Appendix - Supplementary information

Statistical methodology

The cox proportional hazards model was chosen to compare survival probabilities between cohorts as the event rates fluctuated over time. Testing the assumptions of proportional hazards by calculating the Schoenfeld residuals revealed non-significant results both for the unadjusted and weighted models.

Adjusting for confounders was performed using an inverse probability of treatment weighting score. A probability of treatment (propensity) score was generated using logistic regression. Covariates chosen for this model 'a priori' were entered into a univariable analysis to examine their individual effects. The results are shown in supplementary table 1.

Supplementary table 1. Univariable adjusted hazard estimates with 95% confidence intervals

Variable	Skin	Shingles
Unadjusted	2.1 (1.5-3.0)	1.9 (1.4-2.7)
Age	2.4 (1.7-3.4)	2.1 (1.6-3.0)
Gender	2.1(1.5-3.0)	1.9 (1.4-2.6)
DAS	2.1 (1.4-3.0)	1.9 (1.3-2.7)
HAQ	1.6 (1.1-2.2)	1.5 (1.1-2.2)
Baseline corticosteroid	2.0 (1.4-2.8)	1.8 (1.3-2.5)
Disease duration	2.0 (1.4-2.8)	1.9 (1.4-2.7)
Smoking	2.1 (1.5-3.0)	1.9 (1.4-2.7)
Diabetes	2.2 (1.6-3.1)	2.0 (1.4-2.7)
COPD	2.2 (1.5-3.0)	1.9 (1.4-2.6)
Entry year into study	1.9 (1.3-2.7)	1.7 (1.2-2.4)

Abbreviations: DAS, Disease Activity Score; HAQ, Health assessment questionnaire; COPD, chronic obstructive pulmonary disease

As several covariates were associated with treatment likelihood in a non-linear pattern, or demonstrated an interaction with other covariates, sequential analyses were performed to identify the components of the model. If a non-linear relationship was identified, orthogonal polynomial transformations of the covariates were added to the model until a suitable fit was obtained. The final propensity score included polynomials for age, DAS28 score and HAQ score; interactions were identified between age and DAS28 score, entry year and

DAS28 score, co-morbidity and DAS28 score, steroid exposure and disease duration, entry year and disease duration, and entry year and steroid exposure. The inverse of the probability (or 1 minus the inverse of the probability in the nbDMARD cohort) was then used as the treatment weight in the analysis. Truncation of weights greater than 20 was used to prevent a small number of larger weights de-stabilising the model. The balancing of the cohorts using the weighted model was tested by comparing standardised differences between cohorts. The weighted means and standardised differences are shown in supplementary table 2. As imbalance remained between covariates using the weighted model alternative models were explored. Propensity scores were used to stratify the cohort into deciles. Analysis using this approach revealed similar estimates of hazard to the IPTW approach. As the marginal model required the use of weights, this IPTW analyses have been presented throughout the manuscript.

Supplementary table 2. Comparison of baseline covariates within weighted cohort

Variable	DMARD	Anti-TNF	Standardised difference
Age (mean)	58.8	57.3	0.122
Gender (% female)	76	74	0.040
DAS	6.37	6.10	0.272
HAQ	2.01	1.93	0.074
Baseline steroid (%)	43	35	0.169
Disease duration	13.1	12.4	0.074
Smoking (% smokers)	22	23	0.023
Diabetes	6.7	7.4	0.028
COPD	5.6	9.3	0.161

Abbreviations: DAS, Disease Activity Score; HAQ, Health assessment questionnaire; COPD, chronic obstructive pulmonary disease

Numbers needed to treat to harm were calculated as the reciprocal of the (failure probability)^{hr} -(failure probability), with failure being either SSSI or shingles, and 'hr' is the respective hazard ratio[19].

Multiple imputation was performed in Stata using the ICE command. Missing data were present in the following variables: age, disease duration, baseline HAQ, and baseline DAS28 score. The imputation model was constructed separately for the nbDMARD and anti-TNF cohorts. Age, gender, disease duration, baseline HAQ, baseline DAS28 score, co-morbidity, smoking status, entry year, and baseline steroid exposure were all included as predictors

within the imputation model. Twenty imputation cycles were performed and the resulting data were analysed using Rubin's rules with the MIM command.

The amount of missing data for each covariate is shown in supplementary table 3.

Supplementary table 3. Proportion of missing data amongst baseline covariates

Variable	DMARD n (%)	Anti-TNF n (%)
DAS	111 (3)	55 (0.5)
HAQ	729 (20)	590 (5)
Disease duration	87 (2)	23 (0.2)
Smoking (% smokers)	18 (0.5)	77 (0.6)
Diabetes	17 (0.5)	64 (0.5)
COPD	20 (0.5)	111 (0.9)

Abbreviations: DAS, Disease Activity Score; HAQ, Health assessment questionnaire; COPD, chronic obstructive pulmonary disease

Supplementary information regarding patients dropping out of follow-up

Supplementary table 4 shows a breakdown of the reasons why patients did not reach 3 years of follow up. A significant difference in rate of dropout was seen firstly because of switching onto or between a biologic agent (nbDMARD cohort 12.9%; anti-TNF cohort 0.7%), and secondly for moving region (DMARD cohort 2.9%; anti-TNF cohort 0.9%). Patients in the anti-TNF cohort who moved region would be followed by their new consultant whereas nbDMARD treated patients may have moved to a region which was not recruiting to the comparison cohort.

Supplementary table 4. Reasons for patients not reaching 3 years of follow up

Reason	nbDMARD n=3673	Anti-TNF n=11881	p value*
Died, n (%)	217 (5.9)	608 (5.1)	0.062
Switched therapy, n (%)	473 (12.9)	77 (0.7)	<0.001
Moved region, n (%)	105 (2.9)	109 (0.9)	<0.001
Withdrew consent, n (%)	34 (0.9)	115 (1.0)	0.818
No reason documented, n (%)	245 (6.7)	787 (6.6)	0.922
Not yet reached 3 years of follow up, n (%)	783 (21)	1301 (11)	<0.001

*p value calculated using χ^2