# Patient-reported Outcomes as Predictors of Change in Disease Activity and Disability in Early Rheumatoid Arthritis: Results from the Yorkshire Early Arthritis Register

Sarah Twigg, Elizabeth M.A. Hensor, Paul Emery, Alan Tennant, Ann W. Morgan, and the Yorkshire Early Arthritis Register Consortium

**ABSTRACT.** *Objective.* To assess patient-reported variables as predictors of change in disease activity and disability in early rheumatoid arthritis (RA).

*Methods.* Cases were recruited to the Yorkshire Early Arthritis Register (YEAR) between 1997 and 2009 (n = 1415). Predictors of the 28-joint Disease Activity Score (DAS28) and the Health Assessment Questionnaire-Disability Index (HAQ-DI) at baseline and change over 12 months were identified using multilevel models. Baseline predictors were sex, age, symptom duration, autoantibody status, pain and fatigue visual analog scales (VAS), duration of early morning stiffness (EMS), DAS28, and HAQ-DI.

*Results.* Rates of change were slower in women than men: DAS28 fell by 0.19 and 0.17 units/month, and HAQ-DI by 0.028 and 0.023 units/month in men and women, respectively. Baseline pain and EMS had small effects on rates of change, whereas fatigue VAS was only associated with DAS28 and HAQ-DI at baseline. In patients recruited up to 2002, DAS28 reduced more quickly in those with greater pain at baseline (by 0.01 units/mo of DAS28 per cm pain VAS, p = 0.024); in patients recruited after 2002, the effect for pain was stronger (by 0.01 units/mo, p = 0.087). DAS28 reduction was greater with longer EMS. In both cohorts, fall in HAQ-DI (p = 0.006) was greater in patients with longer EMS duration, but pain and fatigue were not significant predictors of change in HAQ-DI.

*Conclusion.* Patient-reported fatigue, pain, and stiffness at baseline are of limited value for the prediction of RA change in disease activity (DAS28) and activity limitation (HAQ-DI). (First Release July 1 2017; J Rheumatol 2017;44:1331–40; doi:10.3899/jrheum.161214)

*Key Indexing Terms:*
RHEUMATOID ARTHRITIS      COHORT STUDY      OUTCOMES      PAIN

The use of patient-reported outcomes (PRO) to assess treatment response in rheumatoid arthritis (RA) is well established. The core set of outcomes recommended for assessment of RA treatment by the Outcome Measures in Rheumatology Clinical Trials (OMERACT) group includes patient-reported variables such as pain and fatigue[1,2]. Measurement of these subjective indicators of health status can aid clinical assessment[3] and there is evidence that they can be useful to help predict RA remission. For example, a study involving 103 patients with RA from Japan found an

*S. Twigg, MD, Clinical Lecturer, LIRMM, School of Medicine, University of Leeds, and NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust;*

*E.M. Hensor, PhD, Biostatistician, LIRMM, School of Medicine, University of Leeds, and NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust; P. Emery, PhD, Arthritis Research UK Professor of Rheumatology, LIRMM, School of Medicine, University of Leeds, and NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust; A. Tennant, PhD, Senior Advisor, LIRMM, School of Medicine, University of Leeds, and Swiss Paraplegic Research; A.W. Morgan, PhD, Professor of Molecular Rheumatology/Hon. Consultant Rheumatologist, LIRMM, School of Medicine, University of Leeds, and NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust.*

*Address correspondence to Dr. A.W. Morgan, Professor of Molecular Rheumatology/Hon. Consultant Rheumatologist, Leeds Institute Rheumatic and Musculoskeletal Medicine, Wellcome Trust Brenner Building, St. James's University Hospital, Leeds LS9 7TF, UK.*
*E-mail: a.w.morgan@leeds.ac.uk*

*Accepted for publication April 26, 2017.*

inverse association between remission and greater pain and fatigue at baseline after a 7-year followup[4]. Similarly, greater baseline pain was associated with reduced odds of remission at 6 and/or 12 months in the French early inflammatory arthritis ESPOIR cohort[5]. Thus, patient-reported measures may be rapid and cost-effective tools for the prediction of outcome in RA. However, before these variables can be useful in a clinical setting, further evidence to support their application is needed. Our present study evaluated patient-reported measures [fatigue, pain, and early morning stiffness (EMS)] alongside traditional predictors of outcome to investigate their value in predicting the rate of change in disease activity and disability in an early RA cohort.

## MATERIALS AND METHODS

*Subjects.* The Yorkshire Early Arthritis Register (YEAR) is an observational inception cohort whose subjects were aged over 18 years with a consultant-made diagnosis of recent-onset RA. Our present study used data from 1415 participants recruited to the YEAR between 1997 and 2009 with inflammatory symptom durations of ≤ 24 months. Details of the YEAR were published previously[6]. Briefly, data on patients with RA were collated from 14 rheumatology outpatient centers across Yorkshire, UK. Participants were treated according to a regionally agreed protocol that recommended sequential escalation of treatment with disease-modifying antirheumatic drugs (DMARD). When data collection began in 1997, the first-line DMARD was sulfasalazine, but this changed to methotrexate (MTX), with a one-off dose of intramuscular methylprednisolone (120 mg) given at baseline, when the data collection and treatment protocols were altered in 2002. Deviations from the treatment protocol were made at the discretion of the treating rheumatologist. For our present analysis, patient data were not included if the symptom duration exceeded 24 months, or was missing. All patients provided written consent for inclusion into the study and ethical approval was granted by the Northern and Yorkshire Research Ethics Committee (MREC /99/3/48).

*Data collection.* Data were collected at baseline, 3, 6, 9 (after 2002), and 12 months by a clinician or research nurse. Identified details included sex, date of birth, date of symptom onset, swollen and tender joint counts from a score of 28 (SJC and TJC), and duration of EMS in minutes. Participants completed self-assessment tools, which included visual analog scales (VAS) to indicate their assessment of pain (0–100 mm scale, where 0 = no pain and 100 = pain as bad as it can be) and fatigue (0 = no abnormal fatigue and 100 = fatigue as bad as it can be). The disability index component of the Health Assessment Questionnaire (HAQ) was completed at each visit and is referred to as the "HAQ-DI" from here onward. The SJC and TJC of 28 joints and C-reactive protein (CRP) were used to calculate the 3-variable Disease Activity Score (DAS28-CRP)[7] for each visit. Laboratory analyses undertaken at individual recruitment centers included CRP at all visits and IgM rheumatoid factor (RF) at baseline. RF was measured using standard nephelometric assays, and anticitrullinated protein antibodies (ACPA) were determined retrospectively on stored samples using previously described methods[8].

*Data analysis.* Baseline demographic and disease characteristics were summarized in terms of means and SD (continuous variables) and percentages (categorical variables). Multilevel models (random intercepts, fixed slopes) were constructed to evaluate baseline predictors of DAS28 and HAQ-DI measured at baseline, 6 months, and 12 months. These were 2-level models in which repeated measurements over time (level 1) were nested within patients (level 2). These models included an indicator for "cohort" (before or after 2002, when the treatment protocol changed) and a variable indicating month, which was treated as a continuous covariate. Interactions were added between each predictor and cohort to show whether associations with baseline DAS28 or HAQ-DI differed by cohort, and between each

predictor and month, to show whether the predictor was associated with change in DAS28 over time. Additionally, 3-way interaction terms between each predictor, month, and cohort were added to analyze whether changes over time differed by cohort. The interaction terms were sequentially discarded in order of least significance until only significant terms (where p ≤ 0.1) remained in the model; 2-way interactions were retained irrespective of significance if both variables were included in a significant 3-way interaction. Linear change was assumed over time. Pseudo-adjusted $R^2$ was calculated as the adjusted $R^2$ between observed and predicted values of each outcome. $R^2$ estimates obtained in each imputed dataset were averaged after using Fisher's r-to-z transformation. We considered whether random slopes were more suitable than fixed slopes. Formally testing for random slopes using the standard likelihood ratio approach is not currently supported for multiply-imputed datasets in our chosen analysis package. We compared the coefficients between models that included fixed or random slopes for time and found them to be very similar, and the conclusions regarding which main effects and interactions were statistically significant remained unaffected; therefore, we opted to retain the simpler model.

Continuous rather than dichotomous outcomes (e.g., remission or non-remission, HAQ above or below a threshold value) were used to retain statistical power. To this end, because HAQ-DI represents an ordinal scale[9], this variable was transformed using Rasch analysis so that it could be analyzed as an interval-scaled variable[10]. As well as traditionally reported predictors of RA outcome — including sex, antibody status, and age — patient-reported pain, fatigue, and duration of EMS were also included as predictors in the models. Continuous variables were centered at the mean prior to analysis. EMS was not normally distributed and was therefore divided into 5 about equal-sized groups: < 30, 30–59, 60–119, 120–179, and ≥ 180 min. Correlation between RF and ACPA status was 0.56 and considered low enough for both variables to be included in the models simultaneously.

*Missing data.* Missing data were accounted for using multiple imputation (MI) by chained equations and 50 imputed datasets, the results from which were combined according to Rubin's rules[11]. Predictive mean matching with 10 nearest neighbors was used to impute continuous variables; for RF and ACPA, logistic regression was used. Fifty imputations were chosen for our analysis to achieve ≥ 95% relative efficiency of the MI estimates[11,12,13], given the amount of missing data (40% missing and 43% missing for the 6- and 12-month analyses, respectively). Auxiliary variables were selected from the dataset and included in the imputation models if they correlated with predictor or outcome variables (Pearson correlation ≥ 0.7), or predicted missingness (significant predictors in logistic regression analyses). The order of imputation (which included auxiliary variables) was TJC28; SJC28; CRP at baseline, 3, 6, and 12 months; HAQ at baseline, 6, and 12 months; and baseline pain VAS, fatigue VAS, EMS, age, sex, symptom duration, RF, and ACPA. Summary statistics of the imputed datasets were examined and compared with those of the complete dataset to check that imputed values were reasonable.

All analyses were conducted using Stata 13 (Stata Statistical Software: Release 14.1).

## RESULTS

*Baseline characteristics and missing data.* Numbers of cases recruited to the YEAR and included in the final analysis are shown in Figure 1. From a total of 1415 cases, 690 were recruited between 1997 and 2002 and 725 were recruited after 2002. Baseline characteristics and rates of missingness for variables included in the analysis are given in Table 1. The YEAR may be considered consistent with other early RA cohorts, with 66% of patients being female, an average age at onset of 58 years, and 71% RF-positive. These summary statistics were similar for cases recruited before and after
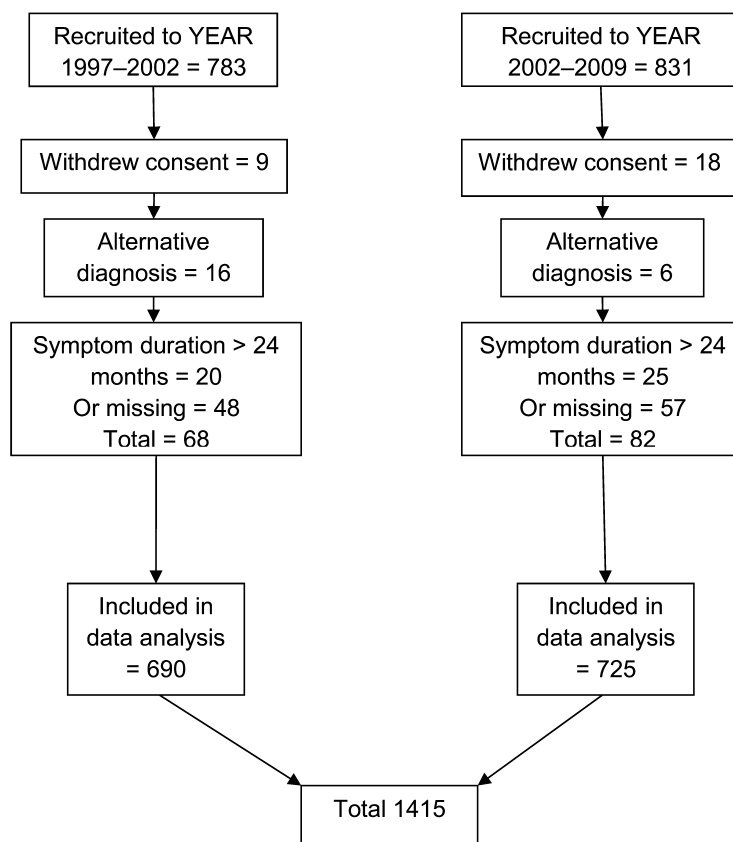
Figure 1. Cases recruited and included in analysis. YEAR: Yorkshire Early Arthritis Register.

2002; however, mean baseline DAS28 was lower for cases recruited after 2002 (4.8 compared with 5.4) and similar differences were seen in baseline HAQ-DI (1.18 compared with 1.28). Baseline pain and fatigue VAS were also slightly higher in the earlier cohort, with mean pain VAS 6.3 cm pre-2002 and 5.3 cm post-2002, and fatigue VAS 4.8 cm and 4.5 cm, respectively. In 21% of cases, some variables were missing at baseline. Cases with no missing data were slightly older (58.6 vs 57.7 yrs) and reported slightly more baseline fatigue with higher DAS28 and HAQ-DI values.

*Change in DAS28*. Table 2 gives the results of the multilevel model of change in DAS28. Baseline DAS28 was higher in patients recruited prior to 2002, older patients, and those with longer disease duration, greater pain, fatigue, and longer duration of EMS. On average, DAS28 reduced by 0.19 units/month in men and the rate of reduction was 0.02 units/month slower in women. Reduction in DAS28 per month was slightly faster in older patients (by 0.01 units per decade of baseline age). All of the statistically significant effects of baseline variables on change in DAS28 were small. At 12 months, the estimated differences between patients according to sex, age (80 yrs compared with 50 yrs), and cohort (for values of pain VAS ranging from 4 cm to 8 cm)

did not exceed 0.6 DAS28 units. Pseudo-adjusted $R^2$ for the DAS28 model was 0.30 (95% CI 0.28–0.32).

The association of baseline pain and stiffness with change in DAS28 differed depending on whether patients were recruited before or after 2002 (overall test of significance for stiffness was $p = 0.022$ and pain, $p = 0.087$). In both cohorts, greater pain VAS at baseline was associated with a slightly greater fall in DAS28 per month; this trend was stronger for patients recruited after 2002 (Figures 2A and 2B). In the earlier cohort, baseline EMS was not associated with rate of change in DAS28, but in the later cohort, longer duration of EMS was associated with greater reduction in DAS28 (Figures 2C and 2D).

Repeating the final model using only cases with complete data yielded similar results to those obtained through MI, although with lower power. The effect of symptom duration on change over time was reduced in the MI analysis compared with complete cases, whereas the interaction between baseline pain, cohort, and change over time was more apparent in the MI analysis.

*Change in HAQ-DI*. Table 3 shows the results of the multilevel model of HAQ-DI. As shown in Figure 3, higher baseline DAS28 and longer EMS duration were associated

Table 1. Cohort characteristics and rates of missing data from 1415 cases in the Yorkshire Early Arthritis Register.

| Categorical Data | n (%) | No. Cases with Missing Data, n (%) |
|---|---|---|
| Female | 932 (66) | 0 (0) |
| RF-positive | 952 (71) | 82 (6) |
| ACPA-positive | 553 (64) | 553 (39) |
| EMS duration, min | | 45 (3) |
| 0–29 | 271 (20) | |
| 30–59 | 172 (13) | |
| 60–119 | 346 (25) | |
| 120–179 | 236 (17) | |
| ≥ 180 | 345 (25) | |

| Continuous Data | Mean (SD) | No. Cases with Missing Data, n (%) |
|---|---|---|
| Age, yrs | 57.7 (14.2) | 0 (0) |
| Symptom duration, mos* | 7.1 (4.3) | 0 (0) |
| Pain VAS, cm | 6.1 (2.6) | 90 (6) |
| Fatigue VAS, cm | 5.3 (2.9) | 291 (21) |
| DAS28-CRP, baseline | 5.01 (1.33) | 157 (11) |
| HAQ-DI, baseline | 1.22 (0.57) | 72 (5) |
| DAS28-CRP, 6 mos | 3.64 (1.57) | 449 (32) |
| HAQ-DI, 6 mos | 0.97 (0.64) | 302 (21) |
| DAS28-CRP, 12 mos | 3.22 (1.48) | 413 (29) |
| HAQ-DI, 12 mos | 0.89 (0.63) | 344 (24) |

* Cases with missing values for symptom duration were excluded to focus the analysis on outcome in early RA. RF: rheumatoid factor; ACPA: anticitrullinated protein antibodies; EMS: early morning stiffness; VAS: visual analog scale; DAS28-CRP: Disease Activity Score in 28 joints and C-reactive protein; HAQ-DI: Health Assessment Questionnaire-Disability Index.

with slightly greater reduction in HAQ-DI and the effect of pain varied with cohort. Baseline HAQ-DI was higher in women than men by 0.217 units in cases recruited after 2002 and 0.091 units in cases recruited pre-2002, but the rate of change in HAQ-DI by sex was consistent between cohorts: average reduction was 0.028 units/month in men and 0.023 in women. As Figure 3C illustrates, reduction in HAQ-DI was between 0.006 and 0.012 units/month faster in patients with EMS ≥ 30 mins compared with < 30 mins (combined test of significance for all EMS categories p = 0.023), and was 0.004 units/month faster per unit of baseline DAS28. Baseline pain was not associated with reduction in HAQ-DI in patients recruited up to 2002 (0.001 HAQ units/cm), but there was a slightly stronger trend in the later cohort (0.003 HAQ units/cm). Pseudo-adjusted $R^2$ for the HAQ model was 0.24 (95% CI 0.22–0.26).

## DISCUSSION

Our study examined predictors of change in DAS28 and HAQ–DI in early RA, including patient-reported measures (pain, fatigue, and EMS) alongside traditional predictors of prognosis: sex, age, and antibody status. The rate of reduction in DAS28 was greater with increased age at baseline and slower in women than men. It was also faster in those with greater pain or EMS at baseline, especially for our patients recruited after 2002. However, effects attributable to statistically significant variables were small. The measurement error

of DAS is 0.6, and therefore a reduction from baseline of twice this (> 1.20) is considered a good response[14]. In comparison, our present analyses predicted a reduction in DAS28 of 0.05 units/month in cases recruited after 2002 with EMS duration of ≥ 180 compared with < 30 min (about 0.3 units after 6 mos and 0.6 units after 12 mos). Further, fall in DAS28 was only 0.02 units/month faster per cm of baseline pain VAS in the later cohort where the effect was strongest. The effects of predictor variables on decrease in HAQ-DI were also small: the rate of change in HAQ-DI was 0.005 units/month slower in women than men, and 0.004 units/month faster per unit of baseline DAS28. For those who reported ≥ 180 compared with < 30 min of baseline EMS, decrease in HAQ-DI was 0.012 units greater per month. In our present analyses, fatigue did not significantly affect the rate of change in HAQ-DI, and pain had only a limited effect, restricted to the later cohort. Pain, fatigue, and EMS as predictors of change in disease activity and disability are therefore unlikely to have direct clinical applications.

These findings are consistent with previously reported associations of patient-reported symptoms and other outcomes. Recent data from the ESPOIR cohort found only a moderate correlation of fatigue and pain VAS with simultaneous DAS28 measurement, among other PRO[15]. Female sex is frequently identified as an independent predictor of adverse outcome in RA, including non-remission[16] and lesser reduction in DAS28[17], and the results from our present study

*Table 2.* Results of multilevel regression models of change in DAS28. Age was entered into the model as (age in years)/10. For EMS, the referent category was < 30 min.

| Variables | Imputed Data, n = 1415 | | | Complete Case, n = 614 | | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | p | Coefficient | SE | p |
| Predictors of baseline DAS28 | | | | | | |
| Cohort, post-2002 vs pre-2002 | –0.65 | 0.18 | < 0.001 | –0.45 | 0.26 | 0.090 |
| Female | –0.06 | 0.08 | 0.435 | –0.07 | 0.11 | 0.510 |
| Age, yrs | 0.09 | 0.03 | < 0.001 | 0.09 | 0.04 | 0.026 |
| Symptom duration, mos | –0.02 | 0.01 | 0.019 | –0.03 | 0.01 | 0.033 |
| RF-positive | –0.04 | 0.10 | 0.718 | –0.13 | 0.14 | 0.342 |
| ACPA-positive | 0.03 | 0.12 | 0.808 | 0.15 | 0.13 | 0.271 |
| Pain VAS, cm, pre-2002 | 0.14 | 0.02 | < 0.001 | 0.13 | 0.03 | < 0.001 |
| Fatigue VAS, cm | 0.05 | 0.02 | 0.001 | 0.06 | 0.02 | 0.004 |
| EMS duration, pre-2002, min | | | | | | |
| 30–59 | 0.44 | 0.17 | 0.010 | 0.63 | 0.27 | 0.020 |
| 60–119 | 0.53 | 0.15 | < 0.001 | 0.44 | 0.22 | 0.048 |
| 120–179 | 0.72 | 0.18 | < 0.001 | 0.57 | 0.28 | 0.042 |
| ≥ 180 | 0.60 | 0.15 | < 0.001 | 0.53 | 0.25 | 0.035 |
| Cohort differences in predictors of baseline DAS28, post-2002 vs pre-2002 | | | | | | |
| Pain VAS, post-2002 vs pre-2002 | 0.02 | 0.03 | 0.555 | –0.01 | 0.04 | 0.805 |
| EMS duration, post-2002 vs pre-2002, min | | | | | | |
| 30–59 | –0.50 | 0.27 | 0.061 | –0.57 | 0.40 | 0.153 |
| 60–119 | –0.24 | 0.23 | 0.289 | –0.10 | 0.33 | 0.762 |
| 120–179 | –0.49 | 0.25 | 0.055 | –0.30 | 0.38 | 0.433 |
| ≥ 180 | –0.20 | 0.23 | 0.380 | 0.01 | 0.35 | 0.981 |
| Change in DAS28 per mo* | –0.19 | 0.02 | < 0.001 | –0.20 | 0.03 | < 0.001 |
| Predictors of change in DAS28, difference in change per mo | | | | | | |
| Cohort, post-2002 vs pre-2002 | 0.03 | 0.02 | 0.103 | 0.03 | 0.03 | 0.303 |
| Female | 0.02 | 0.01 | 0.016 | 0.04 | 0.01 | 0.005 |
| Age, yrs | –0.01 | 0.00 | 0.001 | –0.01 | 0.00 | 0.041 |
| Symptom duration, mos | 0.00 | 0.00 | 0.116 | 0.00 | 0.00 | 0.031 |
| RF-positive | 0.01 | 0.01 | 0.364 | 0.02 | 0.02 | 0.177 |
| ACPA-positive | 0.02 | 0.01 | 0.106 | 0.02 | 0.02 | 0.247 |
| Pain VAS, cm, pre-2002 | –0.01 | 0.00 | 0.024 | –0.01 | 0.00 | 0.040 |
| Fatigue VAS, cm | 0.00 | 0.00 | 0.224 | 0.00 | 0.00 | 0.479 |
| EMS duration, pre-2002, min | | | | | | |
| 30–59 | 0.02 | 0.02 | 0.489 | –0.01 | 0.03 | 0.711 |
| 60–119 | –0.02 | 0.02 | 0.395 | –0.02 | 0.03 | 0.415 |
| 120–179 | 0.00 | 0.02 | 0.992 | –0.02 | 0.03 | 0.536 |
| ≥ 180 | 0.00 | 0.02 | 0.978 | –0.01 | 0.03 | 0.651 |
| Cohort differences in predictors of change in DAS28 | | | | | | |
| Pain VAS, cm, post-2002 vs pre-2002 | –0.01 | 0.00 | 0.087 | 0.00 | 0.00 | 0.808 |
| EMS duration, post-2002 vs pre-2002, min | | | | | | |
| 30–59 | –0.06 | 0.03 | 0.054 | –0.04 | 0.04 | 0.348 |
| 60–119 | 0.00 | 0.02 | 0.926 | 0.03 | 0.04 | 0.428 |
| 120–179 | –0.04 | 0.03 | 0.145 | –0.03 | 0.04 | 0.431 |
| ≥ 180 | –0.05 | 0.03 | 0.051 | –0.06 | 0.04 | 0.097 |
| Model constant | 4.96 | 0.16 | < 0.001 | 4.95 | 0.24 | < 0.001 |

* In seronegative males with EMS < 30 min and mean age, symptom duration, and pain and fatigue VAS recruited pre-2002. DAS28: Disease Activity Score in 28 joints; EMS: early morning stiffness; RF: rheumatoid factor; ACPA: anticitrullinated protein antibodies; VAS: visual analog scale.

were consistent with this. Although some studies have found an association between increasing age at baseline and non-remission, this effect is not consistent between studies[16] and therefore our findings of only slightly faster reduction in DAS28 with increasing age at baseline were not surprising. There have also been several reported associations of increased age at RA onset with less favorable HAQ-DI[18,19,20], and although we did not find an association between rate of change in HAQ-DI and age, baseline HAQ-DI was higher for older patients.

Whether the findings of our study can be applied in the context of modern RA management is influenced by contemporary treatment approaches. Current treat-to-target recommendations for RA management were published in 2010[21], after recruitment to the YEAR ended. However, our findings may still be applicable for certain patients, for example, those
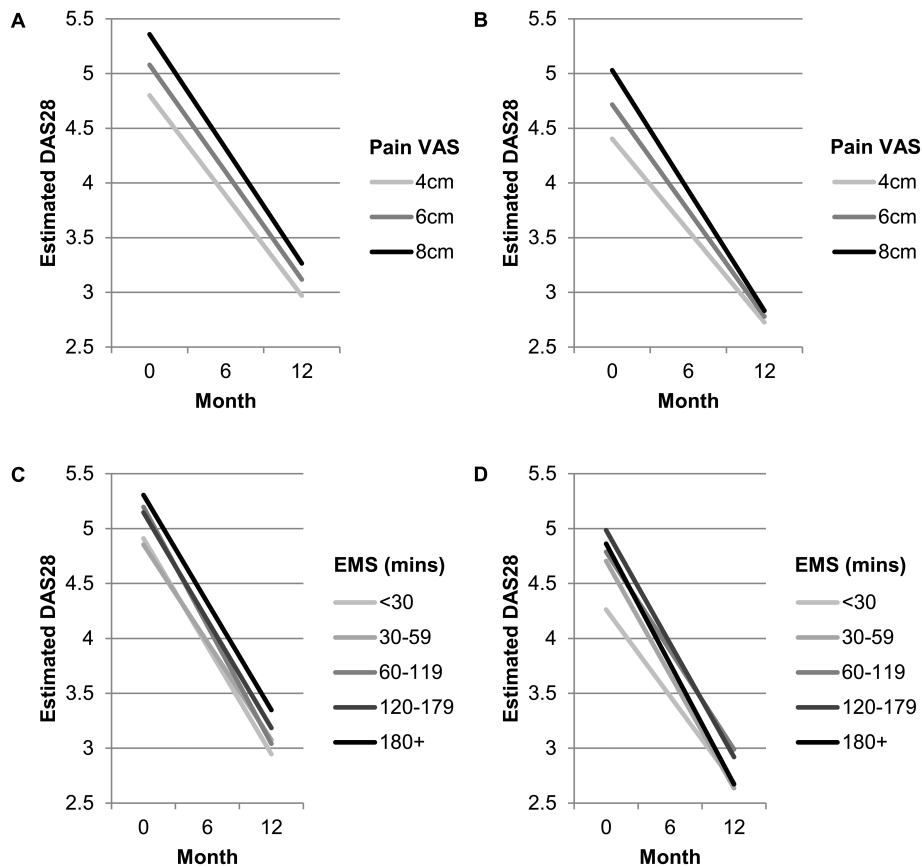
*Figure 2*. Estimated DAS28 at varying values of (A) pain VAS in patients recruited up to 2002, (B) pain VAS in patients recruited after 2002, (C) EMS in patients recruited up to 2002, and (D) EMS in patients recruited after 2002. DAS28: Disease Activity Score in 28 joints; VAS: visual analog scale; EMS: early morning stiffness in minutes.

who cannot take full doses of MTX or other DMARD because of comorbidities or intolerance. The effect of treat-to-target on change in DAS28 and HAQ is an area for further study.

We are not aware of any other studies that have analyzed the use of PRO to predict change in DAS28 and HAQ-DI in RA. However, several studies have highlighted the contribution of noninflammatory pain to overall disease activity scores. The pain index of DAS28 (DAS28-P), described by researchers from the Early Rheumatoid Arthritis Network[22], is the proportion of overall DAS28 derived from its subjective components. Improvement in pain measured using the Medical Outcomes Study Short Form-36 (SF-36) questionnaire after 1 year was less likely in patients with higher baseline DAS28-P[22]. Recently, in patients from a Danish cohort with RA completing the painDETECT questionnaire (designed to classify pain into low, medium, or high likelihood of being non-nociceptive), those whose scores indicated non-nociceptive pain had greater overall DAS28 and DAS28-P, measured at the time of questionnaire completion[23]. Therefore, any association of baseline pain

with subsequent change in DAS28 (as seen predominantly in our post-2002 cohort) may reflect an association with the subjective DAS28 components rather than inflammation alone.

EMS is a disabling symptom that fluctuates with RA disease activity[24], and helps to differentiate patients with RA from noninflammatory arthralgia[25]. In a prospective study that examined the effect of severity of EMS on early retirement, greater EMS at baseline was correlated with simultaneous measurements of DAS28, pain, and function, and those with severe stiffness at baseline were more likely to retire from employment within 3 years of followup[26]. Further findings from our study included an absence of association between EMS and radiographic progression, which was later supported by evidence from the Leiden Early Arthritis Clinic and ESPOIR cohorts in which prolonged EMS (> 60 min) was not associated with poor prognosis in terms of radiographic outcome after 3–7 years, or failure to achieve remission after 5–10 years or followup[27]. Although our study reported an association of greater EMS at baseline with greater rate of reduction in DAS28 in some patients in

*Table 3*. Results of multilevel regression models of change in HAQ-DI. Age was entered into the model as (age in years)/10. For EMS, the referent category was < 30 min.

| Variables | Imputed Data, n = 1416 | | | Complete Case, n = 579 | | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | p | Coefficient | SE | p |
| Predictors of baseline HAQ-DI | | | | | | |
| Cohort, post-2002 vs pre-2002 | –0.120 | 0.048 | 0.012 | –0.039 | 0.076 | 0.608 |
| Female, pre-2002 | 0.091 | 0.041 | 0.027 | 0.068 | 0.061 | 0.266 |
| Age, yrs | 0.028 | 0.011 | 0.009 | 0.026 | 0.018 | 0.136 |
| Symptom duration, pre-2002, mos | 0.006 | 0.005 | 0.205 | 0.012 | 0.007 | 0.106 |
| RF-positive | 0.008 | 0.041 | 0.850 | 0.013 | 0.061 | 0.836 |
| ACPA-positive | –0.014 | 0.045 | 0.748 | 0.004 | 0.057 | 0.942 |
| Pain VAS, cm, pre-2002 | 0.034 | 0.009 | < 0.001 | 0.042 | 0.013 | 0.001 |
| Fatigue VAS | 0.029 | 0.006 | < 0.001 | 0.029 | 0.009 | 0.002 |
| DAS28 | 0.135 | 0.013 | < 0.001 | 0.123 | 0.020 | < 0.001 |
| EMS duration, min | | | | | | |
| 30–59 | 0.006 | 0.053 | 0.912 | 0.025 | 0.085 | 0.765 |
| 60–119 | 0.019 | 0.046 | 0.681 | –0.003 | 0.070 | 0.971 |
| 120–179 | 0.096 | 0.051 | 0.061 | 0.076 | 0.080 | 0.344 |
| ≥ 180 | 0.134 | 0.047 | 0.005 | 0.139 | 0.077 | 0.072 |
| Cohort differences in predictors of baseline HAQ-DI, post-2002 vs pre-2002 | | | | | | |
| Female | 0.126 | 0.054 | 0.021 | 0.121 | 0.086 | 0.160 |
| Symptom duration, mos | –0.012 | 0.006 | 0.063 | –0.015 | 0.010 | 0.131 |
| Pain VAS | 0.008 | 0.011 | 0.461 | –0.008 | 0.017 | 0.661 |
| Change in HAQ-DI per mo* | –0.028 | 0.005 | < 0.001 | –0.031 | 0.007 | < 0.001 |
| Predictors of change in HAQ-DI, difference in change per mo | | | | | | |
| Cohort, post-2002 vs pre-2002 | –0.004 | 0.003 | 0.133 | –0.003 | 0.004 | 0.477 |
| Female | 0.005 | 0.003 | 0.089 | 0.009 | 0.004 | 0.033 |
| Age, yrs | 0.001 | 0.001 | 0.368 | 0.002 | 0.002 | 0.294 |
| Symptom duration, mos | 0.000 | 0.000 | 0.300 | 0.001 | 0.000 | 0.199 |
| RF-positive | 0.001 | 0.004 | 0.877 | –0.005 | 0.005 | 0.392 |
| ACPA-positive | 0.006 | 0.004 | 0.189 | 0.012 | 0.005 | 0.014 |
| Pain VAS, cm, pre-2002 | –0.001 | 0.001 | 0.215 | 0.000 | 0.001 | 0.737 |
| Fatigue VAS, cm | 0.000 | 0.001 | 0.887 | 0.000 | 0.001 | 0.947 |
| DAS28-CRP | –0.004 | 0.001 | < 0.001 | –0.005 | 0.002 | 0.002 |
| EMS duration, min | | | | | | |
| 30–59 | –0.006 | 0.005 | 0.181 | –0.012 | 0.007 | 0.108 |
| 60–119 | –0.002 | 0.004 | 0.684 | 0.002 | 0.006 | 0.749 |
| 120–179 | –0.002 | 0.005 | 0.603 | 0.004 | 0.007 | 0.589 |
| ≥ 180 | –0.012 | 0.004 | 0.006 | –0.014 | 0.007 | 0.037 |
| Cohort differences in predictors of change in HAQ-DI | | | | | | |
| Pain VAS, post-2002 vs pre-2002 | –0.002 | 0.001 | 0.053 | –0.002 | 0.001 | 0.250 |
| Model constant | 1.105 | 0.055 | < 0.001 | 1.096 | 0.085 | < 0.001 |

* In seronegative males with EMS < 30 min and mean age, symptom duration, DAS28, and pain and fatigue VAS recruited pre-2002. HAQ-DI: Health Assessment Questionnaire–Disability Index; EMS: early morning stiffness; RF: rheumatoid factor; ACPA: anticitrullinated protein antibodies; VAS: visual analog scale; DAS28: Disease Activity Score in 28 joints; CRP: C-reactive protein.

conflict with previous reports, the size of the effect was small (up to 0.06 units fall in DAS28 per month) and is unlikely to be clinically significant.

Data on fatigue and pain were identified in the form of VAS. Other methods of assessment are available to measure these variables, but the VAS was chosen because it was simple and quick for patients to complete alongside the other questionnaires that formed part of our study. A systematic review of scales to measure fatigue in RA identified 23 different scales, of which 6, including the VAS, had reasonable evidence of validation[28]. The review found evidence that a VAS performs reasonably well in terms of construct validity and discrimination, but there was little

evidence to demonstrate reliability and a lack of a standardized format. However, although the VAS has its limitations, no other measures are superior in terms of validation, and further, the single-item VAS likely performs as well as other, more detailed measures of fatigue[29]. Therefore, we feel that the use of VAS was justified.

A significant limitation of our study was the quantity of missing data: 40% and 43% cases had missing values for the 6- and 12-month analyses, respectively. Despite clear evidence that modern missing data management techniques such as MI are superior, traditional approaches such as analysis restricted to cases with no missing data (complete case analysis and weighted complete case analysis) are still
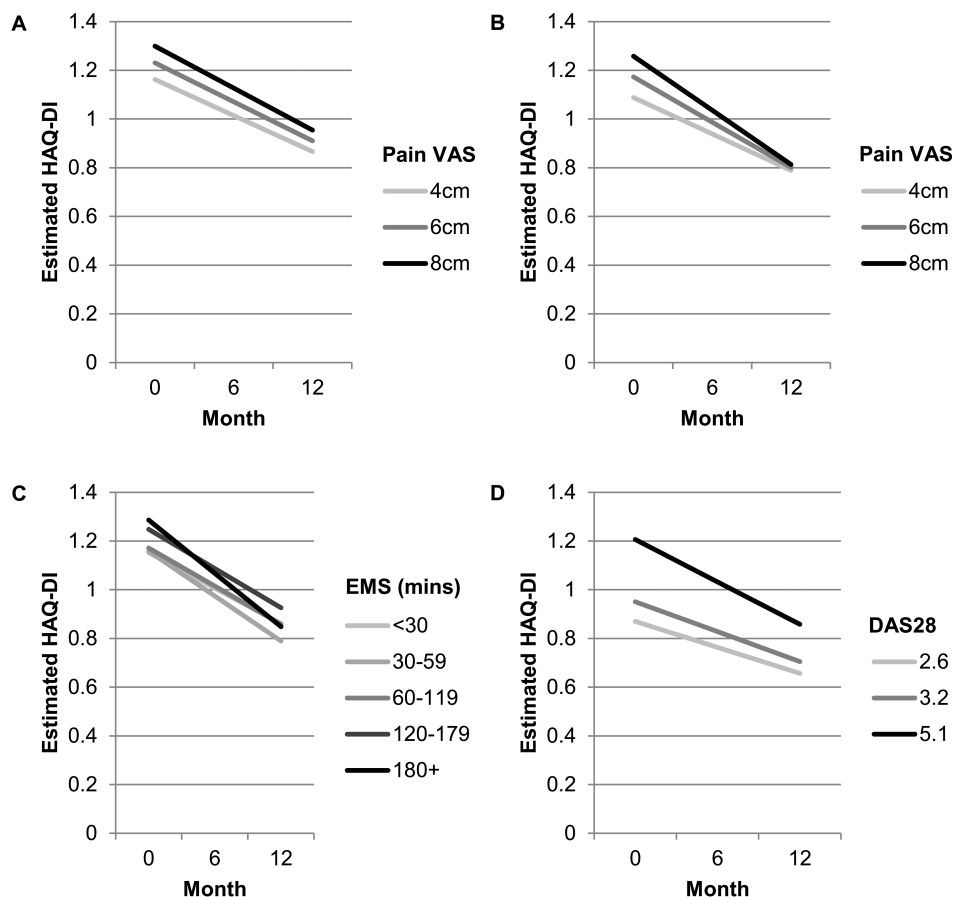
*Figure 3.* Estimated HAQ-DI at varying values of (A) pain VAS in patients recruited up to 2002, (B) pain VAS in patients recruited after 2002, (C) EMS in both cohorts, and (D) DAS28 in both cohorts. HAQ-DI: Health Assessment Questionnaire-Disability Index; VAS: visual analog scale; EMS: early morning stiffness in minutes; DAS28: Disease Activity Score in 28 joints.

reported. Not only does this technique lead to a loss of statistical power when cases with missing data are dropped, complete case analysis is also more likely to give biased estimates[30,31]. The MI models created for our present analyses were carefully constructed, which involved scrutiny of the dataset to identify auxiliary variables, inclusion of all variables in the analysis model within the imputation model, and comparison of results to a complete case model. Because of the large quantity of missing data, we cannot rule out bias in the results of the analyses because of missingness; however, simulation studies have demonstrated that MI is superior to complete case analysis, even when the quantity of missing data is large[32]. Nevertheless, potential bias because of missing data should be considered when interpreting our findings. For example, we found no relationship between RF and ACPA positivity and adverse outcome, in contrast to previous reports that indicated an inverse association between autoantibodies and future remission[33,34]. Evidence for the relationship between autoantibodies and HAQ has been mixed, with some evidence of an association between autoantibodies and worse disability[20,35], and some evidence to indicate there is no relationship between antibodies and HAQ[36,37]. The quantity of missing data was large for ACPA (39% of cases), so this is a potential source of bias.

An additional strength is the use of DAS28 and HAQ-DI as continuous rather than categorical or dichotomous (remission/non-remission) outcomes, thus improving statistical power. Although our study considered 3 separate patient-reported measures as predictors of outcome, it was not possible to assess the prediction value of several other similar variables. These include the RAPID3[38] and SF-36[39], which were not collected in the YEAR, and the VAS of global health status, which was collected in the YEAR but was not included in the statistical models because it was strongly correlated with pain VAS. Our present study was also limited to examining the predictive value of PRO collected at the baseline visit. It is possible that trends in the change of these variables would be more useful as predictors of outcome and therefore could be an area of interest for future study.

Our study showed that PRO at baseline, such as pain,

fatigue, and stiffness, are not useful for the prediction of rate of change in disease activity and disability.

## APPENDIX 1.

List of study collaborators. Yorkshire Early Arthritis Register consortium membership: Management Team: Professor Paul Emery, Professor Philip Conaghan, Professor Ann W. Morgan [Leeds Institute of Rheumatic and Musculoskeletal Medicine (LIRMM), Leeds Teaching Hospitals National Health Service (NHS) Trust], Professor Anne-Maree Keenan, and Dr. Elizabeth M.A. Hensor (LIRMM). Medical Staff: Dr. Mark Quinn (York District Hospital), Dr. Andrew Gough (Harrogate District Hospital), Dr. Michael Green (York District Hospital, Harrogate District Hospital), Dr. Richard Reece (Huddersfield Royal Infirmary), Dr. Lesley Hordon (Dewsbury District and General Hospital), Dr. Philip S. Helliwell (LIRMM, St. Luke's Hospital), Dr. Richard Melsom (St. Luke's Hospital), Dr. Sheelagh Doherty (Hull Royal Infirmary), Dr. Ade Adebajo (Barnsley District General Hospital), Dr. Andrew Harvey, Dr. Steve Jarrett (Pinderfields Hospital), Dr. Gareth Huston (LIRMM), Dr. Amanda Isdale (York District Hospital), Dr. Mike Martin (Leeds Teaching Hospitals NHS Trust), Dr. Zunaid Karim (Pinderfields Hospital), Professor Dennis McGonagle (LIRMM, Calderdale Royal Hospital), Dr. Colin Pease, Dr. Sally Cox (Leeds Teaching Hospitals NHS Trust), Dr. Victoria Bejarano (LIRMM), Dr. Jackie Nam, Dr. Edith Villeneuve, and Dr. Sarah Twigg (LIRMM, Leeds Teaching Hospitals NHS Trust). Nursing Staff: Claire Brown, Christine Thomas, David Pickles, Alison Hammond (LIRMM), Beverley Nevill (York District Hospital, Harrogate District Hospital), Alan Fairclough, Caroline Nunns (Huddersfield Royal Infirmary), Anne Gill, Julie Green (York District Hospital), Belinda Rhys-Evans, Barbara Padwell (Leeds Teaching Hospitals NHS Trust), Julie Madden, Lynda Taylor (Calderdale Royal Hospital), Sally Smith, Heather King (Leeds Teaching Hospitals NHS Trust), Jill Firth (St. Luke's Hospital), Jayne Heard (Hull Royal Infirmary), and Linda Sigsworth (St. Luke's Hospital). Support Staff: Diane Corscadden, Karen Henshaw, Lubna-Haroon Rashid, Stephen G. Martin, Dr. James I. Robinson, Dr. Lukasz Kozera, Dr. Agata Burska, Sarah Fahy, and Andrea Paterson (LIRMM).

## REFERENCES

1. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994 Sept;41:86-9.
2. Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. J Rheumatol 2007;34:1174-7.
3. Pincus T, Gibson KA, Berthelot JM. Is a patient questionnaire without a joint examination as undesirable as a joint examination without a patient questionnaire? J Rheumatol 2014;41:619-21.
4. Kojima M, Kojima T, Suzuki S, Takahashi N, Funahashi K, Asai S, et al. Patient-reported outcomes as assessment tools and predictors of long-term prognosis: a 7-year follow-up study of patients with rheumatoid arthritis. Int J Rheum Dis 2015 Nov 6 (E-pub ahead of print).
5. Castrejón I, Dougados M, Combe B, Fautrel B, Guillemin F, Pincus T. Prediction of remission in a French early arthritis cohort by RAPID3 and other core data set measures, but not by the absence of rheumatoid factor, anticitrullinated protein antibodies, or radiographic erosions. J Rheumatol 2016;43:1285-91.
6. Conaghan PG, Hensor EM, Keenan AM, Morgan AW, Emery P; YEAR Consortium. Persistently moderate DAS-28 is not benign: loss of function occurs in early RA despite step-up DMARD therapy. Rheumatology 2010;49:1894-9.
7. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44-8.
8. Morgan AW, Thomson W, Martin SG; Yorkshire Early Arthritis Register Consortium, Carter AM; UK Rheumatoid Arthritis Genetics Consortium, Erlich HA, et al. Reevaluation of the interaction between HLA-DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. Arthritis Rheum 2009;60:2565-76.
9. Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA. Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 1996;35:574-8.
10. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358-62.
11. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
12. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999;8:3-15.
13. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research—part 2: multiple imputation. Acad Emerg Med 2007;14:669-78.
14. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 1996;39:34-40.
15. Che H, Combe B, Morel J, Cantagrel A, Gossec L, Lukas C. Performance of patient-reported outcomes in the assessment of rheumatoid arthritis disease activity: the experience of the ESPOIR cohort. Clin Exp Rheumatol 2016;34:646-54.
16. Katchamart W, Johnson S, Lin HJ, Phumethum V, Salliot C, Bombardier C. Predictors for remission in rheumatoid arthritis patients: a systematic review. Arthritis Care Res 2010;62:1128-43.
17. Arnold MB, Bykerk VP, Boire G, Haraoui BP, Hitchon C, Thorne C, et al; CATCH Investigators. Are there differences between young- and older-onset early inflammatory arthritis and do these impact outcomes? An analysis from the CATCH cohort. Rheumatology 2014;53:1075-86.
18. Camacho EM, Verstappen SM, Lunt M, Bunn DK, Symmons DP. Influence of age and sex on functional outcome over time in a cohort of patients with recent-onset inflammatory polyarthritis: results from the Norfolk Arthritis Register. Arthritis Care Res 2011;63:1745-52.
19. Combe B, Rincheval N, Benessiano J, Berenbaum F, Cantagrel A, Daurès J, et al. Five-year favorable outcome of patients with early rheumatoid arthritis in the 2000s: data from the ESPOIR cohort. J Rheumatol 2013;40:1650-7.
20. Graell E, Vazquez I, Larrosa M, Rodríguez-Cros JR, Hernández MV, Gratacós J, et al. Disability measured by the modified health assessment questionnaire in early rheumatoid arthritis: prognostic factors after two years of follow-up. Clin Exp Rheumatol 2009;27:284-91.
21. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al; T2T Expert Committee. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631-7.
22. McWilliams DF, Zhang W, Mansell JS, Kiely PD, Young A, Walsh DA. Predictors of change in bodily pain in early rheumatoid arthritis: an inception cohort study. Arthritis Care Res 2012;64:1505-13.

23. Christensen AW, Rifbjerg-Madsen S, Christensen R, Dreyer L, Tillingsøe H, Seven S, et al. Non-nociceptive pain in rheumatoid arthritis is frequent and affects disease activity estimation: cross-sectional data from the FRAME study. Scand J Rheumatol 2016;45:461-9.

24. Orbai AM, Halls S, Hewlett S, Bartlett SJ, Leong AL, Bingham CO 3rd; RA Flare Group Steering Committee. More than just minutes of stiffness in the morning: report from the OMERACT Rheumatoid Arthritis Flare Group Stiffness Breakout Sessions. J Rheumatol 2015;42:2182-4.

25. van Steenbergen HW, Aletaha D, Beaart-van de Voorde LJ, Brouwer E, Codreanu C, Combe B, et al. EULAR definition of arthralgia suspicious for progression to rheumatoid arthritis. Ann Rheum Dis 2017;76:491-6.

26. Westhoff G, Buttgereit F, Gromnica-Ihle E, Zink A. Morning stiffness and its influence on early retirement in patients with recent onset rheumatoid arthritis. Rheumatology 2008;47:980-4.

27. van Nies JA, Alves C, Radix-Bloemen AL, Gaujoux-Viala C, Huizinga TW, Hazes JM, et al. Reappraisal of the diagnostic and prognostic value of morning stiffness in arthralgia and early arthritis: results from the Groningen EARC, Leiden EARC, ESPOIR, Leiden EAC and REACH. Arthritis Res Ther 2015;17:108.

28. Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. Arthritis Rheum 2007;57:429-39.

29. Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. J Rheumatol 2004;31:1896-902.

30. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255-64.

31. Knol MJ, Janssen KJ, Donders AR, Egberts AC, Heerdink ER, Grobbee DE, et al. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. J Clin Epidemiol 2010;63:728-36.

32. Janssen KJ, Vergouwe Y, Donders AR, Harrell FE Jr, Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. Clin Chem 2009;55:994-1001.

33. van der Woude D, Young A, Jayakumar K, Mertens BJ, Toes RE, van der Heijde D, et al. Prevalence of and predictive factors for sustained disease-modifying antirheumatic drug-free remission in rheumatoid arthritis: results from two large early arthritis cohorts. Arthritis Rheum 2009;60:2262-71.

34. van der Woude D, Visser K, Klarenbeek NB, Ronday HK, Peeters AJ, Kerstens PJ, et al. Sustained drug-free remission in rheumatoid arthritis after DAS-driven or non-DAS-driven therapy: a comparison of two cohort studies. Rheumatology 2012;51:1120-8.

35. Quinn MA, Gough AK, Green MJ, Devlin J, Hensor EM, Greenstein A, et al. Anti-CCP antibodies measured at disease onset help identify seronegative rheumatoid arthritis and predict radiological and functional outcome. Rheumatology 2006; 45:478-80.

36. Dirven L, Visser K, Klarenbeek NB, Ewals JA, Han KH, Peeters AJ, et al. Towards personalized treatment: predictors of short-term HAQ response in recent-onset active rheumatoid arthritis are different from predictors of rapid radiological progression. Scand J Rheumatol 2012;41:15-9.

37. Norton S, Sacker A, Dixey J, Done J, Williams P, Young A, et al; Early Rheumatoid Arthritis Study. Trajectories of functional limitation in early rheumatoid arthritis and their association with mortality. Rheumatology 2013;52:2016-24.

38. Pincus T, Bergman MJ, Yazici Y, Hines P, Raghupathi K, Maclean R. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in two abatacept clinical trials: similar results to disease activity score (DAS28) and other RAPID indices that include physician-reported measures. Rheumatology 2008;47:345-9.

39. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. BMJ 1993;306:1437-40.