EXTENDED REPORT

Effect of adherence to European treatment recommendations on early arthritis outcome: data from the ESPOIR cohort

Cécile Escalas,[1, 2] Marie Dalichampt,[2] Bernard Combe,[3] Bruno Fautrel,[4] Francis Guillemin,[5] Pierre Durieux,[6] Maxime Dougados,[7] Philippe Ravaud[1,7,8]

[1]INSERM U738, Paris, France
[2]Assistance Publique-Hopitaux de Paris, Hopital Cochin, Department of Rheumatology; University Paris Descartes –Sorbonne Paris Cité, Paris, France
[3]Department of Immunorhumatologie, CHU Lapeyronie, Montpellier, France
[4]Department of Rheumatology, Pierre et Marie Curie - Paris VI University, Paris, France
[5]Service d'épidémiologie et évaluation cliniques, Hopitaux de Brabois, Nancy, France
[6]Assistance Publique-Hopitaux de Paris, French Cochrane Center, Hopital Hotel Dieu, Paris, France
[7]Assistance Publique-Hopitaux de Paris, Hôpital Hôtel Dieu,Centre d'épidémiologie Clinique ; University Paris Descartes –Sorbonne Paris Cité, Paris, France
[8]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA

Correspondence to
Professor Philippe Ravaud, Hôpital Hôtel-Dieu, Centre d'Epidémiologie Clinique, Paris, France; philipperavaud@gmail.com

ABSTRACT

Objective To assess the association of adherence to the 2007 recommendations of the European League Against Rheumatism (EULAR) for managing early arthritis and radiographic progression and disability in patients

Methods The authors conducted a prospective population-based cohort study. The ESPOIR cohort was a French cohort of 813 patients with early arthritis not receiving disease-modifying antirheumatic drugs (DMARDs). Adherence to the 2007 EULAR recommendations was defined by measuring adherence to three of the recommendations concerning the initiation and early adjustment of DMARDs. The study endpoints were radiographic progression, defined as the presence of at least one new erosion between baseline and 1 year, and disability as a heath assessment questionnaire score ≥1 at 2 years. A propensity score of being treated according to the recommendations was developed.

Results After adjustment for propensity score, treatment centre and the main confounding factors, patients without recommendation adherence were at increased risk of radiographic progression at 1 year, and of functional impairment at 2 years (OR 1.98, (95% CI: 1.08 to 3.62 and OR: 2.36, (95% CI: 1.17 to 4.67), respectively).

Conclusions Early arthritis patients whose treatment adhered to the 2007 EULAR recommendations seemed to benefit from such treatment in terms of risk of clinical and radiographic progression. Using a propensity score of being treated according to recommendations in observational studies may be useful in assessing the potential impact of these recommendations on outcome.

INTRODUCTION

Physicians are increasingly being encouraged to integrate the best available scientific evidence in their routine medical practice. Clinical practice guidelines are a response to the practicing physician requiring assistance to assimilate and apply the exponentially expanding, often contradictory, body of medical knowledge. Guidelines are widely perceived as evidence based and therefore unbiased and valid if they are developed rigorously.[1 2] Most recommendations are based on imperfect evidence even if developed rigorously; therefore, a crucial element is to validate recommendations by demonstrating that adherence to the recommended strategies or recommendations is associated with better patient outcomes. Such studies evaluating the link between recommendation adherence and patient outcome are rarely performed even if considered crucial by some authors.[3–6]

The 2007 European League Against Rheumatism (EULAR) recommendations were developed for managing early arthritis. In managing rheumatoid arthritis (RA), early treatment, even if the diagnosis is still uncertain, has become a key point. Several studies[7–13] from which the EULAR recommendations evolved, highlight the interest in early initiation of disease-modifying antirheumatic drugs (DMARDs) in cases of persistent active disease and adjustment of therapy with remission as the only acceptable goal. We used data from the ESPOIR cohort, a cohort of patients with early arthritis, to assess the impact of adherence to the EULAR recommendations on outcomes (radiographic progression, disability) for patients with early arthritis. We applied the propensity score method. The propensity score has been proposed as a method of adjusting for the bias in treatment assignment in observational studies. Formally, the propensity score for an individual is the probability of receiving a particular treatment conditional on the individual's covariate values at baseline. This method has been used to assess the efficacy of treatment but never, to our knowledge, the impact of adherence to recommendations on outcome.

The objective of this study was to apply a new method of guidelines validation to assess the influence of adherence to EULAR recommendations on outcomes in patients suffering from early arthritis.

METHODS

Patients

The ESPOIR cohort[7] is an ongoing French multicentre prospective cohort of early arthritis patients initiated by the French Society of Rheumatology. Inclusion criteria were: age between 18 and 70 years; arthritis involvement of more than 2 joints; arthritis for at least 6 weeks and less than 6 months with a certain or probable clinical diagnosis of RA; and no DMARD or long-term steroid treatment since the onset of symptoms.

Overall, 813 patients with early arthritis were recruited between 2002 and 2005 in 14 regional centres and in collaboration with local private practitioners. Each centre acted as an observational centre and did not interfere with patient treatment.

Clinical and biological data[7] were prospectively and systematically collected at baseline and every 6 months at each investigating centre. Radiography of the hands and feet was performed once a year. The clinical and radiological data collected were available at 2 years and 1 year, respectively.

Adherence to recommendations

The 2007 EULAR recommendations[14] for managing early arthritis consist of several recommendations concerning diagnosis, treatment and monitoring. We focused on three recommendations concerning the initiation and early adjustment of DMARD therapy: (1) patients at risk of persistent or erosive arthritis should receive DMARDs as early as possible; (2) among the DMARDs, methotrexate should be used first; and (3) the only acceptable goal is to achieve disease remission and regular monitoring should guide decisions on changes in treatment strategies.

These three recommendations were interpreted by two rheumatologists (MD and CE) to determine adherence/non-adherence status for each patient. The definition of adherence was constructed in two steps (see online supplementary appendix 1). In a first step, we developed unambiguous descriptions for items with imprecise definition. Thus, the term remission, or the factors predicting persistent and erosive disease, was precisely defined. In a second step, algorithms were constructed to classify patients into two groups according to whether they received treatment as per each recommendation (yes/no). Then, we used an all-or-none measurement of adherence to the recommendations: adherence was true if treatment was according to all three recommendations.[15] Because the first stage of the treatment could be decisive,[16] adherence to recommendations was assessed over the first 6 months after the patient entered the cohort.

Study endpoints

The main endpoint was radiographic progression. Radiographic damage was scored by the Sharp score as modified by Van der Heijde.[17–20] For each patient, an erosion score was recorded for the hands and feet. Radiographs were read by a single reader, by paired and with information on the chronology of the films but blind to the clinical parameters and treatment received. According to the study sample with early arthritis and the high sensitivity of the chosen reading method, radiographic progression was defined as the occurrence of at least one new erosion between baseline and 1 year.

The secondary endpoint was functional ability as measured by the health assessment questionnaire (HAQ) at 2 years. The HAQ measures self-reported disability by a questionnaire with item scores ranging from 0 to 3 and increasing in steps of 0.125 units.[21] Disability was defined as a HAQ score ≥1 at 2 years.

Propensity modelling

We used the propensity score method,[22 23] which is usually used for estimating the causal effect of a non-randomised treatment on a health outcome with imbalanced baseline characteristics between treatment groups. Formally, the propensity score is the probability of receiving a particular treatment conditional on the individual baseline characteristics. Here, we used the propensity score to assess the probability of being treated according to the 2007 EULAR recommendations. Baseline data were used to construct the logistic regression model to predict the probability of being treated according to the recommendations. These variables were selected as follows. First, baseline variables

potentially associated with the 'recommendation adherence' assignment were analysed with adherence status in a bivariate analysis,[24 25] and those significant at p≤0.20 were included in a logistic regression model. Next, several propensity models were generated using an automatic process of selecting variables and by changing two parameters: the method of selecting variables (backward, stepwise) and the thresholds of entering and staying in the model (0.2, 0.1 or 0.05).[26] The treatment centre was forced into these models. The quality of the models was assessed by the Hosmer-Lemeshow test for 'goodness of fit' and the c-statistic for discriminatory ability. Clinical relevance of the models was assessed by comparing variables included in the final models to those known to be clinically relevant. Finally, the most accurate model was selected according to both its statistical and clinical relevance.[27 28] This model was used to calculate the predicted probability of each patient to be treated according to the recommendations. Patients were then divided into five equal groups by propensity score from lowest (quintile 1) to greatest propensity (quintile 5).[29] Candidate variables with more than 5% missing values from the score modelling were removed from the propensity score.

Statistical analysis

Baseline characteristics of patients were described by propensity score quintiles, frequency (percentage) for qualitative variables and mean (SD) or median (interquartile range) for quantitative variables. The relation between propensity score by quintile and outcome measures was tested by the χ^2 test. A two-tailed p value<0.05 was considered statistically significant.

We used multivariate analyses to estimate the adjusted effect of recommendations adherence on radiographic progression and on disability. Logistic models with mixed effects were used. The treatment centre was introduced as a random effect. Major confounding factors (age, sex, baseline erosions score and baseline HAQ for radiographic progression and disability, respectively) were introduced as fixed effects. Each of the models was performed with and without propensity quintile as an additional fixed effect.[29] Sensitivity analysis involved propensity score weighting and other definitions of outcome measures.

Statistical analysis was performed with SAS software, V. 9.1 (SAS Inst, Cary, North Carolina, USA).

RESULTS
Baseline patient characteristics

In total, 782 patients were included in our analysis. The mean (SD) age of patients was 48.3 (12.5) years and most were women (76.6%). The mean disease activity score for 28 joints (DAS28) was 5.1 (1.3), 55.5% of patients were positive for rheumatoid factor or anti-cyclic citrullinated peptide antibodies, and the mean symptom duration was 10.8 (11.0) weeks (table 1).

Adherence to recommendations

In all, adherence to the three recommendations could be determined for 782 patients (see online supplementary figure S1). The adherence rates were 78.3% (418/534), 66.8% (400/599) and 51.8% (378/730) for recommendations 1, 2 and 3, respectively. Finally, for 178 (22.8%) patients, treatment adhered to all three recommendations. Recommendations adherence differed among centres and ranged from 11.4% to 40.3%.

Propensity models

The most accurate propensity model included nine variables at baseline: centre, body mass index, personal income, professional

Table 1 Baseline characteristics of patients

	All patients n=782
Women, n (%)	599 (76.6)
Age (years), mean (SD)	48.3 (12.5)
Income, n (%):	
<€1220	392 (51.8)
≥€1220	365 (48.2)
Missing value	25 (3.2)
Professional status, n (%):	
Full-time	381 (48.7)
Part-time	82 (10.5)
Retired	152 (19.4)
Homemaker	60 (7.7)
Other	107 (13.7)
Body mass index (kg/cm^2), mean (SD)	25.0 (4.6)
Symptom duration (weeks), mean (SD)	10.8 (11.0)
Morning stiffness (min), median (IQR)	60.0 (15.0 to 120.0)
Pain (VAS), mean (SD)	55.6 (25.7)
Number of tender joints*, median (IQR)	6.0 (3.0 to 13.0)
Number of swollen joints*, median (IQR)	6.0 (3.0 to 10.0)
Disease activity score*, mean (SD)	5.1 (1.3)
State of health (VAS), median (IQR)	60.0 (45.0 to 75.0)
Health assessment questionnaire score, mean (SD)	1.0 (0.7)
C-reactive protein level (mg/l), median (IQR)	9.0 (0.0 to 24.0)
Rheumatoid factor or anti-cyclic citrullinated peptide antibodies positivity, n (%)	434 (55.5)
Human leucocyte antigen DRB*01 or 04 positivity, n (%)	407 (52.0)
Erosion score, median (IQR)	1.0 (0.0 to 3.0)

VAS, visual analogue scale.
*disease activity score on 28 joints: number of tender and swollen joints among 28 joints.

status, vaccination as a triggering factor, duration of morning stiffness, patient's assessment of global health, number of tender joints and symptom duration. The Hosmer-Lemeshow goodness-of-fit test gave a p value of 0.89, so the model fitted the data well. The c-statistic was 0.77 (95% CI 0.73 to 0.81), which indicated good discrimination. The mean propensity scores for adherence or not were 0.36 (0.19; range 0.03–0.88) and 0.19 (0.14; range 0.007–0.73), respectively. Baseline features suggested that the propensity of being treated according to the recommendations was lower for patients with more active, more severe and more recent disease than for others. Furthermore, these patients tended to exhibit socioeconomic disadvantages (table 2).

The propensity score by quintile was not associated with radiographic progression (p=0.75) but was associated with disability (p<0.0001) (table 3). In total, 35% of patients with the lowest quintile had a HAQ score ≥1 at 2 years for strong disability, as compared with 12% of those with the highest quintile.

Effect of recommendations adherence on radiographic progression at 1 year

Radiographs of the hands and feet at baseline and 1 year were available for 725 (92.7%) patients. Patients with missing radiographic data were included in the cohort later, had more active disease and less frequently showed positivity for human leucocyte antigen DRB1*01 or 04 than others. The baseline characteristics associated with radiographic progression were mostly those known to be prognostic factors of radiographic damage in early RA,[31] such as elevated baseline Sharp scores or positivity for rheumatoid factor (data not shown). In total, 159 (21.9%) patients showed 1 additional erosion at 1 year, 25 (14.6%) of those with adherence to the recommendations and 134 (24.2%) without adherence. After adjusting for treatment centre as a

Table 2 Baseline characteristics by propensity score quintiles of patients

	Quintiles				
	1	2	3	4	5
Women, n (%)	126 (85.1)	121 (81.8)	105 (70.9)	108 (73.0)	105 (70.9)
Age, mean (SD)	48.8 (13.4)	47.2 (13.3)	48.7 (12.4)	48.5 (11.5)	47.0 (12.0)
Body mass index, mean (SD)	27.5 (5.4)	25.2 (4.4)	24.6 (4.3)	24.4 (4.2)	23.6 (3.7)
Symptom duration, mean (SD)	60.4 (51.6)	65.9 (51.1)	60.3 (50.7)	81.1 (62.5)	108.9 (129.6)
Income, n (%)					
<€610	72 (48.6)	29 (19.6)	13 (8.8)	16 (10.8)	11 (7.4)
€610–1220	57 (38.5)	74 (50.0)	55 (37.2)	33 (22.3)	21 (14.2)
€1220–1830	13 (8.8)	28 (18.9)	48 (32.4)	53 (35.8)	55 (37.2)
€1830–2440	5 (3.4)	12 (8.1)	12 (8.1)	29 (19.6)	35 (23.6)
>€2440	1 (0.7)	5 (3.4)	20 (13.5)	17 (11.5)	26 (17.6)
Professional status, n (%)					
Full-time	37 (25.0)	76 (51.4)	71 (48.0)	88 (59.5)	93 (62.8)
Part-time	19 (12.8)	9 (6.1)	19 (12.8)	12 (8.1)	19 (12.8)
Retired	32 (21.6)	31 (20.9)	34 (23.0)	24 (16.2)	21 (14.2)
Homemaker	6 (4.1)	11 (7.4)	9 (6.1)	17 (11.5)	12 (8.1)
Other	54 (36.5)	21 (14.2)	15 (10.1)	7 (4.7)	3 (2.0)
Morning stiffness, n (%)					
No	1 (0.7)	3 (2.0)	2 (1.4)	14 (9.5)	36 (24.3)
<1 h	52 (35.1)	46 (31.1)	62 (41.9)	64 (43.2)	69 (46.6)
>1 h	95 (64.2)	99 (66.9)	84 (56.8)	70 (47.3)	43 (29.1)
Pain (VAS), mean (SD)	60.9 (23.9)	62.5 (25.6)	53.8 (27.0)	53.6 (22.4)	48.3 (25.6)
Number of tender joints*, median (IQR)	11.0 (7.0 to 16.5)	9.0 (5.5 to 15.0)	6.0 (4.0 to 13.5)	5.0 (3.0 to 10.5)	2.0 (1.0 to 5.0)
Number of swollen joints*, median (IQR)	8.0 (5.0 to 12.0)	7.0 (4.0 to 12.0)	6.0 (3.0 to 10.0)	6.0 (3.0 to 10.5)	3.0 (2.0 to 6.0)
Disease activity score*, mean (SD)	5.8 (1.1)	5.5 (1.1)	5.2 (1.3)	4.9 (1.3)	4.3 (1.3)

(Continued)

Table 2 (Continued)

	Quintiles				
	1	2	3	4	5
State of health, median (IQR)	50.0 (49.5–75.0)	55.0 (40.0 to 75.0)	60.0 (40.0 to 79.5)	67.0 (46.0 to 75.0)	70.0 (50.0 to 75.0)
HAQ, mean (SD)	1.3 (0.7)	1.1 (0.7)	1.0 (0.7)	0.9 (0.7)	0.7 (0.6)
C-reactive protein, median (IQR)	12.0 (5.0 to 26.0)	9.0 (0.0 to 24.5)	8.5 (0.0 to 25.5)	8.0 (0.0 to 24.5)	9.0 (0.0 to 21.0)
Rheumatoid factor or anti-CCP antibodies positivity, n (%)	80 (54.1)	83 (56.1)	85 (57.4)	84 (56.8)	82 (55.4)
Human leukocyte antigen DRB*01 or 04 positivity, n (%)	71 (48.0)	78 (52.7)	76 (51.4)	74 (50.0)	86 (58.1)
Erosion score, median (IQR)	1.0 (0.0 to 4.0)	1.0 (0.0 to 3.0)	2.0 (0.0 to 5.0)	1.0 (0.0 to 3.0)	1.0 (0.0 to 3.0)
Missing values	13	17	9	10	13

Quintile 1 represents patients with the lowest probability of being treated in accordance with the recommendations; quintile 5 represents patients with the highest probability of being treated in accordance with the recommendations, according to baseline characteristics.
anti-CCP, anti-cyclic citrullinated peptide; HAQ, health assessment questionnaire; VAS, visual analogue scale.
*disease activity score on 28 joints: number of tender and swollen joints among 28 joints.

Table 3 Relation between propensity score and radiographic progression and disability

Propensity score quintile	One additional erosion at 1 year		HAQ score ≥1 at 2 years	
	Observed probability (n/N)	χ^2 p level	Observed probability (n/N)	χ^2 p level
1	0.21 (28/135)	0.75	0.35 (44/126)	<0.0001
2	0.22 (29/133)		0.24 (30/124)	
3	0.24 (33/139)		0.19 (24/129)	
4	0.24 (33/138)		0.16 (21/131)	
5	0.18 (25/139)		0.12 (16/130)	

Quintile 1 represents patients with the lowest probability of being treated in accordance with the recommendations; quintile 5 represents patients with the highest probability of being treated in accordance with the recommendations, according to baseline characteristics.

Table 4 Effect of recommendation adherence on radiographic progression and disability

	One additional erosion at 1 year n=684*		HAQ score ≥1 at 2 years n=640*	
	OR (95% CI)	p Value	OR (95% CI)	p Value
Unadjusted	1.79 (1.11 to 2.89)	0.02	3.01 (1.70 to 5.33)	0.0002
Adjusted for confounding factors† and centre	1.94 (1.09 to 3.45)	0.03	2.55 (1.31 to 4.96)	0.009
Adjusted for propensity score	1.83 (1.10 to 3.05)	0.02	2.28 (1.25 to 4.18)	0.007
Adjusted for propensity score, confounding factors† and centre	1.98 (1.08 to 3.62)	0.03	2.36 (1.17 to 4.67)	0.02

Mixed logistic regression with confounding factors as fixed effects and treatment centre as a random effect.
*684 patients with a calculated propensity score among the 725 patients with the radiographic progression outcome available; 640 patients with a calculated propensity score among the 679 patients with the disability outcome available.
†Confounding factors included:
  for radiographic progression: age, sex and erosions score at baseline.
  for HAQ: age, sex and HAQ score at baseline.
HAQ, health assessment questionnaire.

random effect, confounding factors (age, sex, baseline erosion score) and propensity score as fixed effects, patients without adherence were at increased risk of radiographic progression (OR 1.98, 95% CI 1.08 to 3.62). Adjustment for only confounding factors and centre slightly increased the strength of the recommendations adherence effect on radiographic progression, whereas adjustment for only propensity score had no impact on this effect (table 4). The results remained the same after propensity score weighting or defining radiographic progression as a change in erosion score ≥1.

Effect of recommendations adherence on disability at 2 years

Data to evaluate disability at 2 years were available for 679 (86.8%) patients. Patients with missing data had significantly more pain, more socioeconomic disadvantages and less frequent positivity for human leucocyte antigen DRB1*01 or 04 than others. In total, 145 (21.3%) patients had a HAQ score ≥1 at 2 years, 17 (10.6%) with adherence to the recommendations and 128 (24.7%) without adherence. At baseline, patients with adherence had a lower HAQ score than those without adherence (0.8 (0.7) vs 1.0 (0.7)). After adjustment for treatment centre as a random effect, confounding factors (age, sex, baseline HAQ score) as fixed effects and propensity score, patients without adherence were at increased risk of an HAQ score ≥1 at 2 years (OR 2.36, 95% CI 1.17 to 4.67) (table 4). The results were similar after propensity score weighting or using a lower cut-off to define disability (0.7).

DISCUSSION

Our study suggests that adherence to three recommendations of the 2007 EULAR recommendations reduces the risk of radiographic progression at 1 year and disability at 2 years in patients with early arthritis. We propose a new way of validating clinical practice recommendations based on the use of the propensity score, which seems particularly interesting in the area of recommendations assessment because evidence for the impact of recommendations on treatment cannot be obtained by randomised trials.

Practical guidelines are frequently used as a source of performance measures framed as imperatives and designed to convey what must be done rather than what should be done for disease management. However, most treatment recommendations are based on imperfect evidence even if developed rigorously. Gaps in evidence are inevitable, so they must be filled with judgments and experts' opinions. For example, even though statins are generally considered to reduce the risk of vascular events, many specifics remain debatable, such as the intensity of the therapy and when therapy should be started. Moreover, many recommendations are based on low levels of evidence or on expert opinion.[32][33] Surprisingly, the number of recommendations that lack conclusive evidence remains static. As an example, over the last 25 years, only 11% of recommendations from the American College of Cardiology and the American Heart Association[32] are quoted with a level of evidence A, and this percentage has remained relatively constant over time. The most common grade

of evidence remains a C. Similar results have been observed for Infectious Diseases Society of America guidelines.[33] In the field of rheumatology, despite the efforts of EULAR to improve the methodology permitting to propose recommendations, because of the frequent lack of large trials in different clinical situations, the evidence-based formulation of recommendations might be considered as weaker. Finally, even if evidence from randomised trials is available, external validity or applicability of the trials' supporting evidence could be debatable. So, assessing the impact of practical guidelines in daily practice is crucial for validation.

Many well-designed intervention studies have evaluated the impact of recommendation implementation strategies on processes of care or patient outcomes. Such studies cannot distinguish between the implementation strategy tested and the validity of the recommendations used. Even valid recommendations may not have any impact on patient outcomes if the implementation strategy fails. By contrast, even if the implementation strategy leads to perfect adherence, invalid recommendations will not improve patient outcomes.

Initially, the propensity score was established as the probability of assigning a certain treatment conditional on the individual characteristics observed at baseline. The underlying idea is that use of the probability that a subject would have been treated to adjust the estimate of the treatment effect results in a 'quasi-randomised' experiment. Until now, this method has been used to assess the efficacy of treatment on outcome but not, to our knowledge, the impact of adherence to recommendations on outcomes. The factors influencing adherence to a recommendation are complex, related both to the severity of disease and to external parameters such attitude and knowledge of patients and physicians; these external barriers are all the more important that makes the patient management decisions complex. The propensity of a patient to receive treatment according to recommendations could vary according to all these parameters. Thus, the propensity score method has further relevance when we assess the impact of adherence to recommendations on outcomes.

The recommendations in the current EULAR recommendations were developed by an expert committee using an evidence-based approach. Among the 12 recommendations of this guideline, the first 2 we selected were derived from a meta-analysis of randomised trials (level of evidence Ia) and the third from at least one randomised trial (level of evidence Ib).[14] No study has assessed the impact of these recommendations in daily practice. Previous observational studies[8 27 34 35] and one cohort study of RA[36] analysed the effect of early treatment on early RA outcome, and their results are consistent with our findings. In the ESPOIR cohort, we noted differences in baseline characteristics between patients with or without adherence to the recommendations, which justified a propensity score adjustment. Patients without recommendation adherence had socioeconomic disadvantages and also more severe and active disease than others, as was previously found for coronary disease.[37] Adherence rates for each of the three recommendations were between 52% and 78%, but the adherence rate for all three recommendations was only 23%. Each one of these recommendations did not fit well with the others and recommendation three seemed to be the most selective.

Our study has several limitations. First, despite the large number of baseline variables available in the ESPOIR cohort, the possibility of unobserved confounding covariates was reduced but cannot be ruled out. This situation is the main limitation in the use of the propensity score as compared with the randomised clinical trial in which the randomisation procedure should balance known and unknown confounders between the groups. Indeed, a number of factors such as patient factors (comorbidity, compliance with treatment and monitoring) and clinician characteristics were not available for inclusion in the propensity model. Moreover, we did not assess adherence to all the EULAR recommendations but only to the ones concerning first-line treatment with DMARDs and change in treatment strategies. We did not assess adherence to other recommendations concerning, for example, non-pharmacological treatment, safety or regular monitoring. Furthermore, we chose to define recommendations adherence over only the first 6 months because early stages of the treatment can be decisive for outcome in early arthritis and even prevent development of RA.[16] However, the quality of management after 6 months may also influence outcome.

In summary, patients with early arthritis whose treatment adhered to the 2007 EULAR recommendations seemed to have a clinical and radiological benefit. In observational studies, using a propensity score of adherence to treatment recommendations may be useful in assessing the potential impact of these recommendations on outcome.

REFERENCES

1. **Cabana MD**, Rand CS, Powe NR, *et al*. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999;**282**:1458–65.
2. **Grol R**. Has guideline development gone astray? Yes. *BMJ* 2010;**340**:c306.
3. **Avins AL**. When clinical practice guidelines meet the black box. *Arch Intern Med* 2010;**170**:1013–14.
4. **Krumholz HM**. Guideline recommendations and results: the importance of the linkage. *Ann Intern Med* 2007;**147**:342–3.
5. **McAlister FA**, van Diepen S, Padwal RS, *et al*. How evidence-based are the recommendations in evidence-based guidelines? *PLoS Med* 2007;**4**:e250.
6. **Powers JH**. Practice guidelines: belief, criticism, and probability. *Arch Intern Med* 2011;**171**:15–17.
7. **Combe B**, Benessiano J, Berenbaum F, *et al*. The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. *Joint Bone Spine* 2007;**74**:440–5.
8. **Lard LR**, Visser H, Speyer I, *et al*. Early versus delayed treatment in patients with recent-onset rheumatoid arthritis: comparison of two cohorts who received different treatment strategies. *Am J Med* 2001;**111**:446–51.
9. **Boers M**, Verhoeven AC, Markusse HM, *et al*. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet* 1997;**350**:309–18.
10. **Landewé RB**, Boers M, Verhoeven AC, *et al*. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. *Arthritis Rheum* 2002;**46**:347–56.
11. **Möttönen T**, Hannonen P, Leirisalo-Repo M, *et al*. Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. *Lancet* 1999;**353**:1568–73.
12. **Möttönen T**, Hannonen P, Korpela M, *et al*. Delay to institution of therapy and induction of remission using single-drug or combination-disease-modifying antirheumatic drug therapy in early rheumatoid arthritis. *Arthritis Rheum* 2002;**46**:894–8.
13. **Nell VP**, Machold KP, Eberl G, *et al*. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. *Rheumatology (Oxford)* 2004;**43**:906–14.
14. **Combe B**, Landewe R, Lukas C, *et al*. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2007;**66**:34–45.

15. **Nolan T**, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA* 2006;**295**:1168–70.
16. **Sharp JT**. Speculations on why early treatment of rheumatoid arthritis is uniquely effective. *J Rheumatol* 2008;**35**:2090–3.
17. **van der Heijde DM**, van Riel PL, Nuver-Zwart IH, *et al*. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* 1989;**1**:1036–8.
18. **Sharp JT**, Lidsky MD, Collins LC, *et al*. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. *Arthritis Rheum* 1971;**14**:706–20.
19. **Sharp JT**, Young DY, Bluhm GB, *et al*. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? *Arthritis Rheum* 1985;**28**:1326–35.
20. **van der Heijde D**. How to read radiographs according to the Sharp/van der Heijde method. *J Rheumatol* 1999;**26**:743–5.
21. **Fries JF**, Spitz P, Kraines RG, *et al*. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.
22. **Rosenbaum PR**. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41–55.
23. **Rubin DB**. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;**127**:757–63.
24. **Austin PC**, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med* 2007;**26**:734–53.
25. **Brookhart MA**, Schneeweiss S, Rothman KJ, *et al*. Variable selection for propensity score models. *Am J Epidemiol* 2006;**163**:1149–56.
26. **Weitzen S**, Lapane KL, Toledano AY, *et al*. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf* 2004;**13**:841–53.
27. **Wiles NJ**, Lunt M, Barrett EM, *et al*. Reduced disability at five years with early treatment of inflammatory polyarthritis: results from a large observational cohort, using propensity models to adjust for disease severity. *Arthritis Rheum* 2001;**44**:1033–42.
28. **Park GS**, Wong WK, Oh M, *et al*. Classifying radiographic progression status in early rheumatoid arthritis patients using propensity scores to adjust for baseline differences. *Stat Methods Med Res* 2007;**16**:13–29.
29. **D'Agostino RB** Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;**17**:2265–81.
30. **Benhamou M**, Rincheval N, Roy C, *et al*. The gap between practice and guidelines in the choice of first-line disease modifying antirheumatic drug in early rheumatoid arthritis: results from the ESPOIR cohort. *J Rheumatol* 2009;**36**:934–42.
31. **Combe B**, Dougados M, Goupille P, *et al*. Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study. *Arthritis Rheum* 2001;**44**:1736–43.
32. **Tricoci P**, Allen JM, Kramer JM, *et al*. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009;**301**:831–41.
33. **Lee DH**, Vielemeyer O. Analysis of overall level of evidence behind Infectious Diseases Society of America practice guidelines. *Arch Intern Med* 2011;**171**:18–22.
34. **Bukhari MA**, Wiles NJ, Lunt M, *et al*. Influence of disease-modifying therapy on radiographic outcome in inflammatory polyarthritis at five years: results from a large observational inception study. *Arthritis Rheum* 2003;**48**:46–53.
35. **van Aken J**, Lard LR, le Cessie S, *et al*. Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis. *Ann Rheum Dis* 2004;**63**:274–9.
36. **Tanaka E**, Mannalithara A, Inoue E, *et al*. Efficient management of rheumatoid arthritis significantly reduces long-term functional disability. *Ann Rheum Dis* 2008;**67**:1153–8.
37. **Lee DS**, Tu JV, Juurlink DN, *et al*. Risk-treatment mismatch in the pharmacotherapy of heart failure. *JAMA* 2005;**294**:1240–7.
38. **Visser H**, le Cessie S, Vos K, *et al*. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. *Arthritis Rheum* 2002;**46**:357–65.