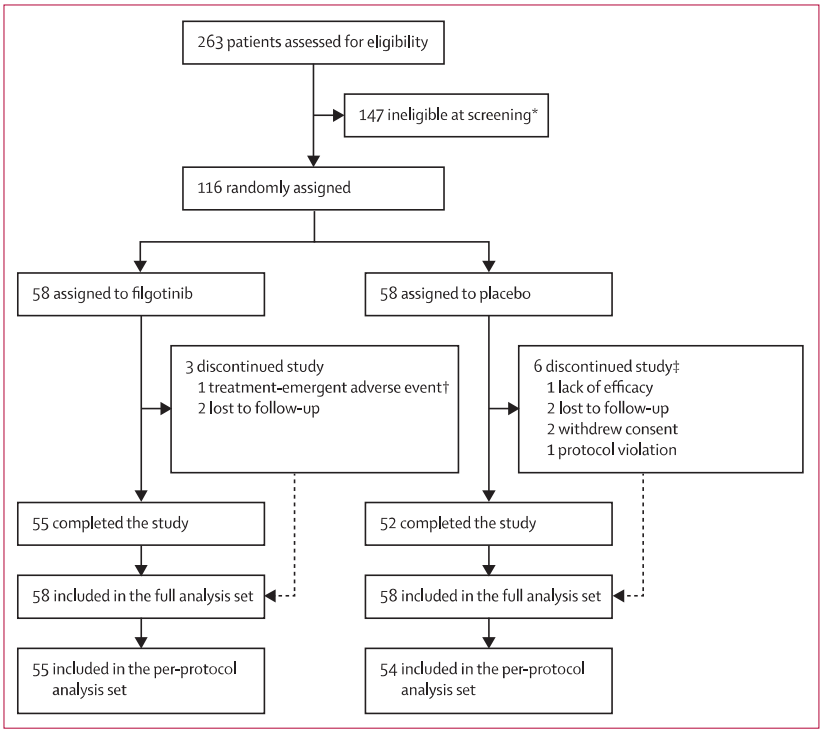

Figure 1: **Trial profile**
*Patients could be ineligible for more than one reason, the most common reasons being not fulfilling criteria for active ankylosing spondylitis (n=67; of whom 58 did not fulfil the Modified New York criteria based on the central reading and nine did not fulfil the diagnosis or criteria for another reason); having concentrations of high-sensitivity C-reactive protein <3·0 mg/L (n=45); having positive serology for HIV-1, HIV-2, hepatitis B virus (HBV), or hepatitis C virus (HCV), or any history of infection with HBV or HCV (n=30); having out-of-range laboratory values (n=13); and having untreated or inadequately treated tuberculosis infection (n=9). †Case of grade 3 pneumonia in a woman aged 49 years who was a current smoker. ‡One patient temporarily discontinued treatment because of an adverse event (grade 3 neutropenia) but restarted treatment and completed all study visits.

Safety endpoints were the incidence of adverse events, serious adverse events, and adverse events of special interest (appendix p 94); treatment discontinuations due to adverse events; and changes in laboratory results, electrocardiograms, physical examination results, and vital signs over time. The severity of adverse events was graded with the modified Common Terminology Criteria for Adverse Events (CTCAE), version 4.03; if CTCAE criteria did not exist, grades were allocated according to definitions provided in the appendix (p 14).

Statistical analysis

We calculated that a total sample size of 100 patients would have 81% power to detect a difference of –0·6 in the primary endpoint between filgotinib and placebo. This calculation was based on an unequal variances *t* test, with a two-sided significance level of 5%, and assumed, on the basis of previous studies, that the mean change from baseline to week 12 would be –0·65 (SD 0·83) in the placebo group and –1·25 (1·2) in the filgotinib group.

We analysed the primary endpoint and other continuous variables (ie, changes from baseline) using an ANCOVA model that included factors for treatment, baseline values, and stratification factors. Normality assumptions were met for all changes in ASDAS from baseline at all timepoints in both groups except for the placebo group at week 1. ANCOVA models produced adjusted least squares means, SDs, and 95% CIs for between-group comparisons. Two-sided p values are provided for between-group comparisons at all timepoints. Binary endpoints (proportions of patients who had a response) were compared between treatment groups using the Cochran-Mantel-Haenszel test for general association, controlling for stratification factors. Proportions of patients who had a response in each treatment group and differences in the proportions of patients who had a response between treatment groups were summarised with point estimates. Missing data for continuous variables (including the primary endpoint) were assigned with the last observation carried forward method. Missing data for binary endpoints were handled with the non-responder imputation method. For both continuous and binary endpoints, a predefined secondary analysis was performed using observed cases only. Adherence to treatment was recorded on the patient's diary card and confirmed by recording numbers of study drugs that were dispensed and returned.

All efficacy and safety analyses were done in the full analysis set (ie, all randomised patients who received at least one dose of study drug, which was equal to the intention-to-treat set). Safety analyses were based on actual treatment received. The primary endpoint and selected secondary endpoints (ASAS20 and ASAS40) were additionally analysed in the per-protocol set, which included all patients in the full analysis set who did not experience a major protocol deviation relevant to efficacy. SAS version 9.4 was used for all statistical analyses. The full statistical analysis plan is available in the appendix (pp 139–216). The trial is registered with ClinicalTrials.gov (NCT03117270).

Role of the funding source

The study sponsor supervised study design, study conduct, data collection, statistical analyses, data interpretation, and writing of the manuscript. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.

Results

263 patients were screened for eligibility between March 7, 2017, and July 2, 2018. Of these, 116 were enrolled and randomly assigned to receive filgotinib 200 mg (n=58) or placebo (n=58). 55 (95%) patients in the filgotinib group and 52 (90%) in the placebo group completed the study. Reasons for discontinuation are shown in figure 1. Demographic and baseline disease characteristics were similar between the treatment groups, apart from the mean baseline SPARCC spine score, which was higher in the filgotinib group than in the placebo group (table 1). 56 (97%) patients in the

	Filgotinib (n=58)	Placebo (n=58)
Age (years)	41 (11·6)	42 (9·0)
Sex		
Female	13 (22%)	17 (29%)
Male	45 (78%)	41 (71%)
Weight (kg)	75 (11·9)	77 (18·2)
Body-mass index (kg/m²)	25·3 (3·7)	26·4 (5·2)
Time since diagnosis (years)	6 (5·5)	8 (7·6)
HLA-B27 positivity	51 (88%)	51 (88%)
ASDAS	4·2 (0·6)	4·2 (0·8)
BASDAI	6·9 (1·2)	7·0 (1·3)
BASFI	7·0 (1·5)	6·9 (1·6)
BASMI (linear)	5·1 (1·7)	5·3 (1·6)
High-sensitivity CRP (mg/L)	19·6 (13·3)	21·2 (23·0)
High-sensitivity CRP ≥ULN*	41 (71%)	34 (59%)
MRI SPARCC spine	19·0 (19·7)	13·8 (19·9)
MRI SPARCC sacroiliac joint	6·8 (10·9)	5·3 (6·9)
Enthesitis at baseline†	47 (81%)	48 (83%)
MASES enthesitis	4·9 (2·8)	4·1 (2·9)
csDMARD use	23 (40%)	22 (38%)
Methotrexate	9 (16%)	4 (7%)
Sulfasalazine (oral)	14 (24%)	18 (31%)
NSAID use	43 (74%)	38 (66%)
Steroid use	7 (12%)	10 (17%)
Previous TNF inhibitor therapy	4 (7%)	7 (12%)

Data are mean (SD) or n (%). ASDAS=ankylosing spondylitis disease activity score. BASDAI=Bath ankylosing spondylitis disease activity index. BASFI=Bath ankylosing spondylitis functional index. BASMI=Bath ankylosing spondylitis metrology index. csDMARD=conventional synthetic disease-modifying anti-rheumatic drug. CRP=C-reactive protein. MASES=Maastricht ankylosing spondylitis enthesitis score. NSAID=non-steroidal anti-inflammatory drug. SPARCC=Spondyloarthritis Research Consortium of Canada. TNF=tumour necrosis factor. ULN=upper limit of normal. *The ULN for high-sensitivity CRP is 10 mg/L. †Data are shown for patients with one or more tender enthesis at baseline.

Table 1: **Baseline patient and disease characteristics (full analysis set)**

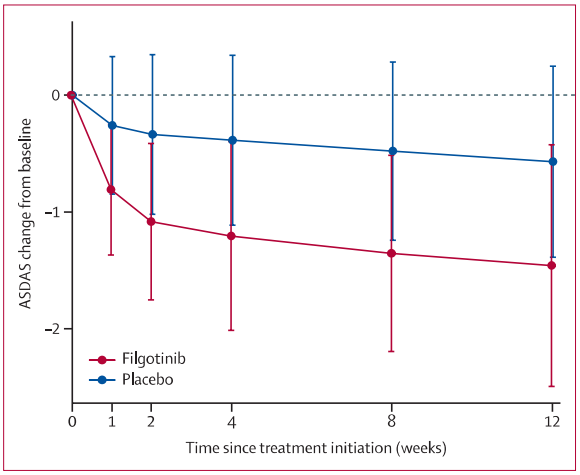

Figure 2: **ASDAS change over time (full analysis set)**
Mean values are shown with SDs. p<0·0001 for the difference between groups at all timepoints. ASDAS=ankylosing spondylitis disease activity score.

filgotinib group and 55 (95%) in the placebo group continued on at least one concomitant medication; the most common concomitant medications were NSAIDs (table 1). Mean on-treatment adherence during the study was 99·3% (SD 5·9) for the filgotinib group and 99·2% (3·5) for the placebo group.

The mean change from baseline to week 12 in ASDAS was –1·47 (SD 1·04) in the filgotinib group and –0·57 (0·82) in the placebo group (figure 2), with a least squares mean difference between groups of –0·85 (95% CI –1·17 to –0·53; p<0·0001; appendix p 15). Analysis of the primary outcome in the per-protocol population confirmed this result: the mean change from baseline to week 12 in the per-protocol population was –1·4 (SE 0·13) in the filgotinib group and –0·5 (0·10) in the placebo group (least squares mean difference –0·88, 95% CI –1·19 to –0·57; p<0·0001).

The difference between groups in the effect on ASDAS was significant as of week 1 (figure 2). A major improvement in ASDAS at week 12 was observed in 19 (33%) of 58 patients in the filgotinib group and in one (2%) of 58 patients in the placebo group (difference 31%, 95% CI 18 to 44; p<0·0001; figure 3). A clinically significant improvement in ASDAS at week 12 was observed in 38 (66%) patients in the filgotinib group compared with 15 (26%) patients in the placebo group (40%, 22 to 54; p<0·0001). Inactive disease at week 12 was achieved in three (5%) patients treated with filgotinib and in no patients treated with placebo (5%, –2 to 14; p=0·092; figure 3). Values for all primary and secondary efficacy endpoints at baseline and week 12 are shown in the appendix (p 15–17).

At week 12, an ASAS20 response was achieved by 44 (76%) of 58 patients assigned to filgotinib and by 23 (40%) of 58 patients assigned to placebo (difference 36%, 95% CI 18 to 51; p<0·0001; figure 3). ASAS40 was achieved by 22 (38%) patients assigned to filgotinib and by 11 (19%) patients assigned to placebo (19%, 3 to 34; p=0·019; figure 3). ASAS5/6 was achieved in 34 (59%) patients in the filgotinib group and in 12 (21%) patients in the placebo group (38%, 20 to 52; p<0·0001; figure 3), and ASAS partial remission in seven (12%) patients in the filgotinib group and in two (3%) patients in the placebo group (9%, –2 to 20; p=0·10; figure 3). Analysis of ASAS20 and ASAS40 in the per-protocol population confirmed these results (appendix p 18).

The mean change from baseline to week 12 in 44 tender joint counts was –2·85 (SD 3·00) in the filgotinib group (n=41) and –1·49 (2·49) in the placebo group (n=47; least squares mean difference –0·79, 95% CI –1·68 to 0·11; p=0·085). The mean change from baseline to week 12 in 44 swollen joint counts was –1·67 (1·88) for the filgotinib group (n=15) and –1·75 (1·65) for the placebo group (n=20; –0·31, –0·76 to 0·15; p=0·18).

At week 12, the BASDAI score had significantly decreased in the filgotinib group compared with the

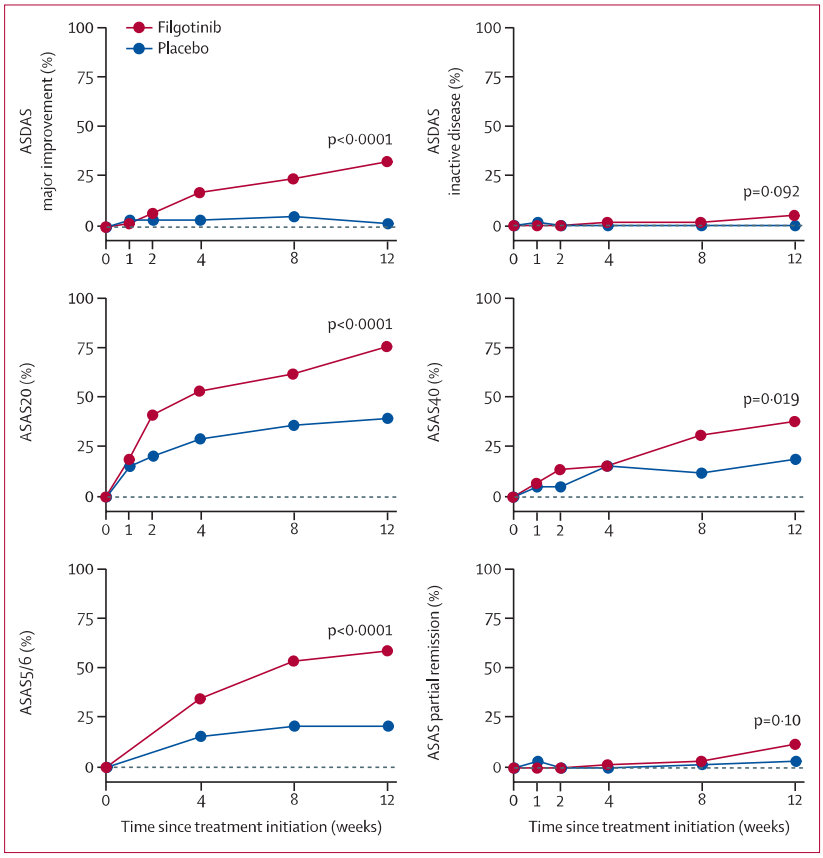

Figure 3: **Proportions of patients reporting major improvements, inactive disease, and fulfilment of ASAS response criteria over time (full analysis set)**
Definitions of ASAS response criteria are in the appendix (p 12). p values for the difference between groups at week 12 are shown; p values for all other timepoints are in the appendix (p 16). ASAS=Assessment of SpondyloArthritis international Society. ASDAS=ankylosing spondylitis disease activity score.

placebo group (mean change from baseline –2·41 [SD 2·01] *vs* –1·44 [2·02]; least squares mean difference –1·00, 95% CI –1·69 to –0·30; p=0·0052), and this difference was significant from week 8 onwards (figure 4). The results for the individual components of the BASDAI are in the appendix (p 15).

There was also a significant improvement in the overall BASFI at week 12 in the filgotinib group compared with the placebo group (–2·45 [SD 1·90] *vs* –1·23 [1·88]; –1·11, 95% CI –1·78 to –0·43; p=0·0015); the difference was significant from week 8 (figure 4). Spinal mobility, as assessed with the BASMI, improved significantly from baseline to week 12 in the filgotinib group compared with the placebo group (–0·75 [1·02] *vs* –0·39 [0·70]; –0·39, –0·68 to –0·10; p=0·0093), and the difference was significant from week 4 onwards (figure 4). SPARCC spine (–5·76 [11·13] *vs* 0·52 [7·47]; –5·69, –9·75 to –1·62; p=0·0066) and SPARCC sacroiliac joint (–3·52 [7·31] *vs* 0·06 [3·51]; –2·33, –4·20 to –0·46; p=0·0150) scores were also significantly decreased in the filgotinib group at week 12 compared with the placebo group (figure 5).

The change from baseline to week 12 in high-sensitivity CRP concentrations was –10·84 mg/L (SD 13·91) in the filgotinib group and –2·24 mg/L (17·35) in the placebo group, with a least squares mean difference between groups of –9·32 mg/L (95% CI –14·01 to –4·62; p<0·0001). The effect of filgotinib on concentration of high-sensitivity CRP was significant compared with placebo at all timepoints (figure 4). The proportion of patients whose high-sensitivity CRP concentration changed from high at baseline to normal at 12 weeks was significantly higher in the filgotinib group than in the placebo group (66% [27/41] *vs* 18% [6/34]; difference 48%, 95% CI 26 to 64; p<0·0001).

At week 12, patients in the filgotinib group also had significantly improved scores on the ASQoL and the physical components of the SF-36 compared with patients in the placebo group (appendix p 15). Mean changes in ASQoL scores were –4·76 (SD 4·50) in the filgotinib group and –2·24 (3·97) in the placebo group, with a least squares mean difference between groups of –2·35 (95% CI –3·92 to –0·77; p=0·0038). The mean change from baseline in the SF-36 physical component score was 8·44 (SD 8·18) for the filgotinib group versus 3·84 (7·10) for the placebo group, with a least squares mean difference between groups of 4·41 (1·88 to 6·93; p=0·0008). The mean change from baseline in the SF-36 mental component score was 3·95 (SD 7·05) for the filgotinib group versus 1·00 (9·83) for the placebo group (least squares mean difference 2·54, 95% CI –0·21 to 5·29; p=0·070).

The proportion of patients who had at least one treatment-emergent adverse event was the same in both groups (31% [18/58] in both; table 2). These events were generally mild or moderate in severity, with only two events reported as grade 3 or higher, both in the filgotinib group (appendix p 19). The most common treatment-emergent adverse event was nasopharyngitis (two patients in the filgotinib group and four in the placebo group; appendix p 19). The one serious treatment-emergent adverse event was a case of grade 3 pneumonia in a woman aged 49 years in the filgotinib group who was a current smoker; she discontinued the study drug and recovered after antibiotic treatment in hospital. The only other treatment-emergent adverse event to lead to permanent discontinuation of study drug, high creatine kinase, was in the placebo group (table 2). There was one other treatment-emergent adverse event of special interest reported: a non-serious, grade 2 deep vein thrombosis in the calf musculature of a man aged 53 years who had a heterozygous factor V Leiden mutation, diagnosed 3 days after the patient's last dose of filgotinib. There were no malignancies (including lymphomas), opportunistic infections, cases of active tuberculosis, extra-articular manifestations (inflammatory bowel disease, psoriasis, or uveitis), or deaths reported in the study. Reports of any infection did not differ significantly between the groups (12% [7/58] of patients in both groups).

Key laboratory parameters monitored in this study are listed in the appendix (p 20). Compared with patients in