

## **Supplementary Material**

### **Statistical analyses**

#### **Univariate analysis**

The relation between onset of RA-ILD and binary covariates [MTX exposure, gender, baseline: rheumatoid factor (RF, positive/negative), smoking status (ever/never), presence of nodules, extra articular RA major, minor and combined, respiratory major, minor and combined, presence of erosions] was analysed by Chi Squared test, and continuous covariates [age of RA onset, baseline: disease activity score (DAS), ESR, tender joint count (TJC), swollen joint count (SJC), patient global assessment (PGA), Hb, BMI, health assessment questionnaire score (HAQ), number of comorbidities major, minor and combined (including respiratory), number of major comorbidities (excluding respiratory), time from RA symptom onset to first outpatient appointment (months), time from first outpatient appointment to first DMARD (months), time from RA symptom onset to first DMARD (months)] was analysed by t test.

#### **Multivariate analysis model**

Binary logistic models were fitted to the data, using the diagnosis of ILD as the outcome variable to examine possible differences within the variables. The initial covariates included were: MTX exposure (Y/N), gender (M/F), age at onset of RA (years) and presence of rheumatoid factor at baseline (+ve/-ve). Additional covariates were added to the model in turn, to assess whether they improved the model fit. It was not possible to use log likelihood ratio tests (LRTs) to compare models because of varying levels of missing data. Instead, the p-value from each covariate's Wald test was considered, to determine whether there was evidence to include them in the model, as well as assessing whether the odds ratios for other covariates had changed. All covariates significant ( $p < 0.05$ ) and borderline significant ( $p < 0.1$  and  $> 0.05$ ) in the univariate analysis were tested and also some non significant covariates in the univariate analysis were included because they might be of clinical interest (DAS, Hb, TJC,

SJC and PGA, all at baseline). Anti-cyclic citrullinated peptide (anti CCP) antibody data was only collected at a few centres and was insufficient to include in any model.

### **Multivariate time-varying analysis**

Cox proportional hazards models were applied to the data, with the diagnosis of ILD as the binary outcome variable at each year (measured from baseline) that a patient participated in the study. 54 (53.5%) patients of the 101 with ILD had known diagnosis dates or could be approximated to the nearest year. This was up to a maximum of 15 years, to match the time-varying data available for analysis. Of the 54, 48 were diagnosed during the study and so were included within the fitted models. The fixed covariates considered were: gender (M/F), age at onset of RA (years), maximum follow-up (years) and smoking status at baseline (ever/never). Time varying co-variates were DAS, patient global assessment, tender joint count, swollen joint count, ESR, health assessment questionnaire, haemoglobin, BMI, erosions, rheumatoid nodules, major comorbidities. Age at onset of RA, baseline smoking status and maximum follow-up were found to be statistically significant in combination, with all three hazard ratios >1 and all corresponding p-values <0.03. However, maximum follow-up was discounted from the model since adding any time-varying covariates removed its significance. Sex was selected a priori to be in the model to improve precision.

Next, time-varying covariates were considered: MTX exposure (Y/N), DAS (and individually PGA, TJC, SJC and ESR), HAQ, haemoglobin (g/L), presence of rheumatoid factor (RF, +ve/-ve), presence of erosions (Y/N), presence of nodules (Y/N), BMI and number of major comorbidities (excluding respiratory). The model was built by individually adding each covariate, then examining its hazard ratio and corresponding p-value to assess for statistical significance. The components of DAS were assessed both individually and in combination. Covariates were retained if  $p < 0.1$ . Akaike information criterion (AIC) was used to compare the fit of each interim model to those previously fitted, with the final model having the lowest AIC value.

### **Cox regression time to event analysis**

The relation between time of ILD diagnosis after first RA symptoms was explored in the MTX exposed ILD and non-MTX exposed ILD groups using Cox regression time to event analysis, adjusting for the same confounders as in the multivariate model. The time of onset of ILD in relation to first symptoms of RA was taken as the first record of this on the ERAS/ERAN CRF.

Where ILD was recorded on the death certificate but not on the last CRF this time was recorded as last CRF + 1 year if the interval was less than 2 years, and as last CRF + 2 years if this interval was 2 years or longer. In patients with ILD at baseline the time of ILD onset was taken to be synchronous with first RA symptoms.

Supplementary Table 1

ERAS ERAN reasons for discontinuation from follow up

Reason for discontinuation	All ERAS + ERAN		Cohort			
	n	%	ERAS		ERAN	
	n	%	n	%	n	%
Reason for discontinuation	2701	100%	1465	100%	1236	100%
Died	491	18%	360	25%	131	11%
Lost to follow up	370	14%	187	13%	183	15%
Pt Choice	119	4%	70	5%	49	4%
Moved discharged	154	6%	114	8%	40	3%
Comorbidity	28	1%	21	1%	7	1%
Remission	30	1%	28	2%	2	0%
Closure	1509	56%	685	47%	824	67%

**Supplementary Table 2a.**

**Primary univariate analysis showing association of incident RA-ILD with MTX exposure and baseline binary covariates, excluding cases with ILD at baseline prior to any csDMARD use.**

		n	ILD		No ILD		Odds Ratio	CI	Chi-squared test p-value
			n	%	n	%			
<b>Total</b>			67		2600				
<b>MTX</b>	Yes	2667	39	58.2%	1539	59.2%	0.96	(0.57, 1.63)	0.872
	No		28	41.8%	1061	40.8%			
<b>Gender</b>	M	2667	30	44.8%	844	32.5%	1.69	(1.00, 2.83)	0.034
	F		37	55.2%	1756	67.5%			
<b>Rheumatoid Factor</b>	Positive	2588	52	77.6%	1568	60.3%	2.11	(1.16, 4.05)	0.010
	Negative		15	22.4%	953	36.7%			
<b>Anti CCP (ever)</b>	Positive	330	6	9.0%	225	8.7%	1.29	(0.23, 13.31)	0.755
	Negative		2	3.0%	97	3.7%			
<b>Smoker Status</b>	Ever	2106	41	61.2%	1081	41.6%	2.29	(1.25, 4.41)	0.004
	Never		16	23.9%	968	37.2%			
<b>Rheumatoid Nodules</b>	None	2667	57	85.1%	2436	93.7%	2.61	(1.16, 5.27)	0.005
	Nodules		10	14.9%	164	6.3%			
<b>Extra-Articular RA features</b>	Yes	2667	11	16.4%	286	11.0%	1.59	(0.74, 3.11)	0.164
	No		56	83.6%	2314	89.0%			
<b>Respiratory comorbidities</b>	Yes	2667	7	10.4%	134	5.2%	2.15	(0.81, 4.82)	0.056
	No		60	89.6%	2466	94.8%			
<b>Erosions at baseline</b>	Yes	2560	18	26.9%	673	25.9%	0.99	(0.54, 1.75)	0.981
	No		49	73.1%	1820	70.0%			

**Key**

csDMARD: conventional synthetic disease modifying anti-rheumatic drug, MTX: methotrexate, CCP: anti-cyclic citrullinated peptide antibody

Extra-Articular RA features: Tendon sheath disease, Sjogren's syndrome, ocular rheumatoid disease, Raynaud's

Respiratory comorbidities: History of chronic obstructive pulmonary disease, asthma, pneumonia, tuberculosis, pleural disease

Supplementary Table 2b.

Primary univariate analysis showing association of incident RA-LD with baseline continuous covariates, excluding cases with ILD at baseline prior to any DMARD use.

	n	ILD			No ILD			Difference	t-test p-value
		N	Mean	CI	N	Mean	CI		
Age of RA onset (years)	2667	67	61.01	(58.45, 63.58)	2600	55.87	(55.32, 56.42)	5.14	0.004
DAS28	2607	66	4.43	(4.04, 4.82)	2541	4.38	(4.32, 4.45)	0.05	0.821
ESR	2473	63	45.51	(38.76, 52.26)	2410	36.87	(35.77, 37.96)	8.64	0.014
TJC	2656	66	10.74	(7.89, 13.60)	2590	10.14	(9.77, 10.52)	0.6	0.623
SJC	2659	67	11.46	(8.70, 14.22)	2592	12.21	(11.76, 12.67)	-0.75	0.608
PGA	2608	66	48.35	(42.08, 54.62)	2542	43.63	(42.62, 44.64)	4.72	0.146
Hb	2627	67	12.93	(12.55, 13.31)	2560	12.85	(12.79, 12.91)	0.08	0.659
Onset-OPD (months)	2642	67	9.78	(7.96, 11.59)	2575	8.22	(7.98, 8.47)	1.56	0.053
OPD-csDMARD (months)	2357	61	8.13	(2.66, 13.60)	2296	5.57	(4.95, 6.19)	2.56	0.197
Onset-csDMARD (months)	2361	62	17.76	(12.27, 23.25)	2299	14.34	(13.58, 15.09)	3.42	0.151
Comorbidities (major)	2667	67	0.43	(0.30, 0.57)	2600	0.49	(0.46, 0.52)	-0.06	0.532
Comorbidities (minor)	2667	67	0.46	(0.31, 0.61)	2600	0.33	(0.31, 0.35)	0.13	0.053
Comorbidities (combined)	2667	67	0.9	(0.69, 1.10)	2600	0.82	(0.87, 0.86)	0.08	0.542

<b>Comorbidities major excluding respiratory</b>	2667	67	0.33	(0.21, 0.45)	2600	0.47	(0.44, 0.50)	-0.14	0.127
<b>BMI</b>	2356	56	25.98	(24.70, 27.25)	2300	26.56	(26.35, 26.76)	-0.58	0.390
<b>HAQ</b>	2625	67	1.28	(1.10, 1.45)	2558	1.11	(1.08, 1.14)	0.17	0.090

### Key

DAS28: 28 joint disease activity score, ESR: erythrocyte sedimentation rate, TJC: tender joint count, SJC: swollen joint count, PGA: patient global assessment, Hb: haemoglobin, BMI: body mass index, HAQ: health assessment questionnaire, Onset-OPD: time from first RA symptoms to first secondary care outpatient visit, OPD-csDMARD: time from first secondary care outpatient visit to start of conventional synthetic disease modifying anti-rheumatic drug therapy, Onset-csDMARD: time from first RA symptoms to start of conventional synthetic disease modifying anti-rheumatic drug therapy, Co-morbidities major and minor: as per ICD 10 definitions.





Supplementary Table 3a.

Univariate analysis showing association of incident RA-ILD with MTX exposure and baseline binary covariates, extended data set including RA-ILD cases recorded at baseline

		n	ILD		No ILD		Odds Ratio	CI	Chi-squared test p-value
			n	%	n	%			
<b>MTX</b>	Yes	2692	39	2.5%	1539	97.5%	0.51	(0.32, 0.79)	0.001
	No		53	4.8%	1061	95.2%			
<b>Gender</b>	M	2692	44	5.0%	844	95.0%	1.91	(1.23, 2.96)	0.002
	F		48	2.7%	1756	97.3%			
<b>Rheumatoid Factor</b>	Positive	2610	65	4.0%	1568	96.0%	1.65	(1.01, 2.77)	0.038
	Negative		24	2.5%	953	97.5%			
<b>Anti-CCP (ever)</b>	Positive	333	9	3.8%	225	96.2%	1.94	(0.39, 18.74)	0.394
	Negative		2	2.0%	97	98.0%			
<b>Smoker Status</b>	Ever	2692	56	4.9%	1081	95.1%	2.18	(1.31, 3.74)	0.002
	Never		23	2.3%	968	97.7%			
<b>Rheumatoid Nodules</b>	None	2692	79	3.1%	2436	96.9%	2.44	(1.22, 4.54)	0.003
	Nodules		13	7.3%	164	92.7%			

<b>Extra-Articular RA features</b>	Yes	2692	11	3.7%	286	96.3%	1.10	(0.52,2.10)	0.774
	No		81	3.4%	2314	96.6%			
<b>Respiratory co-morbidities</b>	Yes	2692	7	5.0%	134	95.0%	1.52	(0.58, 3.35)	0.299
	No		85	3.3%	2466	96.7%			
<b>Erosions at baseline</b>	Yes	2582	26	3.7%	673	96.3%	1.12	(0.67, 1.81)	0.644
	No		63	3.3%	1820	96.7%			

### Key

MTX: methotrexate, CCP: anti-cyclic citrullinated peptide antibody

Extra-Articular RA features: Tendon sheath disease, Sjogren's syndrome, ocular rheumatoid disease, Raynaud's

Respiratory comorbidities: History of chronic obstructive pulmonary disease, asthma, pneumonia, tuberculosis, pleural disease

Supplementary Table 3b.

Univariate analysis showing association of incident RA-ILD with baseline continuous covariates, extended data set including RA-ILD cases recorded at baseline

	N	ILD			No ILD			Difference	t-test p-value
		n	Mean	CI	N	Mean	CI		
<b>Age RA onset (years)</b>	2692	92	62.8	(60.68, 64.95)	2600	55.87	(55.32, 56.42)	6.93	<0.0001
<b>DAS28</b>	2495	91	4.51	(4.17, 4.85)	2541	4.38	(4.32, 4.45)	0.13	0.445
<b>ESR</b>	2495	85	47.38	(41.33, 53.42)	2410	36.87	(35.77, 37.96)	10.51	0.001
<b>TJC</b>	2681	91	10.92	(8.55, 13.30)	2590	10.14	(9.77, 10.52)	0.78	0.455
<b>SJC</b>	2684	92	12.32	(9.64, 14.99)	2592	12.21	(11.76, 12.67)	0.11	0.935
<b>PGA</b>	2632	90	46.66	(41.35, 51.96)	2542	43.63	(42.62, 44.64)	3.03	0.2786
<b>Hb</b>	2652	92	12.8	(12.48, 13.13)	2560	12.85	(12.79, 12.91)	-0.05	0.798
<b>Onset-OPD (months)</b>	2666	91	9.16	(7.69, 10.64)	2575	8.22	(7.98, 8.47)	0.94	0.174
<b>OPD-csDMARD (months)</b>	2379	83	7.14	(2.93, 11.36)	2296	5.57	(4.95, 6.19)	1.57	0.356
<b>Onset-csDMARD (months)</b>	2383	84	16.49	(12.13, 20.85)	2299	14.34	(13.58, 15.09)	2.15	0.294
<b>Comorbidities major</b>	2692	92	0.48	(0.35, 0.60)	2600	0.49	(0.46, 0.52)	-0.01	0.863
<b>Comorbidities minor</b>	2692	92	0.4	(0.28, 0.52)	2600	0.33	(0.31, 0.35)	0.07	0.212
<b>Comorbidities combined</b>	2692	92	0.8	(0.63, 0.98)	2600	0.82	(0.87, 0.86)	-0.02	0.949

<b>Comorbidities major excluding respiratory</b>	2692	92	0.4	(0.28, 0.52)	2600	0.47	(0.44, 0.50)	-0.07	0.394
<b>BMI baseline</b>	2377	77	25.9	(24.82, 26.98)	2300	26.56	(26.35, 26.76)	-0.66	0.256
<b>HAQ at baseline</b>	2650	92	1.22	(1.07, 1.37)	2558	1.11	(1.08, 1.14)	0.11	0.2024

### Key

DAS28: 28 joint disease activity score, ESR: erythrocyte sedimentation rate, TJC: tender joint count, SJC: swollen joint count, PGA: patient global assessment, Hb: haemoglobin, BMI: body mass index, HAQ: health assessment questionnaire, Onset-OPD: time from first RA symptoms to first secondary care outpatient visit, OPD-csDMARD: time from first secondary care outpatient visit to start of conventional synthetic disease modifying anti-rheumatic drug therapy, Onset-csDMARD: time from first RA symptoms to start of conventional synthetic disease modifying anti-rheumatic drug therapy, Co-morbidities major and minor: as per ICD-10 definitions.

**Supplementary Table 4**

Multivariate analysis stratified by smoking; showing effects of baseline co-variables on incident RA-ILD in smokers, non-smokers and those with missing smoker status at baseline.

	Overall		Non-smokers		Smokers		Missing	
	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
N	2015		949		1066		552	
Methotrexate exposed	0.85 (0.49, 1.49)	0.578	0.24 (0.08, 0.70)	0.009	1.56 (0.74, 3.29)	0.240	1.35 (0.33, 5.52)	0.681
Male gender	1.44 (0.83, 2.48)	0.193	0.70 (0.20, 2.52)	0.587	1.8 (0.95, 3.41)	0.073	1.38 (0.38, 5.02)	0.629
Age of RA onset	1.04 (1.02, 1.06)	<0.001	1.03 (0.99, 1.07)	0.097	1.05 (1.02, 1.08)	0.001	1.03 (0.99, 1.08)	0.172
Baseline RF	2.02 (1.07, 3.82)	0.029	1.88 (0.64, 5.57)	0.254	2.20 (1.00, 4.86)	0.051	2.83 (0.58, 13.85)	0.199
Onset-OPD (months)	1.04 (1.00, 1.07)	0.027	1.005 (0.93, 1.08)	0.902	1.05 (1.01, 1.09)	0.012	0.98 (0.87, 1.10)	0.699
Baseline Major Comorbidities (Excl Resp)	0.62 (0.40, 0.95)	0.027	0.43 (0.15, 1.24)	0.119	0.64 (0.4, 1.04)	0.070	0.30 (0.04, 2.39)	0.254
Baseline Smoker Status	2.21 (1.21, 4.03)	0.010						
Constant	0.0009 (0.0002, 0.005)	<0.001	0.005 (0.0004, 0.07)	<0.001	0.0006 (0.00008, 0.005)	<0.001	0.002 (0.00006, 0.05)	<0.001

**Key**

RA: Rheumatoid arthritis, RF: rheumatoid factor

Supplementary Table 5

Multivariate time varying analysis best fit model showing the association of fixed and time-varying co-variates on incident RA-ILD.

		<b>Hazard Ratio (95% CI)</b>	<b>p</b>
Fixed	Age of RA onset	1.07 (1.02, 1.11)	0.002
	Baseline Smoker Status	1.52 (0.61, 3.79)	0.365
	Gender (male)	1.19 (0.47, 2.99)	0.712
Time- varying	Methotrexate	0.96 (0.82, 1.12)	0.629
	Rheumatoid Factor	1.05 (0.96, 1.15)	0.279
	HAQ	1.15 (1.04, 1.26)	0.007
	ESR	1 (1, 1.01)	0.01
	SJC	0.99 (0.98, 1)	0.058

Legend

HAQ Health Assessment Questionnaire; ESR erythrocyte sedimentation rate; SJC swollen joint count