Online supplementary file

Methods

Statistical analysis

Analysis populations

Efficacy analyses were conducted in the full analysis set, meeting the requirements of the intention-to-treat principle. The main analysis of the primary endpoint was supplemented by a sensitivity analysis restricted to the per protocol population that comprised all patients in the full analysis set with primary endpoint data available and no major protocol violations. Safety analyses were conducted in the safety population. See supplementary table S1 for details of each analysis population.

Primary outcome

The primary analysis was conducted using a Pearson's Chi-square test. The point estimate and 95% confidence interval for the odds of achieving clinical remission is reported.

A number of sensitivity analyses of the primary outcome were conducted. These consisted of available case analysis, best case (response assumed if missing for reasons other than lack of efficacy) analysis, worst case (non-response assumed if missing for reasons other than lack of efficacy) analysis. We also carried forward the last available on-treatment observation. Additional sensitivity analyses made different assumptions about the missing data in each treatment arm: we assumed all patients with missing data in the early ETN arm would not have achieved remission, while all such patients in the delayed ETN would have achieved remission (designated 'near split' imputation) or vice versa (designated 'far split' imputation). These were designed to test the impacts of the most extreme possible deviations from the missing at random assumption required for multiple imputation. In addition, the primary analysis was repeated in the per protocol set.

Secondary outcomes

Proportions of patients achieving remission, ACR or EULAR response at 96 weeks were compared between groups using Pearson's Chi-squared tests. Changes in continuous variables over time were compared between groups using a linear multilevel modelling approach, mirroring the mixed between-within ANOVA model but with less restrictive assumptions about covariance structure. Optimal covariance pattern was identified using Akaike Information Criterion (AIC) values after inspection of correlations between repeated observations. An autoregressive structure was found to be optimal. Robust standard errors were used to address minor deviations from normality in the

residuals. Following an overall test of a significant difference between the groups over time, comparisons were then made between groups at each time-point. Severely skewed variables were analysed using quantile regression; this was required for the ultrasound variables and total Sharp score which were heavily right-skewed. The analysis plan specified that median regression would be used; however, because of the high prevalence of zero scores for power Doppler and erosion, differences between groups at the 90th percentile have been presented instead.

Time to sustained remission was compared between groups using logrank tests.

The Holm correction (modified Bonferroni) was used on a family-wise basis to control for multiple comparisons of secondary outcomes. This set the critical P-value for testing significance at the 5% level to p<0.00088. The correction included secondary outcomes but excluded exploratory outcomes; by definition, exploratory analyses are considered hypothesis generating rather than confirmatory.

All analyses have been conducted in Stata v15.

Rasch analysis of Patient Reported Outcomes

Quasi-interval-scaled scores were obtained for HAQ-DI, RAQoL and RAWIS using analysis of fit to the Rasch model. For HAQ scores, scores were converted using a published conversion table. For RAQoL and RAWIS, analysis in RUMM2030 was used to test for overall fit to the model, evidence of local dependency, differential item functioning by age, sex and timepoint, multidimensionality and targeting. After adopting a testlet approach, both scales were found to fit the model with no evidence of specific issues of misfit, although targeting was relatively poor at post-baseline visits as many patients' scores were at the minimum possible. Having demonstrated absence of differential item functioning by timepoint, estimates in the full dataset were anchored to the solution obtained at baseline, where targeting was best (and threshold estimates were therefore most accurate).

Other planned analyses

Proportions of patients in each arm requiring escalation of therapy to triple therapy and to biologic therapy have been summarised descriptively, as has cumulative IM steroid dose up to week 48. Comparisons have been made between the response in the early ETN group over the first 24 weeks, and the response in the delayed ETN group over the 24-48 period in those who were escalated to etanercept at week 24.

Unplanned analyses

To illustrate trends in disease activity, mean DAS28-ESR (95% CI) has been plotted over time in each group; mean (SD) DAS28-ESR and change in DAS28-ESR has been tabulated at each of the primary and secondary time-points.

Because all medians were 0 at all time-points in both groups for US power Doppler and erosion scores, quantile regression analysis has been performed in which the 90th percentile is the target rather than the (pre-specified) median (50th percentile).

Area under the curve DAS28-ESR has been calculated during the first 48 weeks and over the whole 96 weeks of the trial. To calculate this, values obtained at 12-weekly intervals were used.

In response to recent interest in sustained remission, and alternative definitions of remission, with respect to predicting successful cessation of biologic therapy, sustained remission (defined as remission at both the current and immediately preceding 12-weekly visit) has been calculated at weeks 48 and 96 for DAS28 remission (<2.6), deep DAS28-ESR remission (<1.98), Boolean remission (tender joint count<=1 & swollen joint count<=1 & CRP<=10 mg/L & patient disease activity VAS <=10mm). For calculation of Boolean remission, 28 joint counts were augmented with assessments of the ankles and MTPs, as recommended by the developers of the criteria. Odds ratios and 95% confidence intervals have been calculated for the odds of achieving DAS28-ESR remission at week 96 according to the level of remission at week 48, in the group as a whole (irrespective of treatment group). The odds of achieving DAS28-ESR remission at week 96 according to DAS28 remission status at week 48 have also been calculated within each treatment arm, and within a subgroup of MTX-TT; MTX-TTb (delayed ETN arm) defined by ETN escalation status at week 24.

Handling of missing data

Missing values at baseline were imputed using screening values, where available.

For response variables (including the primary outcome) patients who discontinued study medication for lack of efficacy were considered non-responders from that point forward.

In all other instances, missing data were addressed using multiple imputation by chained equations. Predictive mean matching (10 nearest neighbours) was used to impute all variables. We imputed 50 datasets and combined the results of our analyses according to Rubin's rules.

Imputation models were as follows:

For all disease activity outcomes that included 28 joint counts:

[SJC28, TJC28, lnCRP, lnESR, physician global VAS, pain VAS, disease activity VAS, early morning stiffness and HAQ-DI] at 0, 12, 24, 36*, 48, 72 & 96 weeks, age at baseline, sex, symptom duration, treatment group and whether or not the patient was escalated to etanercept at week 24

*Excluding HAQ-DI which was not collected at week 36

For all disease activity outcomes that included 44 swollen joint count and RAI:

[SJC44, RAI, lnCRP, lnESR, physician global VAS, pain VAS, disease activity VAS, early morning stiffness and HAQ-DI] at 0, 12, 24, 36*, 48, 72 & 96 weeks, age at baseline, sex, symptom duration, treatment group and whether or not the patient was escalated to etanercept at week 24

*Excluding HAQ-DI which was not collected at week 36

For all patient-reported outcomes:

[HAQ-DI, RAQoL, EQ5D, RAWIS, DAS28-ESR, disease activity VAS, early morning stiffness] at 0, 12, 24, 48, 72 & 96 weeks, age at baseline, sex, symptom duration, whether or not the patient was in paid employment at baseline (relevant for RAWIS), DAS28-ESR, treatment group and whether or not the patient was escalated to etanercept at week 24

For all imaging outcomes:

[lnCRP, early morning stiffness, SJC28, TJC28, lnESR, disease activity VAS, physician global VAS, total GS score, total PD score, total erosion score, total tenosynovitis score, total osteophyte score] at weeks 0, 12, 24 & 48, [lnCRP, SJC28, TJC28, lnESR, disease activity VAS] at week 36, [total modified Sharp score] at weeks 0, 48 & 96, age at baseline, sex, symptom duration, treatment group and whether or not the patient was escalated to etanercept at week 24

Missing visit patterns

Visits that occurred more than 2 weeks from the scheduled date were considered to be missing and data were imputed as detailed above. This was necessary for six visits in six patients. Two were at the primary endpoint 48 weeks, three were at 24 weeks, one was at week 96.

Results

Table S1: Analysis populations, screening criteria and per protocol exclusions

Analysis population	**All**	**ETN+MTX**	**MTX-TT**
Screening population	177	60	60
Full analysis set	120	60	60
Safety population	120	60	60
Per protocol efficacy	84/120 (70%)	45/60 (75%)	39/60 (65%)
In/exclusion criteria met (in screening population)	**Failed (n=57)**	**ETN+MTX (n=60)**	**MTX-TT (n=60)**
IN: Patient aged 18-80	57	60	60
IN: 2010 ACR/EULAR criteria	57	60	60
IN: Symptom onset <12months	54	60	60
IN: Active disease	49	60	60
IN: ACPA or PD positive	56	60	60
IN: No previous DMARDs	57	60	60
IN: Use contraception	57	60	60
EX: Corticosteroid IA/IM	0	0	0
EX: Oral steroid use	1	0	0
EX: NSAID use or change	2	0	0
EX: Imaging	19	0	0
EX: Pregnant/breastfeeding	0	0	0
EX: Other contraindications*	16	1	0
EX: Medical history	7	0	0
EX: Planned surgery	0	0	0
Reason for exclusion from per protocol efficacy population (in full analysis set)	**All (n=120)**	**ETN+MTX (n=60)**	**MTX-TT (n=60)**
Received prohibited concomitant medications	4	0	4
Did not meet inclusion criteria (bar contraception) or met exclusion criteria (bar safety-related criteria)	7	4	3
Study treatment(s) was/were paused for >28 days	1	0	1
Did not receive randomised treatment**	2	0	2
Had >1 visit (up to week 48) outside the 7 day window	3	2	1
Withdrew from study treatment for any reason	16	8	8
Did not have primary endpoint data available	3	1	2

*The patient in ETN+MTX who met exclusion criteria for other contraindications was randomised in error before being discovered to have a positive quantiferon test; they were withdrawn after 1 dose of ETN

**Two patients in the delayed ETN group were not fully escalated to ETN despite being eligible; one declined, another felt unwell after 1 dose and discontinued

Table S2: Primary outcome DAS28-ESR remission (<2.6)

Population: Full analysis set

Visit	**ETN+MTX (n=60)**	**MTX-TT (n=60)**	**Odds ratio (95% CI)**	**Chi-square, P value**
Week 12	39%	17%	3.18 (1.35, 7.50)	Chi-sq=7.0 p=0.008
Week 24	38%	33%	1.25 (0.58, 2.70)	Chi-sq=0.3 p=0.565
Week 48	52%	38%	1.73 (0.81, 3.70)	Chi-sq=2.0 p=0.160
Week 96	47%	42%	1.20 (0.56, 2.60)	Chi-sq=0.2 p=0.641

Proportions estimated following multiple imputation. CI=confidence interval

Table S3: Primary outcome DAS28-ESR remission (<2.6) at 48 weeks - sensitivity analyses

Population: Full analysis set (with exception of per protocol analysis, conducted in per protocol set)

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Best case	63% (38/60)	50% (30/60)	1.73 (0.83, 3.58)	Chi-sq=2.2 p=0.140
Worst case	50% (30/60)	33% (20/60)	2.00 (0.96, 4.18)	Chi-sq=3.4 p=0.063
Split: near	50% (30/60)	50% (30/60)	1.00 (0.49, 2.05)	Chi-sq=0.0 p=1.000
Split: far	63% (38/60)	33% (20/60)	3.45 (1.63, 7.32)	Chi-sq=11.0 p=0.001
LOCF	52% (31/60)	35% (21/60)	1.99 (0.95, 4.13)	Chi-sq=3.4 p=0.065
Complete case	58% (30/52)	40% (20/50)	2.05 (0.93, 4.50)	Chi-sq=3.2 p=0.073
Per protocol	58% (26/45)	44% (17/39)	1.77 (0.74, 4.21)	Chi-sq=1.7 p=0.194

Best case imputation: remission assumed if missing. Worst case imputation: non-remission assumed if missing. Split imputation: if 'near', non-remission assumed for ETN+MTX, remission for MTX-TT. If 'far': remission assumed for ETN+MTX, non-remission for MTX-TT. CI=confidence interval; LOCF=last on-treatment observation carried forward

Table S4: Secondary outcome DAS28-CRP(<2.6)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	53%	32%	2.46 (1.17, 5.20)	
Week 24	54%	48%	1.27 (0.60, 2.67)	
Week 48	75%	64%	1.70 (0.74, 3.87)	
Week 96	61%	59%	1.09 (0.49, 2.41)	Chi-sq=0.0 p=0.837

Proportions estimated following multiple imputation. CI=confidence interval

Table S5: Secondary outcome DAS44-ESR remission (<1.6)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	31%	13%	2.86 (1.13, 7.23)	
Week 24	40%	30%	1.57 (0.72, 3.43)	
Week 48	54%	39%	1.86 (0.87, 4.00)	
Week 96	46%	44%	1.07 (0.50, 2.32)	Chi-sq=0.0 p=0.858

Proportions estimated following multiple imputation. CI=confidence interval

Table S6: Secondary outcome DAS44-CRP remission (<1.6)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	47%	27%	2.46 (1.14, 5.33)	
Week 24	47%	34%	1.74 (0.80, 3.79)	
Week 48	61%	53%	1.38 (0.65, 2.94)	
Week 96	51%	51%	1.03 (0.47, 2.23)	Chi-sq=0.0 p=0.946

Proportions estimated following multiple imputation. CI=confidence interval

Table S7: Table of DAS scores and changes at each visit (unplanned analysis)

Population: Full analysis set

Outcome		Mean score at visit (SD)		Mean change from baseline (SD)	
		ETN+MTX (n=60)	MTX-TT (n=60)	ETN+MTX (n=60)	MTX-TT (n=60)
DAS28-ESR	week 12	3.06 (1.16)	3.82 (1.23)	-2.71 (1.40)	-1.74 (1.36)
	week 24	3.15 (1.50)	3.42 (1.37)	-2.62 (1.75)	-2.14 (1.36)
	week 48	2.69 (1.19)	2.90 (1.07)	-3.08 (1.46)	-2.65 (1.45)
	week 96	2.69 (1.29)	2.92 (1.24)	-3.08 (1.67)	-2.64 (1.72)
DAS28-CRP	week 12	2.64 (1.20)	3.24 (1.24)	-2.57 (1.38)	-1.67 (1.33)
	week 24	2.76 (1.35)	2.96 (1.34)	-2.46 (1.67)	-1.94 (1.37)
	week 48	2.24 (1.16)	2.39 (0.90)	-2.98 (1.47)	-2.52 (1.43)
	week 96	2.45 (1.05)	2.44 (1.07)	-2.76 (1.51)	-2.46 (1.66)
DAS44-ESR	week 12	1.91 (0.80)	2.44 (0.83)	-1.81 (1.00)	-1.21 (0.93)
	week 24	1.99 (1.05)	2.21 (0.93)	-1.73 (1.23)	-1.45 (0.97)
	week 48	1.61 (0.84)	1.87 (0.74)	-2.11 (1.05)	-1.78 (0.98)
	week 96	1.69 (0.93)	1.81 (0.85)	-2.02 (1.14)	-1.84 (1.14)
DAS44-CRP	week 12	1.71 (0.82)	2.17 (0.84)	-1.75 (1.00)	-1.18 (0.92)
	week 24	1.80 (0.98)	2.00 (0.92)	-1.65 (1.20)	-1.34 (0.97)
	week 48	1.40 (0.84)	1.63 (0.69)	-2.06 (1.06)	-1.71 (0.98)
	week 96	1.58 (0.84)	1.59 (0.81)	-1.88 (1.08)	-1.76 (1.15)

Means and SDs estimated following multiple imputation. SD=Standard deviation

Table S8: Proportion in DAS28ESR remission or LDA (DAS28ESR<=3.2) (unplanned analysis)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)
Week 12	57%	37%	2.27 (1.09, 4.75)
Week 24	62%	54%	1.42 (0.66, 3.03)
Week 48	72%	62%	1.59 (0.70, 3.59)
Week 96	65%	61%	1.17 (0.53, 2.59)

CI=confidence interval; ETN=etanercept; MTX=methotrexate; TT=treat-to-target

Table S9: Proportion in ACR/EULAR Boolean remission (unplanned analysis)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)
Week 12	20%	13%	1.59 (0.59, 4.25)
Week 24	17%	12%	1.54 (0.54, 4.40)
Week 48	29%	23%	1.37 (0.59, 3.17)
Week 96	29%	26%	1.14 (0.50, 2.62)

CI=confidence interval; ETN=etanercept; MTX=methotrexate; TT=treat-to-target

Table S10: Secondary outcome time to sustained remission

Population: Full analysis set

Sustained remission	Proportion of patients			Survival time 25th percentile		Logrank chi-square, P value
	All (n=120)	ETN+MTX (n=60)	MTX-TT (n=60)	ETN+MTX (n=60)	MTX-TT (n=60)	
DAS28-ESR	34% (41/120)	42% (25/60)	27% (16/60)	24	36	Chi-sq=4.46 p=0.035
DAS28-CRP	54% (65/120)	57% (34/60)	52% (31/60)	13	24	Chi-sq=2.06 p=0.151
DAS44-ESR	33% (40/120)	38% (23/60)	28% (17/60)	24	36	Chi-sq=2.48 p=0.115
DAS44-CRP	46% (55/120)	48% (29/60)	43% (26/60)	13	24	Chi-sq=1.44 p=0.230
SDAI	26% (31/120)	33% (20/60)	18% (11/60)	24	-	Chi-sq=4.38 p=0.036
CDAI	28% (33/120)	33% (20/60)	22% (13/60)	24	-	Chi-sq=3.07 p=0.080

Note that because fewer than 50% of patients achieved sustained remission for the majority of remission definitions, medians and 75th percentiles could not be calculated for survival time; only the 25th percentiles are presented, where it was possible to calculate these. CRP=C-reactive protein; ESR=erythrocyte sedimentation rate.

Table S11: Secondary outcome EULAR moderate or good response

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	92%	75%	3.59 (1.21, 10.66)	
Week 24	88%	86%	1.18 (0.37, 3.77)	
Week 48	94%	89%	1.88 (0.43, 8.18)	
Week 96	94%	87%	2.39 (0.58, 9.82)	Chi-sq=1.5 p=0.226

Proportions estimated following multiple imputation. CI=confidence interval

Table S12: Secondary outcome EULAR good response

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	55%	33%	2.46 (1.17, 5.17)	
Week 24	61%	49%	1.63 (0.77, 3.46)	
Week 48	70%	61%	1.48 (0.66, 3.31)	
Week 96	63%	61%	1.09 (0.50, 2.39)	Chi-sq=0.0 p=0.832

Proportions estimated following multiple imputation. CI=confidence interval

Table S13: Secondary outcome ACR20 response

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	75%	55%	2.45 (1.10, 5.46)	
Week 24	69%	67%	1.13 (0.50, 2.54)	
Week 48	83%	73%	1.79 (0.69, 4.60)	
Week 96	77%	72%	1.28 (0.52, 3.16)	Chi-sq=0.3 p=0.591

Proportions estimated following multiple imputation. CI=confidence interval

Table S14: Secondary outcome ACR50 response

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	66%	36%	3.46 (1.60, 7.50)	
Week 24	53%	47%	1.29 (0.60, 2.76)	
Week 48	69%	56%	1.78 (0.81, 3.92)	
Week 96	64%	55%	1.45 (0.65, 3.21)	Chi-sq=0.8 p=0.362

Proportions estimated following multiple imputation. CI=confidence interval

Table S15: Secondary outcome ACR70 response

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	36%	21%	2.03 (0.88, 4.67)	
Week 24	42%	28%	1.90 (0.85, 4.22)	
Week 48	58%	44%	1.76 (0.82, 3.75)	
Week 96	50%	42%	1.40 (0.64, 3.07)	Chi-sq=0.7 p=0.397

Proportions estimated following multiple imputation. CI=confidence interval

Table S16: Secondary outcome HAQ-DI score

Population: Full analysis set

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)			
Week 12	-0.60 (-0.70, -0.50)	-0.41 (-0.51, -0.32)	0.19 (0.04, 0.33)	t=2.57, p=0.010	F=1.79, p=0.148
Week 24	-0.59 (-0.71, -0.47)	-0.48 (-0.61, -0.36)	0.11 (-0.07, 0.28)	t=1.19, p=0.233	
Week 48	-0.65 (-0.77, -0.53)	-0.62 (-0.74, -0.50)	0.03 (-0.14, 0.20)	t=0.35, p=0.730	
Week 96	-0.63 (-0.75, -0.51)	-0.66 (-0.79, -0.52)	-0.02 (-0.20, 0.16)	t=-0.26, p=0.797	

Means estimated following multiple imputation. CI=Confidence interval; HAQ-DI=Health assessment questionnaire disability index

Table S17: Secondary outcome HAQ-DI normalisation (HAQ-DI<=0.5)

Population: Full analysis set

Visit	ETN+MTX (n=60)	MTX-TT (n=60)	Odds ratio (95% CI)	Chi-square, P value
Week 12	56%	48%	1.36 (0.66, 2.80)	
Week 24	59%	56%	1.13 (0.53, 2.39)	
Week 48	63%	65%	0.93 (0.43, 2.05)	
Week 96	65%	72%	0.72 (0.31, 1.66)	Chi-sq=0.6 p=0.442

Proportions estimated following multiple imputation. CI=confidence interval

Table S18: Secondary outcome EQ5D-3L

Population: Full analysis set

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)			
Week 12	0.30 (0.24, 0.35)	0.19 (0.11, 0.26)	-0.11 (-0.20, -0.02)	t=-2.46, p=0.014	F=0.40, p=0.751
Week 24	0.32 (0.27, 0.36)	0.22 (0.15, 0.29)	-0.10 (-0.17, -0.02)	t=-2.58, p=0.010	
Week 48	0.33 (0.27, 0.40)	0.27 (0.20, 0.35)	-0.06 (-0.15, 0.03)	t=-1.39, p=0.165	
Week 96	0.36 (0.31, 0.41)	0.29 (0.22, 0.36)	-0.06 (-0.14, 0.02)	t=-1.58, p=0.114	

Means estimated following multiple imputation. CI=confidence interval ; EQ5D-3L=Euroqol health index (5 dimensions 3 levels)

Table S19: Secondary outcome pain visual analogue scale (mm)

Population: Full analysis set

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)			
Week 12	-30.52 (-36.44, -24.60)	-21.51 (-27.82, -15.20)	9.01 (0.30, 17.73)	t=2.03, p=0.043	F=1.76, p=0.152
Week 24	-25.42 (-32.29, -18.55)	-24.18 (-30.49, -17.88)	1.24 (-8.02, 10.50)	t=0.26, p=0.793	
Week 48	-32.77 (-38.03, -27.51)	-32.84 (-38.62, -27.07)	-0.07 (-7.89, 7.75)	t=-0.02, p=0.986	
Week 96	-32.79 (-38.45, -27.13)	-26.47 (-33.66, -19.27)	6.32 (-2.85, 15.50)	t=1.35, p=0.177	

Means estimated following multiple imputation. CI=Confidence interval

Table S20: Secondary outcome disease activity visual analogue scale (mm)

Population: Full analysis set

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)			
Week 12	-31.36 (-37.20, -25.52)	-23.39 (-29.61, -17.17)	7.97 (-0.58, 16.52)	t=1.83, p=0.068	F=0.89, p=0.447
Week 24	-28.92 (-35.15, -22.69)	-25.55 (-32.21, -18.90)	3.37 (-5.80, 12.53)	t=0.72, p=0.472	
Week 48	-35.37 (-40.87, -29.87)	-34.86 (-40.80, -28.92)	0.51 (-7.64, 8.66)	t=0.12, p=0.902	
Week 96	-33.88 (-39.91, -27.85)	-27.67 (-34.63, -20.71)	6.21 (-3.00, 15.42)	t=1.32, p=0.186	

Means estimated following multiple imputation. CI=Confidence interval

Table S21: Secondary outcome Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire

Population: Full analysis set

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)			
Week 12	-7.78 (-9.35, -6.22)	-4.79 (-6.33, -3.25)	2.99 (0.79, 5.19)	t=2.67, p=0.008	F=2.30, p=0.075
Week 24	-7.43 (-9.15, -5.71)	-6.31 (-7.96, -4.65)	1.12 (-1.27, 3.51)	t=0.92, p=0.358	
Week 48	-8.36 (-10.28, -6.43)	-7.90 (-9.77, -6.02)	0.46 (-2.23, 3.15)	t=0.34, p=0.737	
Week 96	-8.21 (-10.13, -6.29)	-8.25 (-10.16, -6.34)	-0.04 (-2.74, 2.66)	t=-0.03, p=0.978	

Means estimated following multiple imputation. CI=Confidence interval

Table S22: Secondary outcome Rheumatoid Arthritis Work Instability Scale (RAWIS)

Population: Full analysis set (subset*)

Visit	Mean change (95% CI)		Difference (95% CI)	T value, P value	Overall F value, P value
	ETN+MTX (n=49)	MTX-TT (n=39)			
Week 12	-7.34 (-9.50, -5.17)	-4.32 (-6.62, -2.01)	3.02 (-0.13, 6.17)	t=1.88, p=0.060	F=1.09, p=0.354
Week 24	-7.05 (-9.39, -4.70)	-5.87 (-8.21, -3.53)	1.17 (-2.11, 4.45)	t=0.70, p=0.483	
Week 48	-7.69 (-10.01, -5.37)	-7.50 (-10.14, -4.86)	0.19 (-3.29, 3.68)	t=0.11, p=0.914	
Week 96	-7.50 (-9.99, -5.00)	-7.47 (-10.34, -4.60)	0.02 (-3.73, 3.77)	t=0.01, p=0.990	

Means estimated following multiple imputation. CI=Confidence interval

*This scale is only validated for use in people who are in paid work therefore this analysis has been restricted to those who reported being in paid work at baseline

Table S23: Secondary outcome total modified Sharp score (hands and feet)

Population: Full analysis set

Visit	% TSS>0		Estimated median change (95% CI)		Difference* (95% CI)	T value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)	ETN+MTX (n=60)	MTX-TT (n=60)		
Week 48	49%	48%	0.20 (-0.34, 0.73)	0.21 (-0.40, 0.81)	0.01 (-0.79, 0.81)	t=0.03, p=0.980
Week 96	54%	57%	0.36 (-0.26, 0.99)	0.60 (-0.11, 1.31)	0.24 (-0.71, 1.19)	t=0.50, p=0.621

Proportions and medians estimated following multiple imputation. CI=Confidence interval; TSS=Total Sharp score

Table S24: Secondary outcome total ultrasound erosions

Population: Full analysis set

Visit	Estimated % E>0		Estimated 90th percentile* (95% CI)		Difference (95% CI)	T value, P value
	ETN+MTX (n=60)	MTX-TT (n=60)	ETN+MTX (n=60)	MTX-TT (n=60)		
Baseline	16%	12%	1.00 (-1.19, 3.19)	1.00 (0.12, 1.88)		
Week 12	14%	9%	1.00 (-0.75, 2.75)	0.00 (-1.32, 1.32)	-1.00 (-2.94, 0.94)	t=-1.02, p=0.310
Week 24	9%	11%	0.10 (-2.18, 2.38)	0.48 (-0.79, 1.75)	0.20 (-2.17, 2.57)	t=0.17, p=0.867
Week 48	11%	13%	0.38 (-1.65, 2.41)	0.78 (-0.83, 2.39)	0.02 (-1.91, 1.95)	t=0.02, p=0.984

Proportions and percentiles estimated following multiple imputation. CI=Confidence interval; E=Erosion

*Median was 0 in both groups at all visits. Unplanned use of 90th percentile instead of median as point of comparison.

Table S25: Secondary outcome cumulative intramuscular steroid dose

Population: Full analysis set

Cumulative steroid dose	All (n=119)	ETN+MTX (n=59)	MTX-TT (n=60)
Up to week 48 (uncorrected)	240.0 (240.0, 240.0), 120-720	240.0 (120.0, 240.0), 120-720	240.0 (240.0, 240.0), 120-600
Up to week 48 (corrected*)	5.0 (5.0, 6.2), 2-18	5.0 (2.6, 7.5), 2-15	5.0 (5.0, 7.4), 2-18

All results presented as median (binomial 95% confidence interval), range. Note that one patient in ETN+MTX was withdrawn after baseline; this patient has been excluded from this analysis.
*Per week of follow-up.

Table S26: Escalation of disease-modifying anti-rheumatic drugs (DMARDs)

Population: Full analysis set

Treatment started	Time interval	ETN+MTX (n=60)	MTX-TT (n=60)
ETN	Before week 12	100% (60)	- (0)
	Weeks 12 to 23	- (0)	- (0)
	Weeks 24 to 47	- (0)	52% (31)
	At/after week 48	- (0)	2% (1)
ADA	At/after week 48	3% (2)	2% (1)
ABA	At/after week 48	2% (1)	- (0)
>=2 concurrent csDMARDs	Before week 12	2% (1)	73% (44)
	Weeks 12 to 23	2% (1)	15% (9)
	Weeks 24 to 47	2% (1)	2% (1)
	At/after week 48	42% (25)	3% (2)
3 concurrent csDMARDs	Before week 12	- (0)	72% (43)
	Weeks 12 to 23	- (0)	13% (8)
	Weeks 24 to 47	2% (1)	2% (1)
	At/after week 48	10% (6)	- (0)

Note that the patients in the ETN+MTX arm who were escalated to double or triple conventional synthetic DMARD therapy prior to week 48 had all been withdrawn from study therapy and were being followed-up observationally. Two were withdrawn from ETN due to AEs and received double therapy after cessation of ETN, one was an ETN non-responder and received triple therapy. ABA=abatacept; ADA=adalimumab; ETN=etanercept

Table S27: Adverse event summary

Population: Safety population

	All (n=120; PY=205.3)	ETN+MTX (n=60; PY=104.2)	MTX-TT (n=60; PY=101.1)
AE			
Total AE, n	946	431	515
Unique AE, n	778	360	418
AE per 100 PY, n	460.8	413.6	509.6
Patients with >=1 AE, n	118	58	60
Discontinuation due to AE, % (n)	7% (8/120)	8% (5/60)	5% (3/60)
AE by category			
Gastrointestinal, n per 100 PY (n)	63.8 (131)	43.2 (45)	85.1 (86)
General, n per 100 PY (n)	43.8 (90)	45.1 (47)	42.5 (43)
Infections, n per 100 PY (n)	163.7 (336)	144.9 (151)	183.1 (185)
Musculoskeletal, n per 100 PY (n)	20.0 (41)	19.2 (20)	20.8 (21)
Nervous system, n per 100 PY (n)	34.1 (70)	29.7 (31)	38.6 (39)
Respiratory, n per 100 PY (n)	29.7 (61)	22.1 (23)	37.6 (38)
Skin, n per 100 PY (n)	29.7 (61)	33.6 (35)	25.7 (26)
Other, n per 100 PY (n)	76.0 (156)	75.8 (79)	76.2 (77)
AE severity*			
Mild, n per 100 PY (n)	264.5 (543)	235.1 (245)	294.9 (298)
Moderate, n per 100 PY (n)	106.7 (219)	101.7 (106)	111.8 (113)
Severe, n per 100 PY (n)	6.3 (13)	5.8 (6)	6.9 (7)
Life-threatening, n per 100 PY (n)	1.5 (3)	2.9 (3)	0 (0)
AE by relation to study drug*			
Not related, n per 100 PY (n)	52.1 (107)	51.8 (54)	52.4 (53)
Unlikely, n per 100 PY (n)	87.2 (179)	72.9 (76)	101.9 (103)
Possible, n per 100 PY (n)	174.9 (359)	155.4 (162)	194.9 (197)
Probable, n per 100 PY (n)	58.0 (119)	56.6 (59)	59.4 (60)
Definite, n per 100 PY (n)	6.3 (13)	8.6 (9)	4.0 (4)
SAE			
Total SAE, n	14	9	5
SAE per 100 PY, n	6.8	8.6	4.9
Patients with >=1 SAE, n	12	7	5

PY=Patient years

*Recurrent AEs counted once at maximum severity/relation reported.

Figure S1: DAS28ESR plotted over time within subgroups defined according to DAS28ESR remission (<2.6) status at week 24

Solid lines indicate ETN+MTX; dashed lines indicate MTX-TT. Green lines indicate those in remission at week 24; red lines indicate those not in remission at week 24. In figure S1b the ETN+MTX arm has also been split according to whether patients had DAS28ESR<2.6 (ETN+MTXa) or DAS28ESR>=2.6 (ETN+MTXb) at 24 weeks.

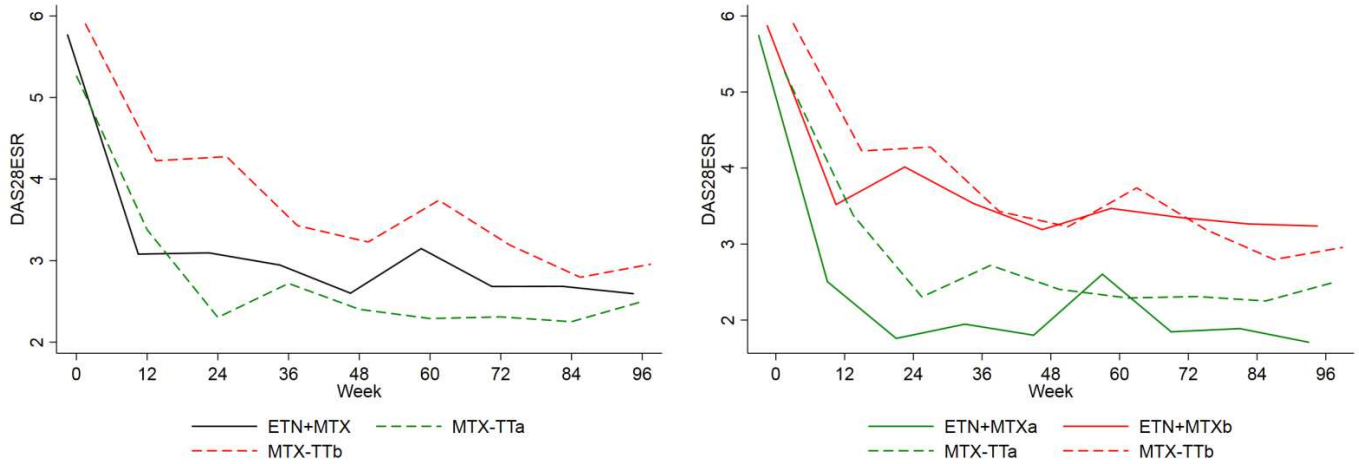

Figure S2: Cumulative probability plot of total van der Heijde-modified Sharp score at baseline and week 48

Population: Full analysis set (available case only)

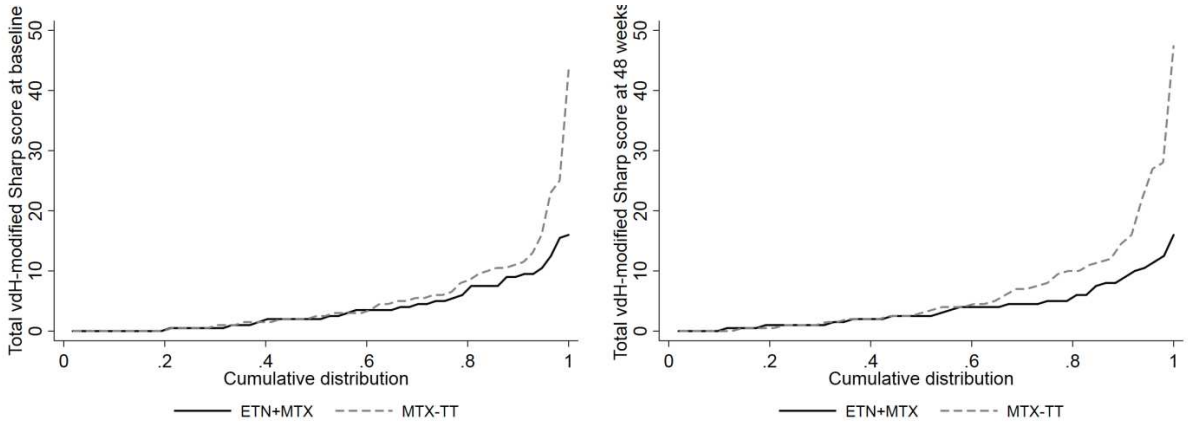

Figure S3: Cumulative probability plot of ultrasound grey scale score at baseline and week 48

Population: Full analysis set (available case only)

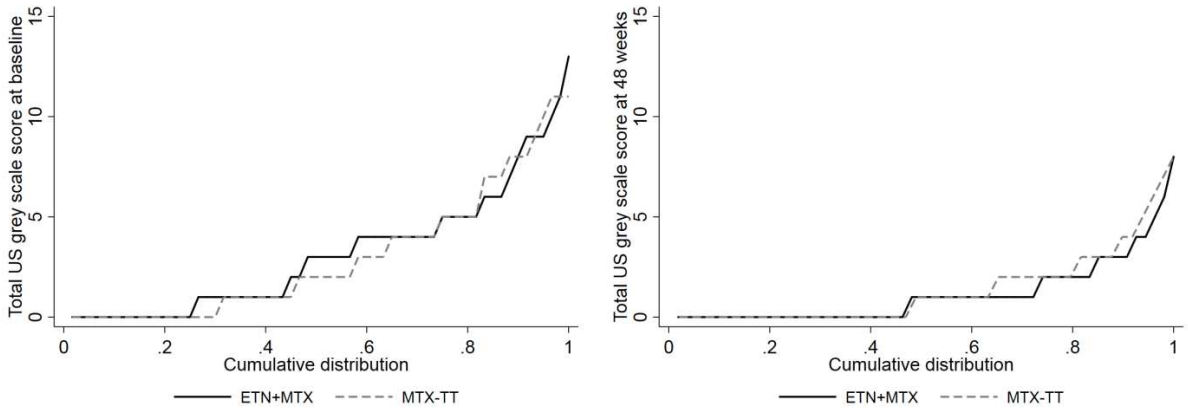

Figure S4: Cumulative probability plot of ultrasound power Doppler score at baseline and week 48

Population: Full analysis set (available case only)

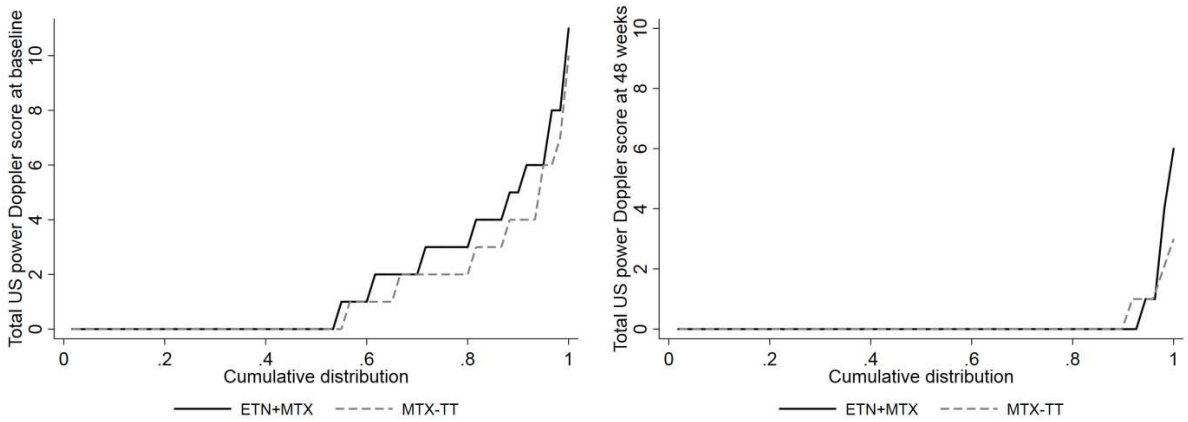

Figure S5: Cumulative probability plot of ultrasound erosion score at baseline and week 48

Population: Full analysis set (available case only)

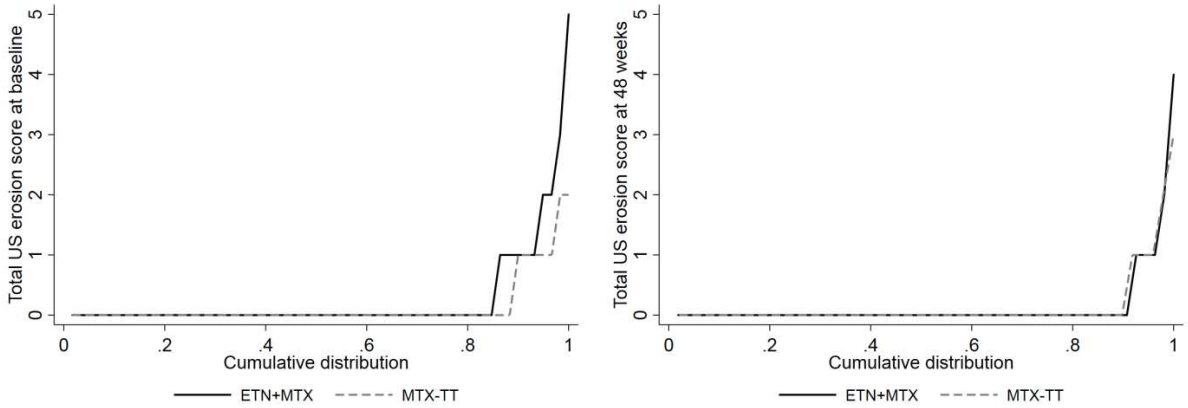

Figure S6: DAS28 components and early morning stiffness plotted over time (unplanned analysis)

Population: Full analysis set (available case only)

Medians in the ETN+MTX and MTX-TT arms plotted over time for a) TJC28 b) SJC28 c) Erythrocyte sedimentation rate (ESR) d) C-reactive protein (CRP) e) patient disease activity visual analogue scale (VASDA) f) early morning stiffness.

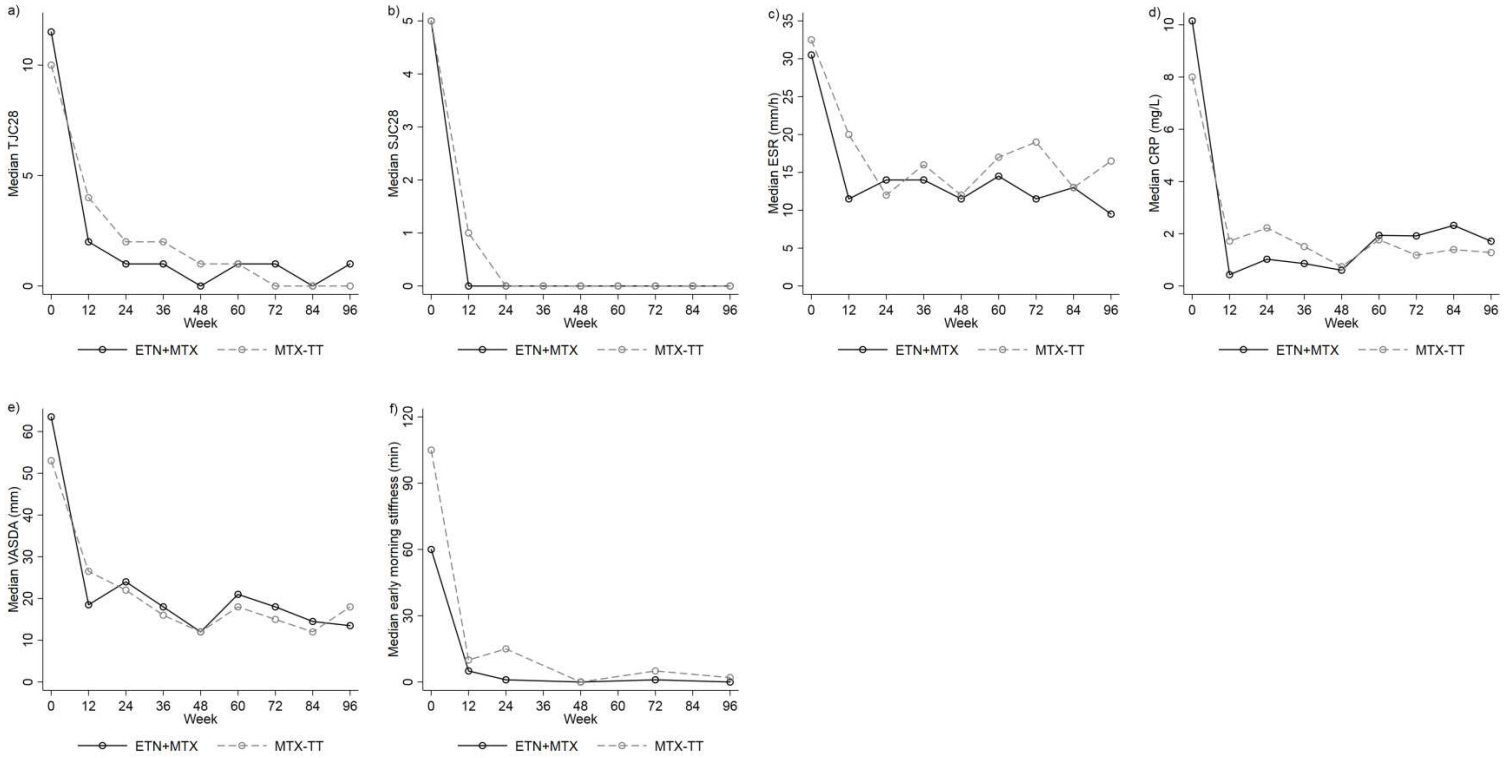

Figure S7: Different DAS28 definitions plotted over time (unplanned analysis)

Population: Full analysis set (available case only)

Medians in the ETN+MTX and MTX-TT arms plotted over time for a) DAS28ESR (objective components; [0.28 x sqrtSJC28] + [0.70 x ln(ESR)]) b) DAS28CRP (objective components; [0.28 x sqrtSJC28] + [0.36 x ln(CRP+1)] + 0.96) c) DAS28 (subjective components; [0.56 x sqrtTJC28] + [0.014 x patient disease activity visual analogue scale]) d) Four component DAS28ESR e) Four component DAS28CRP f) Two component DAS28CRP.