OPEN ACCESS

EXTENDED REPORT

Evaluation of the change in structural radiographic sacroiliac joint damage after 2 years of etanercept therapy (EMBARK trial) in comparison to a contemporary control cohort (DESIR cohort) in recent onset axial spondyloarthritis

Maxime Dougados,[1] Walter P Maksymowych,[2] Robert B M Landewé,[3] Anna Moltó,[1] Pascal Claudepierre,[4] Manouk de Hooge,[5] Robert G Lambert,[6] Randi Bonin,[7] Jack F Bukowski,[7] Heather E Jones,[8] Isabelle Logeart,[9] Ron Pedersen,[10] Annette Szumski,[11] Bonnie Vlahos,[7] Désirée van der Heijde[5]

Handling editor Tore K Kvien
For numbered affiliations see end of article.

Correspondence to
Professor Maxime Dougados, Paris Descartes University, Rheumatology Department, Cochin Hospital, Assistance Publique - Hôpitaux de Paris, INSERM (U1153): Clinical Epidemiology and Biostatistics, PRES Sorbonne Paris-Cité, France;
maxime.dougados@aphp.fr

Received 30 June 2017
Revised 1 September 2017
Accepted 19 September 2017
Published Online First
29 September 2017

ABSTRACT

Objective To compare 2 years of radiographic sacroiliac joint (SIJ) changes in patients with recent onset axial spondyloarthritis (axSpA) receiving etanercept in a clinical trial (EMBARK) to similar patients not receiving biologics in a cohort study (DESIR).

Methods Endpoints were changes at week 104 per the modified New York (mNY) grading system in total SIJ score (primary endpoint) and net percentage of patients with progression defined three ways. Treatment effect was analysed with and without adjustment for baseline covariates.

Results At 104 weeks, total SIJ score improved in the etanercept group (n=154, adjusted least-squares mean change: −0.14) and worsened in the control group (n=182, change: 0.08). The adjusted difference between groups (etanercept minus control) was −0.22 (95% CI −0.38 to −0.06), p=0.008. The net percentage of patients with progression was significantly lower in the etanercept versus the control group for two of three binary endpoints: −1.9% versus 1.6% (adjusted difference for etanercept minus control: −4.7%,95% CI −9.9 to 0.5, p=0.07) for change in mNY criteria; −1.9% versus 7.8% (adjusted difference: −18.2%,95% CI −30.9 to −5.6, p=0.005) for change ≥1 grade in ≥1 SIJ; and −0.6% versus 6.7% (adjusted difference: −16.4%,95% CI −27.9 to −5.0, p=0.005) for change ≥1 grade in ≥1 SIJ, with shift from 0 to 1 or 1 to 0 considered no change.

Conclusion Despite the slow radiographic SIJ progression rate over 2 years in axSpA, this study suggests a lower rate of progression in the SIJ with etanercept than without anti-tumour necrosis factor therapy.

Trial registration numbers NCT01258738,NCT01648907; Post-results.

To cite: Dougados M, Maksymowych WP, Landewé RBM, *et al*. *Ann Rheum Dis* 2018;**77**:221–227.

INTRODUCTION

The most frequently observed symptoms in spondyloarthritis (SpA) are axial.[1 2] The various criteria for SpA (eg, Amor, European Spondyloarthropathy Study Group and, more recently, the Assessment of SpondyloArthritis international Society (ASAS) criteria) enable classification of patients in the absence of radiographic structural damage, that is, non-radiographic axial SpA (nr-axSpA).[3–7] In patients with an inadequate response to non-steroidal anti-inflammatory drugs (NSAIDs) with radiographic (r-) or nr-axSpA, anti-tumour necrosis factor (TNF) agents have demonstrated a beneficial effect on symptoms,[8–11] but their structural effect is still unclear.[12–17]

Structural evaluation of axSpA can be performed using conventional radiographs or MRI at the spine or pelvic level. Radiographic axSpA studies have focused on the spine using a radiography scoring system, and data suggest that a structural effect either does not exist[18 19] or requires studies >2 years to be observed.[20 21] Questions exist about the risk of future structural damage, particularly at the sacroiliac joint (SIJ) level, in patients with nr-axSpA. Approximately 10% of patients with nr-axSpA develop SIJ radiographic damage within 2 years and 60% within 10 years.[22–24]

The conventional method for assessing SIJ structural damage on radiography is the modified New York (mNY) grading system, consisting of a semiquantitative scale from 0 (normal) to 4 (total ankylosis).[2] However, this method has been criticised because of its poor reliability.[25] Moreover, this grading system has no accepted method to evaluate change in radiographic damage except the categorisation of a patient as having either nr-axSpA or r-axSpA: r-axSpA is considered to be at least grade 2 bilaterally or at least grade 3 unilaterally. Alternative outcome measures appear to be more sensitive, such as change in the total score over time, and percentage of patients with a change of at least one grade in at least one SIJ.[24 26]

Ideally, a long-term controlled clinical trial would address the structural impact of long-term treatment. Additionally, a robust study should include both a treatment and a control group. However, it is not possible to conduct a study of sufficient length, that is, at least 2 years, with a placebo control.[21]

Another option is to compare a treatment cohort from one study to a control cohort in another study. This technique has been used to

evaluate the structural changes observed at the spine level in r-axSpA in patients receiving an anti-TNF. These patients have been compared with a control group consisting of patients in a study evaluating the natural history of r-axSpA, the OASIS cohort.[13 16 17 27]

All of these considerations prompted us to conduct a study in patients with early axSpA aimed at evaluating the radiographic changes in the SIJ observed after 2 years of etanercept therapy in patients enrolled in a clinical trial (EMBARK) compared with usual care in patients enrolled in an observational cohort (DESIR).

PATIENTS AND METHODS

Details of the EMBARK trial have been described previously.[8 28 29] All patients fulfilled the ASAS criteria for axSpA, but based on a central reading procedure, none of them met the mNY criteria for radiographic status. Patients were aged ≥18 and <50 years with symptoms for >3 months but <5 years, had a Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) score ≥4 of 10, and had symptoms of back pain with an inadequate response to ≥2 NSAIDs. After a 12-week, double-blind, placebo-controlled period, all patients received etanercept 50 mg once weekly during a 92-week open-label period.

The DESIR cohort has been described in detail.[23] The study included patients aged >18 and <50 years with inflammatory back pain for >3 months but <3 years, suggestive of axSpA according to the treating rheumatologist. Patients with a history of treatment with any biological therapy were excluded.

The present analysis included the patients from the EMBARK trial with available baseline and 2-year pelvic radiographs, and patients from the DESIR cohort who met the ASAS criteria for axSpA, did not receive any biological therapy during the first two years of follow-up and had baseline and 2-year pelvic radiographs.

Grading of radiographic sacroiliitis

Radiographic sacroiliitis was graded using the 0–4 grade scale for the left and right SIJ from the mNY grading system.[2] The scale is provided below:

- Grade 0: normal.
- Grade 1: suspicious changes.
- Grade 2: minimal abnormality—small localised areas with erosion or sclerosis, without alteration in the joint width.
- Grade 3: unequivocal abnormality—moderate or advanced sacroiliitis with one or more of erosions, evidence of sclerosis, widening, narrowing or partial ankylosis.
- Grade 4: severe abnormality—total ankylosis.

Reading the radiographs

The SIJ radiographs from the DESIR and EMBARK cohorts were anonymised so that the readers were unaware of the chronology of the films and the original patient cohort. The three trained and experienced readers, who were not readers used for screening in either DESIR or EMBARK, met via videoconference for a calibration session prior to the start of this analysis. They graded each joint at each time point, with a scale from 0 to 4 per the mNY grading system.

Assessments

The primary endpoint was change in total SIJ score at week 104. Total SIJ score was obtained by adding the scores of both SIJs according to the mNY grading system (0–4 per SIJ, range from 0 to 8); thus the change could range from −8 to +8. For this endpoint, the mean change of the three readers' values was used. Three binary endpoints were also evaluated: (1) proportion of patients switching from mNY criteria negative at baseline to mNY criteria positive at week 104 and the proportion of patients switching from mNY criteria positive at baseline to mNY criteria negative at week 104 (based on the central reading for the current analysis); and (2) proportion of patients with change (improvement or worsening in SIJ score of ≥1) in at least one SIJ. The third binary endpoint excluded minimal or doubtful changes (changes from normal appearance (grade 0) to 'suspicious' abnormalities of the SIJ (grade 1)) from the improved or worsened categories: proportion of patients with change (improvement or worsening in SIJ score of ≥1) in at least one SIJ, with a shift from 0 to 1 (in the worsened joint) or from 1 to 0 (in the improved joint) considered no change. For these binary endpoints, improvement or worsening was assigned only if at least two of the three readers agreed on the direction of change.

Other collected data

In both studies, patient demographics and clinical outcome measures of disease activity were collected at baseline and throughout the duration of the follow-up. The baseline SIJ MRI evaluating the presence of inflammation according to the Spondyloarthritis Research Consortium of Canada (SPARCC) method[30] was assessed separately in EMBARK and DESIR using a central reading procedure previously described.[29 31] A score ≥2 was considered an indicator of SIJ inflammation on MRI.[32]

Statistical analysis

This analysis included the completer population, defined as having pelvic radiographs available at baseline and 2 years. Baseline characteristics were analysed using either the Wilcoxon rank-sum or the Mantel-Haenszel χ^2 test. The radiographic analyses were conducted without covariates (unadjusted analysis) and also with the following covariates as potential baseline confounders (adjusted analysis): sex, symptom duration, smoking status, human leucocyte antigen (HLA)-B27 status, Ankylosing Spondylitis Disease Activity Score (ASDAS) with C reactive protein, SPARCC MRI SIJ score and total SIJ score based on the mNY grading system. One-way analysis of variance was used to compare study cohorts for the unadjusted difference, and analysis of covariance was used for the adjusted difference.

The a priori primary outcome measure was the absolute change in total SIJ score adjusted for baseline covariates. For each of the three binary endpoints, the percentage of patients with disease progression (worsening) and the percentage of patients with disease regression (improvement) was determined per group. Additionally, the net percentage of patients with progression was defined as the number of patients with worsening minus the number of patients with improvement, divided by the total study population. The between-group difference in the net percentage of patients with progression was reported for each of the three binary endpoints. A cumulative probability plot was generated to compare the change in SIJ radiography score from baseline to week 104 for the control and etanercept cohorts. Change was defined as the average change of the three readers.

RESULTS

The EMBARK trial included 225 randomised patients; a complete data set was available for 162 patients. The DESIR cohort study enrolled 708 patients; 506 of these patients did not receive a biological therapy during the 2 years of follow-up, 283 of these 506 patients fulfilled the ASAS criteria for axSpA and

Table 1 Demographics and baseline disease characteristics

	Control (DESIR) n=193	Etanercept (EMBARK) n=162	p Value
Age, years	32.2 (7.0)	31.8 (7.7)	0.47*
Male, n/N (%)	100/193 (51.8)	106/162 (65.4)	0.01†
Symptom duration, years	1.7 (1.0)	2.4 (1.8)	<0.001*
Current smoker, n/N (%)	70/192 (36.5)	37/162 (22.8)	0.006†
HLA-B27(+), n/N (%)	162/193 (83.9)	113/156 (72.4)	0.009†
BASDAI (0–10)	3.6 (1.9)	5.9 (1.8)	<0.001*
ASDAS	2.2 (0.9)	3.0 (1.0)	<0.001*
BASFI (0–10 cm VAS)	2.2 (2.0)	4.0 (2.4)	<0.001*
CRP, mg/L	5.4 (7.5)	6.9 (11.2)	0.06*
SPARCC MRI SIJ score (0–72)	5.8 (9.5)	8.4 (11.0)	<0.001*
SPARCC MRI SIJ score ≥2, n/N (%)	78/191 (40.8)	95/159 (59.7)	<0.001†
Total SIJ score (mNY grade 0–8)	1.9 (1.6)	1.5 (1.2)	0.03*
SIJ score met mNY criteria, n/N (%)	39/193 (20.2)	19/162 (11.7)	0.03†

Values are mean (SD) unless otherwise noted.
*Wilcoxon rank-sum.
†Mantel-Haenszel.
ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; CRP, C-reactive protein; HLA, human leucocyte antigen; mNY, modified New York; SIJ, sacroiliac joint; SPARCC, Spondyloarthritis Research Consortium of Canada; VAS, visual analogue scale.

193 had both baseline and 2-year pelvic radiographs available and qualified for this study. Demographics and baseline disease characteristics are provided in table 1.

At baseline, several differences existed between the groups: a higher proportion of males and longer disease duration in the etanercept group, and a higher proportion of smokers and HLA-B27-positive patients in the control group. Because all EMBARK patients were eligible for initiation of anti-TNF therapy and none of the DESIR cohort received an anti-TNF during the 2-year follow-up period, it is not surprising that the disease activity markers of BASDAI, ASDAS and SPARCC MRI SIJ inflammation were significantly higher in the etanercept group at baseline. Conversely, total SIJ score was slightly but significantly higher in the control group.

After 104 weeks, there was a slightly positive change (worsening) in the total SIJ score for the control group versus a slightly negative change (improvement) in the etanercept group in the adjusted analysis (least-squares mean change: 0.08 (95% CI −0.04 to 0.20) vs −0.14 (95% CI −0.26 to −0.01)). The adjusted between-group difference in change (etanercept − control) was significant: −0.22 (95% CI −0.38 to −0.06, p=0.008); the unadjusted between-group difference was not significant: −0.11 (95% CI −0.25 to 0.02, p=0.10).

Figure 1 presents the cumulative probability plot for the change in SIJ radiography score over 104 weeks. The control cohort trended towards worsening, with more patients having a positive score. In contrast, the etanercept cohort trended towards improvement, with more patients having a negative score.

The observed radiographic changes from baseline to week 104 are shown in table 2. For change in mNY criteria, the net percentage of patients with progression was lower in the etanercept versus the control group; however, the difference between the groups was not statistically significant: −1.9% versus 1.6% (adjusted difference for etanercept minus control: −4.7%, 95% CI −9.9 to 0.5, p=0.07). For the other two binary endpoints, the net percentage of patients with progression was significantly lower in the etanercept versus the control group: −1.9% versus 7.8% (adjusted difference: −18.2%, 95% CI −30.9 to −5.6, p=0.005) for change ≥1 grade in at least one SIJ; and −0.6% versus 6.7% (adjusted difference: −16.4%, 95% CI −27.9 to −5.0, p=0.005) for change ≥1 grade in at least one SIJ, with shift from 0 to 1 or 1 to 0 considered no change.

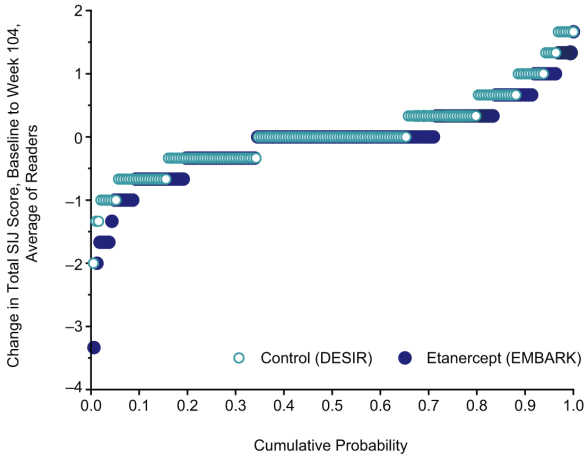

Figure 1 Cumulative probability of change in sacroiliac joint (SIJ) radiography score from baseline to week 104 for the control and etanercept cohorts, average of the readers.

Figure 2 presents the net percentage of patients with progression in the two study groups for the three binary endpoints.

DISCUSSION

This study supports the existence of a small structural effect of anti-TNF therapy in the SIJ using plain pelvic radiographs as the primary assessment tool and the mNY grading system as the outcome measure. It also confirms the relatively slow rate of radiographic progression in the SIJ in terms of shifting from non-radiographic to radiographic status according to the mNY criteria over a 2-year period.

An assessment of 2-year SIJ radiographic progression in early axSpA was also conducted in the German Spondyloarthritis Inception Cohort (GESPIC), a cohort comparable to DESIR.[24] A similar rate of SIJ radiographic progression was observed, with a mean change in the SIJ score of 0.07 (95% CI −0.05 to 0.19) and 0.09 (95% CI −0.03 to 0.21) for the left and right SIJ, respectively.[24] Moreover, in the GESPIC cohort, after 2 years, 11 of the 95 patients with nr-axSpA at baseline met the mNY criteria for r-axSpA (ie, worsened).[24] Additionally, 3 of the 115 patients with r-axSpA at baseline did not fulfil the mNY criteria at year 2 (ie, improved).[24] Calculating the net rate of progression for the full study population results in a rate of 3.8% ((11 − 3)/(95+115)). This is similar to the data observed in the present study for the DESIR patients (control group), with a net progression rate of 1.6% (95% CI −1.3% to 4.4%). The slight difference between the two studies may be due to chance or may be explained by different patient phenotypes, in particular, the proportion of patients with SIJ inflammation on MRI (greater in the GESPIC cohort than in this DESIR subgroup).

When considering an outcome parameter based on a semiquantitative variable (score 0–4 per side), including different types of damage, collected in the left and right joints, some concerns may

Table 2 Observed radiographic changes from baseline to week 104

Endpoint	Cohort	Improved n/N (%)	Worsened n/N (%)	Net % patients with progression,*† unadjusted (95% CI)	Between-group differences in net % patients with progression, Etanercept—Control (95% CI)	
					Unadjusted analysis	Adjusted analysis‡
Δ in mNY criteria	Control Etanercept	3/193 (1.6) 4/162 (2.5)	6/193 (3.1) 1/162 (0.6)	1.6% (−1.3 to 4.4) −1.9% (−4.9 to 1.2)	−3.4% (−7.6 to 0.8) p=0.11	−4.7% (−9.9 to 0.5) p=0.07
Δ ≥1 grade in ≥1 SIJ	Control Etanercept	21/193 (10.9) 19/162 (11.7)	36/193 (18.7) 16/162 (9.9)	7.8% (0.6 to 15.0) −1.9% (−9.7 to 6.0)	−9.6% (−20.3 to 1.0) p=0.08	−18.2% (−30.9 to −5.6) p=0.005
Δ ≥1 grade in ≥1 SIJ; shift from 0 to 1 or 1 to 0 considered no Δ	Control Etanercept	16/193 (8.3) 15/162 (9.3)	29/193 (15.0) 14/162 (8.6)	6.7% (0.3 to 13.2) −0.6% (−7.6 to 6.4)	−7.4% (−16.9 to 2.2) p=0.13	−16.4% (−27.9 to −5.0) p=0.005

Based on two of three readers assigning same category; otherwise considered no change.
Some patients started with lowest possible score and could not improve.
*Net % patients with progression=number of patients with worsening minus the number of patients with improvement, divided by the study population.
†One-way analysis of variance.
‡Adjusted for these covariates at baseline: sex, symptom duration, smoking status, human leucocyte antigen-B27 status, Ankylosing Spondylitis Disease Activity Score, Spondyloarthritis Research Consortium of Canada MRI SIJ score and SIJ mNY grade.
Δ, change; mNY, modified New York; SIJ, sacroiliac joint.

be raised. Semiquantitative scores may not be translated into continuous scores without consideration since it is unknown if the steps in the semiquantitative score are equidistant. While this is a technical limitation of our study, this approach is frequently used in medicine in general and in rheumatology in particular, and we do not believe that it has influenced the results.

Dichotomisation is a frequently used technique to overcome scaling issues related to semiquantitative scores and interpretational concerns from continuous scores. Dichotomisation also assists in the analysis of non-normally distributed data. It is tempting for clinicians to interpret radiographic change scores as dichotomies (those that progress vs those that do not; those that have nr-axSpA vs those that have r-axSpA). However, dichotomisation is a simplification of the truth because it largely ignores measurement error. Measuring radiographic change in patients with SpA is a challenge since the true change ('the signal') in a patient is usually outweighed by spurious change ('the noise') due to differences in technique and inherent rater variability. An observed difference between groups is only credible if the scores have been obtained under unbiased conditions and all possible directions of change have been considered.

'Net percentage of patients with progression' is a concept we explored to combine the advantages of dichotomisation ('progressor' or 'non-progressor') while preserving the option of adjusting for measurement error. It is an artificial concept in terms of interpretation since it appears possible in a single patient to adjust the true signal for the noise of measurement error, which is not the case. Net percentage of patients with progression should be interpreted at the group level. Although more patients had disease progression than disease regression overall, this difference cannot be translated to an individual patient. Therefore, the concept does not elementarily differ from the comparison of group means.

Another potential issue when using the mNY grading system as an outcome measure is that two concepts are mixed: repair (sclerosis) and destruction (joint erosion). One patient may have a change in sclerosis and another may have a change in erosion, and the grade change could be the same. Additionally, the results can vary between readers since the inter-reader reliability of this approach is known to be quite poor.[25 26] In EMBARK there was a greater proportion of patients with regression than progression, resulting in a negative parameter estimate for progression rate. This may be due to measurement error or a true repair process with a reduction in erosions.

The switch from a continuous or semiquantitative to a binary variable (progression yes/no) necessitates choosing a cut-off. Because the conventional yet arbitrary mNY criteria distinguish between radiographic and non-radiographic status, it was tempting to use these to describe a patient at a particular time point and to estimate the natural disease history. It was also tempting to present the results in a simpler, more understandable manner, such as change of ≥1 grade in ≥1 SIJ. We used the approach proposed by the GESPIC investigators. However, because of the difficulty in distinguishing a grade 0 from a grade 1, we modified this system by excluding the change from grade 0 to grade 1 for the worsened joint or from grade 1 to 0 for the improved joint.[26]

These results suggest a significant structural effect of etanercept in the SIJ. The treatment group was not compared with a control group within a prospective randomised controlled trial; rather, it was compared with a contemporary cohort of patients. Consequently, the baseline characteristics differed between the two groups, particularly the disease activity. All patients in EMBARK were eligible for anti-TNF therapy; the DESIR patients in this study did not receive biological therapy. Therefore, we adjusted for covariates that may affect radiographic progression.[24 26 33]

To our knowledge, this is the first study to evaluate the anti-TNF structural effect in the SIJ using plain pelvic radiography as the assessment tool and the mNY grading system as the scoring method. These results should be considered within the context of the literature. Previous studies of radiographic progression in axSpA evaluated the spine since structural damage in the spine correlates with functional impairment. However, study results suggest that a longer period of evaluation is needed to observe a structural anti-TNF effect in the spine.[13 16 17 34 35] The clinical relevance of our study may be more difficult to interpret since the correlation between a change in radiographic SIJ damage and the functional capacity of a patient is usually considered poor. Future studies are needed to better evaluate the predictive validity of this outcome measure.

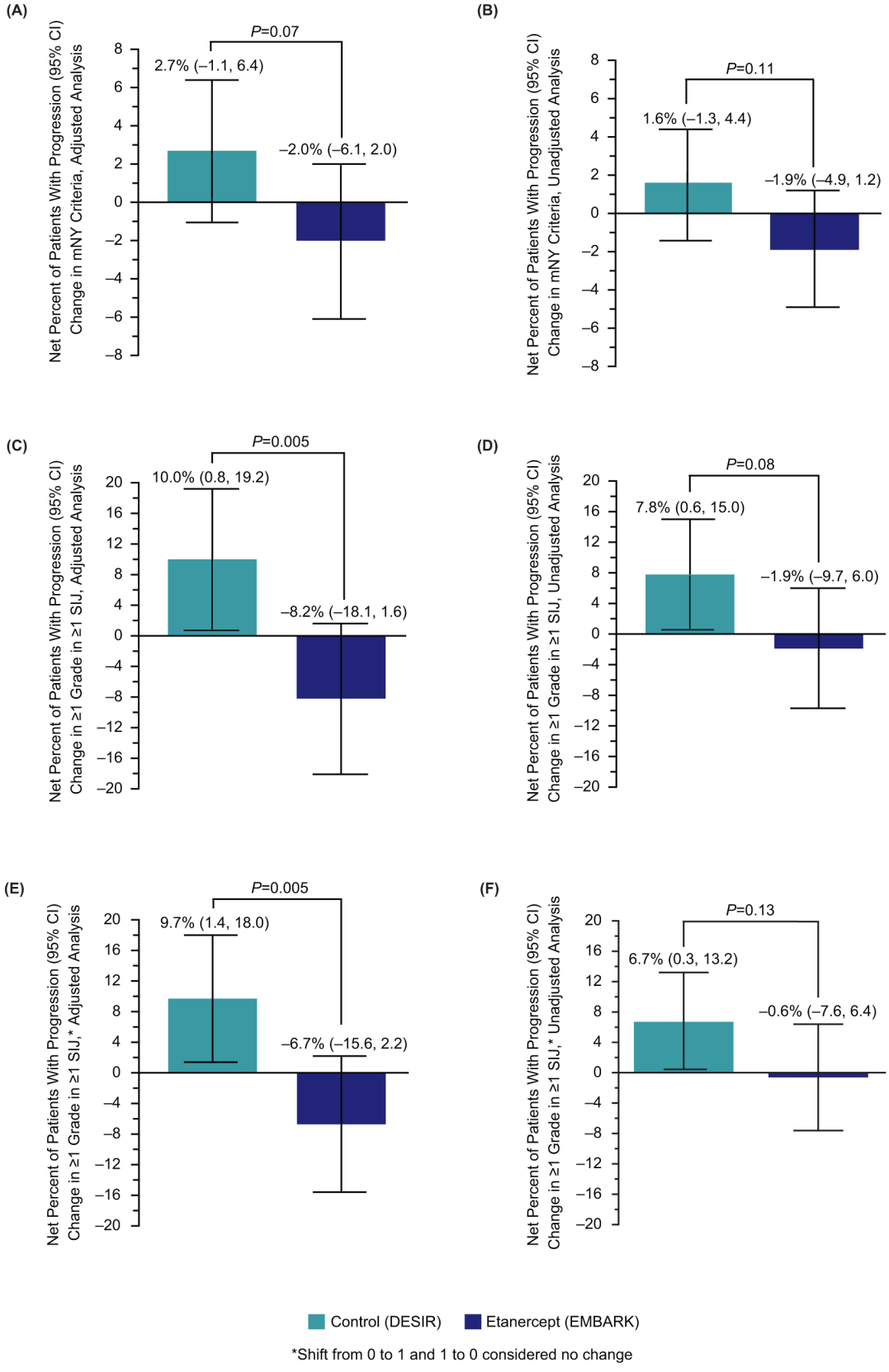

Figure 2 Net percent of patients with progression (number of patients with worsening minus the number of patients with improvement, divided by the study population) from baseline to week 104 in each study group for the three binary endpoints: (A) change in modified New York (mNY) criteria, adjusted analysis and (B) unadjusted analysis; (C) change of ≥1 grade in ≥1 sacroiliac joint (SIJ), adjusted analysis and (D) unadjusted analysis; and (E) change of ≥1 grade in ≥1 SIJ with shift from 0 to 1 and from 1 to 0 considered no change, adjusted analysis and (F) unadjusted analysis.

Our study has several strengths. First, both study cohorts had a large sample size. Second, the scoring methodology was designed to avoid and adjust for bias, that is, the three independent, trained readers were unaware of the chronology of the radiographs and the patient cohort. Third, the study included a control group. Even though both cohorts were not randomised as a whole, the control group was an appropriate comparison for the etanercept group.

These results further support a structural anti-TNF effect in the SIJ.[36] The data are promising, but additional studies are needed to confirm the validity of these outcome measures and to evaluate the structural effect of various therapies in the SIJ using advanced imaging techniques.

Author affiliations
[1]Paris Descartes University, Rheumatology Department, Cochin Hospital, Assistance Publique-Hôpitaux de Paris, INSERM (U1153): Clinical Epidemiology and Biostatistics, PRES Sorbonne Paris-Cité, France
[2]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada
[3]Amsterdam Rheumatology and Immunology Center, Amsterdam, Netherlands
[4]Universite Paris Est Creteil, EA 7379 – EpidermE, AP-HP, Service de Rhumatologie, Hopital Henri Mondor, Creteil, France
[5]Leiden University Medical Center, Leiden, Netherlands
[6]Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Alberta, Canada
[7]Clinical Affairs, Pfizer, Collegeville, Pennsylvania, USA
[8]Medical Affairs, Pfizer, Collegeville, Pennsylvania, USA
[9]Pfizer France, Paris, France
[10]Department of Biostatistics, Pfizer, Collegeville, USA
[11]Biostatistics, inVentiv Health, Princeton, New Jersey, USA

Acknowledgements The authors thank all patients who participated in this study, as well as the investigators and medical staff at all of the participating centres. This paper is based on work that was previously presented at the 2017 Annual Meeting of the European League Against Rheumatism (EULAR); 14-17 June 2017; Madrid, Spain; and was published as a conference abstract: Dougados M, *et al* Ann Rheum Dis. 2017;76:350-1.

Contributors Study conception or design: MD, WPM, RL, DvdH, JFB, HJ, IL, RP, AS. Acquisition of data: MD, WPM, RL, DvdH, AM, PC, MdH, RGL, RB. Analysis or interpretation of data: MD, WPM, RL, DvdH, AM, PC, MdH, RGL, RB, JFB, HJ, IL, RP, AS, BV. All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. All authors agree to be accountable for all aspects of the work.

Funding The EMBARK trial was funded by Pfizer. The DESIR cohort is supported by unrestricted grants from the French Society of Rheumatology and Pfizer. Medical writing support was provided by Jennica Lewis, PharmD, CMPP, of Engage Scientific Solutions and was funded by Pfizer.

Competing interests MD reports grants and personal fees from Pfizer, AbbVie, UCB, Merck, Lilly, Janssen and Novartis during the conduct of the study. WPM reports grants and personal fees from AbbVie and Pfizer, personal fees from Janssen, Lilly, Novartis, Merck and UCB outside the submitted work. RGL reports personal fees from AbbVie and BioClinica outside the submitted work. JFB, HJ, IL, RP and BV are employees of, and own stock in Pfizer. RB was an employee of Pfizer at the time the article was written. AS is an employee of inVentiv Health and was contracted by Pfizer to provide statistical support for the development of this paper. DvdH reports personal fees from AbbVie, Amgen, Astellas, AstraZeneca, Bristol Meyers Squibb, Boehringer Ingelheim, Celgene, Daiichi, Galapagos,Glaxo-Smith-Kline, Janssen, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda and UCB outside the submitted work; and is director of Imaging Rheumatology BV. RL, AM, PC and MdH have no competing interests to declare.

Ethics approval EMBARK: The institutional review board or independent ethics committee at each participating centre reviewed and approved all consent forms and the study protocol; DESIR: Ile de France III Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Rudwaleit M, van der Heijde D, Landewé R, *et al*. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. *Ann Rheum Dis* 2011;70:25–31.
2 Sieper J, Rudwaleit M, Baraliakos X, *et al*. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Ann Rheum Dis* 2009;68 (Suppl 2):1–ii44.
3 Akgul O, Ozgocmen S. Classification criteria for spondyloarthropathies. *World J Orthop* 2011;2:107–15.
4 Dougados M, van der Linden S, Juhlin R, *et al*. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;34:1218–27.
5 Rudwaleit M, Landewé R, van der Heijde D, *et al*. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 2009;68:770–6.
6 Rudwaleit M, van der Heijde D, Landewé R, *et al*. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777–83.
7 Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. *Rev Rhum Mal Osteoartic* 1990;57:85–9.
8 Maksymowych WP, Dougados M, van der Heijde D, *et al*. Clinical and MRI responses to etanercept in early non-radiographic axial spondyloarthritis: 48-week results from the EMBARK study. *Ann Rheum Dis* 2016;75:1328–35.
9 Braun J, Brandt J, Listing J, *et al*. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. *Lancet* 2002;359:1187–93.
10 Baraliakos X, Haibel H, Fritz C, *et al*. Long-term outcome of patients with active ankylosing spondylitis with etanercept-sustained efficacy and safety after seven years. *Arthritis Res Ther* 2013;15:R67.
11 van der Heijde D, Kivitz A, Schiff MH, *et al*. ATLAS Study Group. Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2006;54:2136–46.
12 Baraliakos X, Listing J, Brandt J, *et al*. Radiographic progression in patients with ankylosing spondylitis after 4 yrs of treatment with the anti-TNF-alpha antibody infliximab. *Rheumatology* 2007;46:1450–3.
13 van der Heijde D, Salonen D, Weissman BN, *et al*. Canadian (M03-606) study group, ATLAS study group. Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. *Arthritis Res Ther* 2009;11:R127.
14 Osman MS, Maksymowych WP. An update on the use of tumor necrosis factor alpha inhibitors in the treatment of ankylosing spondylitis. *Expert Rev Clin Immunol* 2017;13:125–31.
15 Song IH, Hermann KG, Haibel H, *et al*. Inflammatory and fatty lesions in the spine and sacroiliac joints on whole-body MRI in early axial spondyloarthritis-3-Year data of the ESTHER trial. *Semin Arthritis Rheum* 2016;45:404–10.
16 van der Heijde D, Landewé R, Einstein S, *et al*. Radiographic progression of ankylosing spondylitis after up to two years of treatment with etanercept. *Arthritis Rheum* 2008;58:1324–31.
17 van der Heijde D, Landewé R, Baraliakos X, *et al*. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. *Arthritis Rheum* 2008;58:3063–70.
18 Ramiro S, van Tubergen A, Stolwijk C, *et al*. Scoring radiographic progression in ankylosing spondylitis: should we use the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) or the Radiographic Ankylosing Spondylitis Spinal Score (RASSS)? *Arthritis Res Ther* 2013;15:R14.
19 Maksymowych WP. Controversies in conventional radiography in spondyloarthritis. *Best Pract Res Clin Rheumatol* 2012;26:839–52.
20 Poddubnyy D, Protopopov M, Haibel H, *et al*. High disease activity according to the Ankylosing Spondylitis Disease Activity Score is associated with accelerated radiographic spinal progression in patients with early axial spondyloarthritis: results from the GErman SPondyloarthritis Inception Cohort. *Ann Rheum Dis* 2016;75:2114–8.
21 van der Heijde D, Landewé R, van der Linden S. How should treatment effect on spinal radiographic progression in patients with ankylosing spondylitis be measured? *Arthritis Rheum* 2005;52:1979–85.
22 Poddubnyy D, Brandt H, Vahldiek J, *et al*. The frequency of non-radiographic axial spondyloarthritis in relation to symptom duration in patients referred because of chronic back pain: results from the Berlin early spondyloarthritis clinic. *Ann Rheum Dis* 2012;71:1998–2001.
23 Dougados M, Etcheto A, Molto A, *et al*. DESIR cohort. Clinical presentation of patients suffering from recent onset chronic inflammatory back pain suggestive of spondyloarthritis: The DESIR cohort. *Joint Bone Spine* 2015;82:345–51.
24 Poddubnyy D, Rudwaleit M, Haibel H, *et al*. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. *Ann Rheum Dis* 2011;70:1369–74.

25 van den Berg R, Lenczner G, Feydy A, *et al*. Agreement between clinical practice and trained central reading in reading of sacroiliac joints on plain pelvic radiographs. Results from the DESIR cohort. *Arthritis Rheumatol* 2014;66:2403–11.
26 Dougados M, Demattei C, van den Berg R, *et al*. Rate and Predisposing Factors for Sacroiliac Joint Radiographic Progression After a Two-Year Follow-up Period in Recent-Onset Spondyloarthritis. *Arthritis Rheumatol* 2016;68:1904–13.
27 Ramiro S, Stolwijk C, van Tubergen A, *et al*. Evolution of radiographic damage in ankylosing spondylitis: a 12 year prospective follow-up of the OASIS study. *Ann Rheum Dis* 2015;74:52–9.
28 Dougados M, van der Heijde D, Sieper J, *et al*. Effects of long-term etanercept treatment on clinical outcomes and objective signs of inflammation in early nonradiographic axial spondyloarthritis: 104-week results from a randomized, placebo-controlled study. *Arthritis Care Res* 2017.
29 Dougados M, van der Heijde D, Sieper J, *et al*. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheumatol* 2014;66:2091–102.
30 Maksymowych WP, Inman RD, Salonen D, *et al*. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. *Arthritis Rheum* 2005;53:703–9.
31 van den Berg R, Lenczner G, Thévenin F, *et al*. Classification of axial SpA based on positive imaging (radiographs and/or MRI of the sacroiliac joints) by local rheumatologists or radiologists versus central trained readers in the DESIR cohort. *Ann Rheum Dis* 2015;74:2016–21.
32 van den Berg R, de Hooge M, Bakker PA, *et al*. Metric Properties of the SPARCC Score of the Sacroiliac Joints - Data from Baseline, 3-month, and 12-month Followup in the SPACE Cohort. *J Rheumatol* 2015;42:1186–93.
33 Poddubnyy D, Sieper J. Similarities and differences between nonradiographic and radiographic axial spondyloarthritis: a clinical, epidemiological and therapeutic assessment. *Curr Opin Rheumatol* 2014;26:377–83.
34 Baraliakos X, Haibel H, Listing J, *et al*. Continuous long-term anti-TNF therapy does not lead to an increase in the rate of new bone formation over 8 years in patients with ankylosing spondylitis. *Ann Rheum Dis* 2014;73:710–5.
35 Haroon N, Inman RD, Learch TJ, *et al*. The impact of tumor necrosis factor α inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum* 2013;65:2645–54.
36 Pedersen SJ, Wichuk S, Chiowchanwisawakit P, *et al*. Tumor necrosis factor inhibitor therapy but not standard therapy is associated with resolution of erosion in the sacroiliac joints of patients with axial spondyloarthritis. *Arthritis Res Ther* 2014;16:R100.