	
DAS28	3.9 (3.1–5.7)	22 (3.2)	5.1 (3.5–6.9)	13 (4.4)
ESR	17 (8–38)	21 (3.1)	29 (16–55)	11 (3.7)
Functional grade				
I	409 (59.5)	23 (3.3)	50 (16.8)	12 (4)
II	249 (36.2)		191 (64.1)	
III + IV	6 (0.9)		45 (15.1)	
Grip strength, mm	210 (140–287)	19 (2.8)	135 (95–185)	8 (2.7)
Hemoglobin	129 (73–181)	19 (2.8)	125 (76–187)	8 (2.7)
HAQ index	0.5 (0–0.9)	19 (2.8)	1.5 (0.8–2)	8 (2.7)
Joint count				
Swollen	4 (0–9)	19 (2.8)	10 (4–17)	10 (3.4)
Tender	6 (2–13)	21 (3.1)	14 (6–24)	10 (3.4)
Morning stiffness, h	< 1 (0.1)	19 (2.8)	1 (0–2)	9 (3)
No. of DMARD				
0	211 (30.7)	0	51 (17.1)	0
1	396 (57.6)		174 (58.4)	
2+	80 (11.7)		73 (24.5)	
Pain score, VAS	20 (3–39)	23 (3.5)	42 (20–60)	11 (3.7)

Larsen damage and erosion scores: median < 2 in both groups. DMARD: disease modifying antirheumatic drugs. For other definitions see Table 1.

ables included the Carstairs deprivation index, hemoglobin, and Larsen score at baseline and DAS28 at Year 1. Table 3 shows the results of the logistic regression.

A slope shrinkage of 0.89 close to 1 indicates the model demonstrates only minor over-fitting. Estimates of the regression coefficients and performance can be corrected by multiplication with the shrinkage factor (Table 3). Reasonably large values for the c-index (0.82) and the Nagelkerke R^2 (0.39) indicate that the set of prognostic factors is explaining the variation in outcome reasonably well, and this implies good prediction for individual patients.

Application of the model. Our final prognostic model may be used to calculate expected probabilities of longterm moderate/severe functional outcome. Figure 2 is a nomogram[34] that enables a clinician to calculate the probability of a patient's prognosis. For each level of a prognostic factor, a number of points are allocated, and the total number of points from all prognostic factors can be converted into a probability of normal/mild or moderate/severe functional outcome.

It is normal clinical practice in the UK for early RA patients who do not respond to nonsteroidal antiinflammatory drugs prescribed in primary care to receive DMARD (e.g., MTX or sulfasalazine) early, after referral to secondary and specialist care, and then to switch drugs or use combination therapies when there is toxicity or lack of efficacy. The calculated probability of a patient having moderate/severe functional outcome according to the prediction rule could be used to select patients earlier for more aggressive drug therapy, for example, TNF inhibitor therapy.

In practice, there would need to be agreement on a certain threshold of risk from the algorithm in order to define a patient with predicted moderate/severe functional outcome. Table 4 shows the results of using different cutoff levels for different risk thresholds. As the threshold increases, the number of patients correctly predicted positive decreases (sensitivity); however, the number of patients incorrectly predicted positive decreases (specificity). Picking the correct risk threshold depends on the tradeoff between sensitivity and specificity. For example, in the first row in Table 4 the scenario is one of giving aggressive therapy to every patient (threshold, 100%). This would therefore identify all patients with actual moderate/severe functional outcome (sensitivity, 100%), but would mean a large proportion of patients without moderate/severe functional outcome would be given unwarranted aggressive therapy. If aggressive therapy is given only to patients whose predicted probability of moderate/severe functional outcome is, for example, above the threshold of 20%, the number of patients given aggressive therapy will be reduced to 65%. However, one of every 10 patients who would develop severe functional disease will be missed (sensitivity, 90%). With increasing cutoff levels, the number of patients given aggressive therapy is increasingly reduced. As a consequence the number of patients that would develop moderate/severe functional outcome but not given aggressive therapy would increase.

The last column in Table 4 describes the percentage of patients who would be selected for aggressive therapy using the algorithm, and who would actually be eligible for TNF

Table 3. Multivariate prognostic model.

Prognostic Variable	Coefficient	SE	Odds Ratio	p	Shrunk Coefficient
Carstairs Deprivation Index					
1	0		1.000		
2	-0.247	0.318	0.781	0.437	-0.225
3	0.363	0.310	1.438	0.243	0.330
4	0.548	0.323	1.730	0.080	0.499
5	0.685	0.339	1.984	0.044	0.624
Functional grade at baseline					
I	0		1.000		
II	0.592	0.272	1.808	0.030	0.538
III + IV	0.568	0.498	1.765	0.254	0.518
Functional grade at Year 1					
I	0		1.000		
II	0.689	0.251	1.992	0.006	0.627
III + IV	1.905	0.829	6.719	0.022	1.736
HAQ at baseline					
Baseline	0.532	0.189	1.702	0.005	0.485
At Year 1	0.894	0.215	2.445	0.000	0.814
Hemoglobin level at baseline	0.015	0.007	1.015	0.025	0.014
DAS28 at Year 1	0.138	0.075	1.148	0.064	0.126
Larsen score at baseline	0.013	0.010	1.013	0.200	0.011
Intercept	-5.647	1.029	0.004	0.000	-5.183

Concordance probability (area under the ROC curve = 0.82), Nagelkerke $R^2 = 0.39$. The multivariate logistic regression model can be written as: Predicted probability of severe disease = $1/1 + e^{-(LP)}$

Where linear predictor LP

$$= -5.647 + \text{Carstairs}\begin{pmatrix}1 & 0\\ 2 & -0.247\\ 3 & 0.363\\ 4 & 0.548\\ 5 & 0.685\end{pmatrix} + \text{FGYr0}\begin{pmatrix}\text{I} & 0\\ \text{II} & 0.592\\ \text{III/IV} & 0.568\end{pmatrix} + \text{FGYr1}\begin{pmatrix}\text{I} & 0\\ \text{II} & 0.689\\ \text{III/IV} & 1.905\end{pmatrix} + (\text{HAQYr0} \times 0.532) +$$

$$(\text{HAQYr1} \times 0.894) + (\text{HB} \times 0.015) + (\text{DAS28Yr1} \times 0.138) + (\text{LarsenYr0} \times 0.013)$$

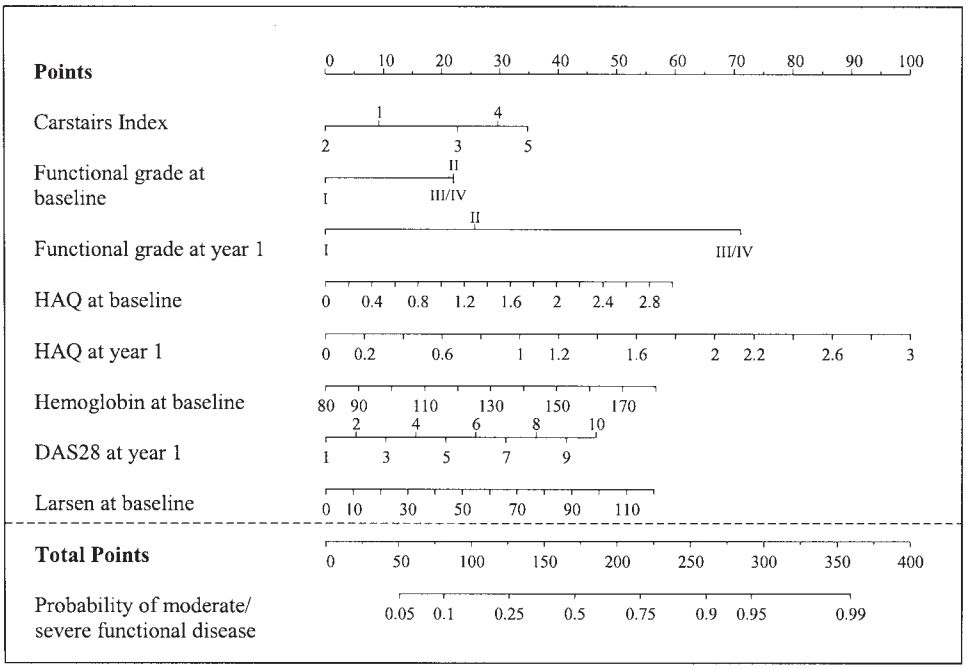

Figure 2. Nomogram for prediction of risk of moderate/severe functional outcome. For example, consider a patient with a Carstairs deprivation index = 3 (~23 points), functional grade at baseline = III, and at Year 1 = II (~22, and ~25 points), HAQ at baseline = 1.8 and at Year 1 = 1.7 (~36, and ~57 points), hemoglobin at baseline = 110 (~17 points), DAS28 at Year 1 = 4 (~15 points), and Larsen at baseline = 10 (~5 points). Total points equal ≈ 200, which equates to a risk ~ 0.65 or 65%.

Table 4. Implications of using the prediction rule in clinical practice. Values are percentages.

Predicted probability (risk threshold) at which aggressive treatment is given	Sensitivity*	Specificity**	Patients to be treated aggressively using algorithm	Of patients identified at risk by algorithm, % that meet current BSR criterion for treatment with TNF inhibitors +
≥ 0	100	0	100	7
≥ 10	97	25	83	8
≥ 20	90	51	65	11
≥ 30	82	67	51	13
≥ 40	73	79	40	16
≥ 50	63	88	31	20
≥ 60	51	94	23	20
≥ 70	42	96	18	23
≥ 80	29	99	12	27
≥ 90	12	100	5	34

* Probability that a patient with moderate/severe functional outcome will be identified positive. ** Probability that a patient with good functional outcome will be identified negative. + British Society of Rheumatology (BSR) guidelines: trial of minimum of 2 disease modifying antirheumatic drugs, one of which is methotrexate, and a Disease Activity Score28 > 5.1 before a tumor necrosis factor (TNF) inhibitor can be used.

inhibitors using the current British Society of Rheumatology guidance[2]. This demonstrates that at a cutoff of 0.6, of the 23% of patients selected for aggressive therapy using the algorithm, only 20% (5% of the sample) would actually be able to start treatment. Even at a cutoff of 0.9, where 5% of patients are selected, only 34% of these patients (1.7% of the sample) would be eligible.

DISCUSSION

The results of our statistical analysis have provided a multivariate tool to predict longterm functional outcome in patients with early RA. Most of the variables in the final equation have previously been shown to have some predictive value. Functional measures at baseline have been shown to consistently predict eventual functional outcome, although HAQ is not robust enough to be used routinely in clinical settings[10]. Education level and socioeconomic status have also been reported as important factors in functional outcome[22,35,36], whereas clinical and laboratory disease activity measures have variable importance[10], and radiological change appears less strongly related[37]. Hemoglobin does not normally perform well as a predictor of outcome compared to ESR and CRP, but we have reported its power in predicting large-joint damage, measured by need for joint replacement surgery[23].

The Carstairs index score was a strong prognostic factor and is a relatively easy item to collect in the UK, as it is based on postal code. The Carstairs index score is a measure of deprivation based on the UK 1991 Census, and is derived from the number of persons per household, rate of male unemployment, social class, and number of overcrowded households. There are criticisms of the index: that the choice of indicators is rather arbitrary, as is the assignment of equal weight to them in the overall deprivation index. Also, the use of car ownership in the index has been questioned on the grounds that this may be a useful indicator of deprivation in urban areas, but is less suitable in rural areas, where a car is seen as essential given the limited availability of public transport. When the Carstairs index is removed from the final predictive model, the c-index decreases from 0.82 to 0.80 and the Nagelkerke R^2 decreases from 0.39 to 0.38. This represents only a minor detriment in predictive ability of the model, and indicates the nomogram is valid without using the Carstairs index.

Most observational studies have reported initial improvement in HAQ, then gradual deterioration with time in patients receiving standard therapies, the subject of a recent review[37]. A relationship between function and radiological change has been described in other studies, but generally in late disease[38], and establishing a definite relationship between function and radiological damage has been a controversial area. We report here an association between baseline Larsen score and 5-year HAQ. In a recent study of radiological progression, the ERAS group reported the reverse association, that is, a weak but definite correlation between first-year HAQ and severity of erosions at 3 years[24]. Clinical trials have demonstrated a reduction in the progression of radiographic damage with TNF-α inhibition compared to MTX alone, but less improvement in functional outcome[39]. Despite this, our findings could be used to justify the use of TNF-α agents in early disease.

Strengths of our study include the inception cohort design, standard assessments in typical clinical settings, and careful followup of large numbers of patients in the long term. Possible concerns include left and right censorship and selection of patients. Because ours is a hospital clinic-based study, milder and early remitting RA may be underrepresented. Conversely, patients who were not included in the analysis because they died before 5-year followup had worse function at baseline and at followup, a not unexpected finding[17]. The risk factors for functional outcome remained the same for this

group. We conclude that our findings reflect hospital clinic-based populations and have to be interpreted accordingly.

There is an abundance of published prognostic models in the literature, but very few are routinely used in clinical practice[40]. One of the problems in the acceptance of the models is in their construction, which predominantly relies on small datasets, due mainly to difficulties maintaining longterm studies, and problems with missing data. These factors, in conjunction with other variations in study design, may explain the different results reported. Small datasets with too many variables are likely to overestimate the prognostic value of the data.

Statistical tests should be applied to account for these inadequacies[41,42], especially when attempting to account for the overestimation of prognostic information in a model that is fitted and evaluated on the same dataset. This can be achieved by various methods of internal validation. Common ways of establishing how well a model might perform are data-splitting or cross-validation; however, these are not often performed. The most sophisticated and accurate method is the use of bootstrapping, as described in our analysis, to estimate shrinkage factors to regression coefficients to counter overoptimism[31].

Producing relatively simple diagrams such as the nomogram presented in Figure 2 provides a realistic attempt to make statistical analyses relevant to clinicians. There are a number of further issues to discuss before such an algorithm might find use in clinical practice. While the algorithm is statistically validated, it has not been clinically validated. This requires the algorithm to be tested with an external dataset to see if it performs satisfactorily. A potential problem is whether the variables used in the prognostic equation are collected in other clinics. A number of the variables, including the HAQ, Larsen and Carstairs index, and functional grade, were collected specifically for the ERAS specifications. Although all are standard and validated measures, they are not performed routinely outside observational studies and clinical trials. It could be argued that their collection can be time-consuming and costly. To become clinically useful, however, it would require clinicians to start collecting such data.

The natural successor to ERAS was initiated in the UK in 2002 and nearly 30 centers are now collecting these items as part of the Early RA Network[43]. Further work is also planned to assess the additional prognostic power of specific variables other than standard clinical measures, for example, genetic data.

A validated model for prediction of severe functional outcome could allow clinicians to target aggressive therapy earlier in a patient's disease course. Two questions remain: How long after diagnosis is early enough, and what would aggressive therapy be? Our prognostic model relies on at least one year of followup before applying the decision rule. While it could be argued that this is too late, it would certainly expedite earlier use of TNF inhibitors in countries where patients now have to fulfil criteria for biologic therapy, including failing a number of DMARD over at least 6 months[7,44]. We also show that current guidelines may not be the most suitable measure for discrimination for aggressive therapies. Of course the question whether aggressive therapy including TNF inhibitors alters the course of function in identified patients remains to be fully determined. External validation of our prognostic index is required clinical practice in the UK and other countries.

ACKNOWLEDGMENT

All members of ERAS have consented to publication of this article. ERAS clinicians and metrologists, respectively, listed by center: Chelmsford: Dr. P. Davies and L. Hill; Birmingham: Dr. A. Gough, Dr. J. Devlin, Prof. P. Emery, and L. Waterhouse; Basingstoke: Dr. P. Prouse and C. Boys, A. Kent; Grimsby: Dr. D. James and H. Tait; Medway: Dr. P. Williams and D. White; Oswestry: Dr. J. Dixey and H. Dart; St. Albans: Dr. A. Young and A. Seymour; Sheffield: Dr. J. Winfield; Winchester: Dr. N. Cox and S. Stafford. Statistics and Health Economics: N. Banstock, A. Brennan, Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield S1 4DA, UK.
ERAS Coordinator: Mrs. Cathy Mayes, ARC Academic Secretary, Rheumatology Department, City Hospital, St. Albans, Herts, UK, AL3 5PN. E-mail: eras@whht.nhs.uk

REFERENCES

1. Felson D, Anderson J, Meenan R. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses. Arthritis Rheum 1990;33:1449-61.
2. Choi EH. Two is better than one? Combination therapy in rheumatoid arthritis [editorial]. Rheumatology Oxford 2004;43:1205-7.
3. Boers M, Verhoeven AC, Markusse HM, et al. Randomized comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1997;350:309-18.
4. Maini RN, Breedveld FC, Kalden JR, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumour necrosis factor alpha monoclonal antibody combined with low dose weekly methotrexate in rheumatoid arthritis. Arthritis Rheum 1998;41:1551-63.
5. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. Ann Intern Med 1999;130:478-86.
6. Provoo MLL, van't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight joint scores. Arthritis Rheum 1995;38:44-8.
7. Guidelines for prescribing TNF alpha blockers in adults with rheumatoid arthritis. Working party of the British Society of Rheumatology; 2001. Available from: www.rheumatology.org.uk/guidelines/clinicalguidelines. Accessed April 26, 2006.
8. Scott DL. The diagnosis and prognosis of early arthritis: rationale for new prognostic criteria [editorial]. Arthritis Rheum 2002;46:286–90.
9. Fries JF, Spitz P, Kraines G, Holman H. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
10. Young A, van der Heijde DFM. Can we predict aggressive disease? Ballieres Clin Rheumatol 1997;11:27-48.
11. Brennan P, Harrison B, Barrett E, et al. A simple algorithm to predict the development of radiological erosions in patients with early rheumatoid arthritis: prospective cohort study. BMJ 1996;313:471-6.
12. Drossaers-Bakker KW, Zwinderman AH, Vlieland TP, et al. Long-term outcome in rheumatoid arthritis: a simple algorithm of

baseline parameters can predict radiographic damage, disability, and disease course at 12-year followup. Arthritis Rheum 2002;47:383-90.
13. Visser H, le Cessie S, Vos K, Breedveld FC, Hazes JM. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46:357–65.
14. Young A, Dixey J, Cox N, et al. How does functional disability in early rheumatoid arthritis (RA) affect patients and their lives? Results of 5 yr follow up in 781 patients from the Early RA Study (ERAS). Rheumatology Oxford 2000;39:603–11.
15. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatoid disorders using the EuroQol. Rheumatology Oxford 1997;36:786-93.
16. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Rheumatology Oxford 1997;36:551-9.
17. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1530-42.
18. Michaud K, Messer J, Choi HK, Wolfe F. Direct medical costs and their predictors in persons with rheumatoid arthritis: a 3-year study of 7,527 patients. Arthritis Rheum 2003;48:2750-63.
19. Ramey DR, Fries JF, Singh G, Spilker B. Quality of life and pharmacoeconomics in clinical trials. In: Spilker B, editor. The Health Assessment Questionnaire 1995 — status and review. Philadelphia: Lippincott-Raven; 1996:227-37.
20. Wiles N, Dunn G, Barrett E, Silman A, Symmons D. Associations between demographic and disease-related variables and disability over the first five years of inflammatory polyarthritis. J Clin Epidemiol 2000;53:988-96.
21. Yelin E, Trupin L, Wong B, Rush S. The impact of functional status and change in functional status on mortality over 18 years among persons with RA. J Rheumatol 2002;29:1851-7.
22. Young A, Wilkinson P, Talamo J, et al. Socio-economic factors in the presentation and outcome of early rheumatoid arthritis. Lessons for the health service? Ann Rheum Dis 2000;59:794-9.
23. James D, Young A, Dixey J, et al. Orthopaedic intervention in early rheumatoid arthritis. Occurrence and predictive factors in an inception cohort of 1029 patients followed for 5 years. Rheumatology Oxford 2004;43:369-76.
24. Dixey J, Solymossy C, Young A. Is it possible to predict radiological damage in early RA? J Rheumatol 2004;31 Suppl 69:48-54.
25. Carstairs V. Deprivation indices: their interpretation and use in relation to health. J Epidemiol Community Health 1995;49:s3-8.
26. Arnett FC, Edworthy SM, Bloch DA, et al. The ARA 1987 revised criteria for classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-2.
27. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. J Am Med Assoc 1949;140:659-62.
28. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn 1977;18:481–91.
29. Solymossy C, Dixey J, Utley M, et al. Larsen scoring of digitized X-ray images. Rheumatology Oxford 1999;38:1127–9
30. Tabachnick BG, Fidell LS. Using multivariate statistics. 2nd ed. London; New York: Harper Collins; 1989.
31. Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics. Multivariate prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing bias. Stat Med 1996;15:361-87.
32. Harrell FE. Regression modelling strategies. New York: Springer-Verlag; 2001.
33. Harrell FE. S-Plus and R libraries, functions, and documentation. Available from: http://lib.stat.cmu.edu/S/Harrell/splus.html. Accessed May 5, 2006.
34. Lubsen J, Pool J, van der Does E. A practical device for the application of a diagnostic or prognostic function. Methods Inf Med 1978;17:127-9.
35. Callahan LF, Pincus T. Formal education level as a significant marker of clinical status in rheumatoid arthritis. Arthritis Rheum 1988;31:1346-57.
36. McEntegart A, Morrison E, Capell HA, et al. Effect of social deprivation on disease severity and outcome in patients with rheumatoid arthritis. Ann Rheum Dis 1997;56:410-3.
37. Young A. Early rheumatoid arthritis. Rheum Dis Clin North Am 2005;31:659-79.
38. Scott DL, Pugner K, Kaarela K, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology Oxford 2000;39:122-32.
39. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 2004;363:675-81.
40. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539-41.
41. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;9:453-73.
42. Visser H, Hazes JM. Prognostics. Best Pract Res Clin Rheumatol 2003;17:403-14.
43. Garwood W. The early RA network. Musculoskeletal Care 2004;2:240-4.
44. Furst DE, Breedveld FC, Kalden JR, et al. Updated consensus statement on biological agents for the treatment of rheumatoid arthritis and other immune mediated inflammatory diseases. Ann Rheum Dis 2003;62 Suppl II:ii2-9.