**Supplementary materials**

for “The impact of smoking on response to TNF inhibitors in axial spondyloarthritis: methodological considerations for longitudinal observational studies”

Theoretical basis for using inverse probability weighting

In most studies, including our own, smoking exposure is simplified into status at baseline (we approximated baseline visit as TNFi initiation). As mentioned in the main text, studying the causal effect of baseline smoking status has conceptual difficulty: we cannot randomly assign an individual to “having smoked for 20 years” at the onset of a hypothetical trial.

To help decide on the analytical approach, it is helpful to consider how causal effects *can* be estimated under ideal conditions. If the full smoking exposure history for each individual were known, it would be possible to “assign” smoking status at each time point – analogous to a sequentially randomised trial. Descriptions of causal inference in observational studies often draw analogy from hypothetical randomised clinical trials. In contrast to conventional trials that are based on a single randomisation, sequentially randomised design allows the study of adaptive treatment strategies that adjust treatment in response to the observed course of disease. Consider the following directed acyclic graph (we omitted other variables for clarity):



Where, $\overbar{A}$(4) = smoking history until time 4

Current smoker: $\overbar{A}$(4) = (a0, a1, a2, a3, 1)

Never smoker: $\overbar{A}$(4) = (0, 0, 0, 0, 0)

Ex-smoker: $\overbar{A}$(4) = (a0, a1, a2, a3, 0) [where a0 + a1 + a2 + a3 ≥ 1 (smoked at some point)]

A(4) = baseline smoking status

L(4) = baseline disease activity

Baseline smoking status, A(4), is confounded by $\overbar{L}$(4) and $\overbar{A}$(3), which are not available in most studies. Decision therefore rests on whether to control for baseline disease activity, L(4).

L(4) is a mediator with respect to $\overbar{A}$(3)

L(4) is a confounder for A(4), and is a time-varying confounder that is affected by past smoking history, $\overbar{A}$(3)

There are thus three options to approximate the effect of $\overbar{A}$(4):

1. Do not account for L(4): entire $\overbar{A}$(4) is confounded.
2. Condition on L(4): A(4) unconfounded; $\overbar{A}$(3) confounded; selection bias since L(4) is a collider.
3. Using IPW to control for L(4) without conditioning: A(4) unconfounded; $\overbar{A}$(3) confounded.

Thus, we chose 3), which may be somewhat less problematic than 1) and 2).

For observational studies of smoking status, the imperfect emulation of a hypothetical clinical trial need not stop us from analysing the available data. The same issue of the exposure trajectory starting before study time 0 – and thus potential for bias – can also be raised for established methods like Mendelian randomisation.

Constructing inverse-probability weights

Stabilised IPW to balance baseline characteristics between smoking status were constructed as follows: the numerator is predicted probability from a multinomial logistic model with smoking (A) as the only variable, and the denominator is the same model conditioned on *a priori* covariates listed in the main text (L).



Participants excluded from the analysis were represented by included participants with the same baseline smoking status and covariates. Stabilised IPCW were constructed as follows: the numerator is the predicted probability from logistic models of not being excluded (C=0) conditioned on smoking status (A), over the same model additionally conditioned on covariates (L). The same approach was used for missing data in analysis 1.



This same modelling approach was used to model missing 3-month responses in analysis 1, and missing responses in the logistic models of BASDAI50/2.

Multiple imputation

To generate the above IPWs, multinomial and logistic models required complete data for all covariates. Multiple imputation was performed using chained equations (-mi impute mice- command in Stata v13). All variables in each IPW model were included in the respective imputation models, with 30 imputed datasets. Logistic (ordinal/multinomial) models were used for categorical variables and predictive mean matching for continuous variables (which accounts for their restricted range).

**Results**

|  |
| --- |
| Supplementary table 1. Baseline characteristics and causes of discontinuation of the 840 patients, according to whether they were eligible for longitudinal analysis or were excluded. |
|  | Included (n=627) | Excluded (n=213) | P-value |
| Smoking status | Never | 234 (37%) | 81 (38%) | 0.980 |
| Ex | 187 (30%) | 63 (30%) |
| Current | 206 (33%) | 69 (32%) |
| Remained on treatment | 480 (77%) | 172 (81%) | 0.540 |
| Stopped treatment | Adverse events | 52 (8%) | 15 (7%) |
| Inefficacy | 43 (7%) | 14 (7%) |
| Other | 52 (8%) | 12 (6%) |
| Age, mean (SD) years | 45.6 (13.9) | 43.8 (12.3) | 0.087 |
| Male | 430 (69%) | 136 (64%) | 0.200 |
| Meets mNY criteria for AS | 389 (62%) | 148 (69%) | 0.051 |
| HLA-B27 positive | 354 (76%) | 115 (77%) | 0.830 |
| Elevated CRP\* | 370 (61%) | 117 (60%) | 0.680 |
| Symptom duration, median (IQR) years | 15.6 (6.4 to 29.1) | 13.2 (5.3 to 23.9) | 0.019 |
| BMI, mean (SD) | 28.0 (5.5) | 27.7 (6.2) | 0.620 |
| Quintiles of Index of Multiple Deprivation | 1, most deprived | 132 (21%) | 35 (16%) | <0.001\*\* |
| 2 | 106 (17%) | 37 (17%) |
| 3 | 118 (19%) | 48 (23%) |
| 4 | 149 (24%) | 49 (23%) |
| 5, most affluent | 122 (19%) | 44 (21%) |
| Highest level of education | Secondary school | 225 (36%) | 65 (31%) | 0.014 |
| Apprenticeship | 62 (10%) | 16 (8%) |
| Further education college | 191 (31%) | 65 (31%) |
| University degree | 112 (18%) | 43 (20%) |
| Further degree | 29 (5%) | 23 (11%) |
| Alcohol status | Current | 456 (73%) | 144 (68%) | 0.380 |
| Ex | 113 (18%) | 46 (22%) |
| Never | 57 (9%) | 22 (10%) |
| Number of comorbidities | 0 | 339 (54%) | 129 (61%) | 0.025\*\* |
| 1 | 183 (29%) | 61 (29%) |
| ≥2 | 101 (16%) | 20 (10%) |
| Data presented as mean (standard deviation), median (interquartile range), number (percentage). Comparisons used t- or Wilcoxon rank-sum tests for continuous variables, Chi-squared test for categorical variables.\*Above upper normal limit.\*\*Non-parametric test for trend across ordered groups.SD, standard deviation; IQR, interquartile range; mNY, modified New York criteria for Ankylosing Spondylitis; BMI, body mass index. |



Supplementary figure 1. Standardised mean differences (SMD) for baseline variables before and after balancing using inverse-probability weights. SMD<0.1 is taken to indicate negligible difference. BASDAI, Bath AS disease activity index; ASDAS, AS disease activity score; BASFI, Bath AS functional index; ASQoL, AS quality of life questionnaire; BASG, Bath AS Global Score; IMD, Index of Multiple Deprivation. Graph produced using R version 3.5.

Analysis 1: comparing response at 3 months according to smoking status



Supplementary figure 2. No statistically significant difference in response to TNF inhibitors at 3 months according to smoking status. Plots show predicted values from weighted generalised estimating equations.

Analysis 2: Comparing response after 6 months in those who remained on treatment

|  |
| --- |
| Supplementary table 2. Baseline characteristics of participants included in analysis 2, according to smoking status. |
|  | Never smoker (n=189) | Ex-smoker (n=149) | Current smoker (n=154) | P-value |
| Age, mean (SD) years | 44.4 (14.9) | 49.5 (13.4) | 42.5 (12.2) | <0.001 |
| Male | 125 (66%) | 98 (66%) | 118 (77%) | 0.060 |
| Meets mNY criteria for AS | 113 (60%) | 97 (65%) | 95 (62%) | 0.600 |
| HLA-B27 positive+ | 109 (76%) | 87 (81%) | 101 (84%) | 0.230 |
| Elevated CRP\* | 109 (59%) | 88 (62%) | 100 (68%) | 0.230 |
| Symptom duration, median (IQR) years | 13.8 (5.4 to 28.9) | 19.2 (9.0 to 32.5) | 13.4 (5.6 to 23.9) | 0.004 |
| BMI, mean (SD) | 27.8 (5.5) | 28.8 (5.3) | 27.5 (5.4) | <0.001 |
| Quintiles of Index of Multiple Deprivation | 1, most deprived | 25 (13%) | 22 (15%) | 43 (28%) | <0.001\*\* |
| 2 | 43 (23%) | 13 (9%) | 28 (18%) |
| 3 | 29 (15%) | 31 (21%) | 30 (19%) |
| 4 | 50 (26%) | 41 (28%) | 29 (19%) |
| 5, most affluent | 42 (22%) | 42 (28%) | 24 (16%) |
| Highest level of education | Secondary school | 58 (31%) | 49 (33%) | 67 (45%) | 0.001 |
| Apprenticeship | 13 (7%) | 14 (10%) | 20 (13%) |
| Further education college | 54 (29%) | 55 (37%) | 43 (29%) |
| University degree | 50 (27%) | 24 (16%) | 16 (11%) |
| Further degree | 13 (7%) | 5 (3%) | 4 (3%) |
| Alcohol status | Current | 149 (79%) | 119 (80%) | 97 (63%) | 0.001 |
| Ex | 21 (11%) | 24 (16%) | 35 (23%) |
| Never | 18 (10%) | 6 (4%) | 22 (14%) |
| Number of comorbidities | 0 | 115 (61%) | 82 (55%) | 82 (53%) | 0.051\*\* |
| 1 | 53 (28%) | 40 (27%) | 50 (32%) |
| ≥2 | 20 (11%) | 26 (18%) | 22 (14%) |
| Disease activity, median (IQR) | BASDAI | 6.3 (4.9 to 7.4) | 6.6 (5.3 to 7.9) | 7.1 (5.6 to 7.8) | 0.013 |
| ASDAS+ | 2.9 (2.3 to 3.4) | 2.9 (2.3 to 3.4) | 3.0 (2.6 to 3.6) | 0.066 |
| Spinal pain | 7.0 (5.0 to 8.0) | 7.0 (5.0 to 8.0) | 7.0 (6.0 to 8.0) | 0.100 |
| BASFI, median (IQR) | 5.7 (4.0 to 7.3) | 6.4 (4.8 to 8.0) | 7.0 (5.4 to 8.4) | <0.001 |
| ASQoL, median (IQR) | 11.0 (7.5 to 14.0) | 13.0 (8.0 to 15.0) | 14.0 (11.0 to 16.0) | <0.001 |
| BASG, median (IQR) | 7.0 (6.0 to 8.0) | 7.0 (5.5 to 8.0) | 7.3 (6.0 to 8.5) | 0.240 |
| Fatigue, median (IQR) | 17.0 (14.0 to 21.0) | 17.0 (13.0 to 21.0) | 18.0 (14.0 to 21.0) | 0.200 |
| Sleep, median (IQR) | 13.0 (8.0 to 17.0) | 14.0 (9.0 to 18.0) | 15.0 (10.0 to 19.0) | 0.045 |
| HADS, median (IQR) | Anxiety | 8.0 (5.0 to 11.0) | 8.0 (5.0 to 11.0) | 11.0 (8.0 to 13.0) | <0.001 |
| Depression | 6.0 (3.0 to 9.0) | 7.0 (4.5 to 9.0) | 9.0 (6.0 to 11.0) | <0.001 |
| Data presented as mean (standard deviation), median (interquartile range), number (percentage). Comparisons used ANOVA or Kruskal–Wallis test for continuous variables, Chi-squared test for categorical variables.+Not all variables had complete data, HLA-B27 status was available for 371 participants, ASDAS for 417.\*Above upper normal limit.\*\*Non-parametric test for trend across ordered groups.SD, standard deviation; IQR, interquartile range; mNY, modified New York criteria for Ankylosing Spondylitis; BMI, body mass index; BASDAI, Bath AS disease activity index; ASDAS, AS disease activity score; BASFI, Bath AS functional index; ASQoL, AS quality of life questionnaire; BASG, Bath AS Global Score; HADS, Hospital Anxiety and Depression Scale. |



Supplementary figure 3. Response to TNF inhibitors after 6 months according to smoking status. Plots show predicted values from weighted generalised estimating equations.

|  |
| --- |
| Supplementary table 3. Descriptions of stabilised inverse-probability weights used in analyses.  |
|  | Mean (SD) | Range |
| Analysis 1 | IPTW to balance baseline covariates between smoking status | 0.97 (0.63) | 0.35 to 6.94 |
| IPCW for excluded | 1.00 (0.14) | 0.79 to 2.09 |
| IPCW for missing 3-month response | 1.00 (0.16) | 0.56 to 2.40 |
| Analysis 2 | IPTW to balance baseline covariates between smoking status | 0.96 (0.58) | 0.29 to 7.59 |
| IPCW for excluded | 0.99 (0.18) | 0.73 to 1.81 |
| Logistic model | missing 3-month response | 0.96 (0.63) | 0.30 to 4.76 |
| missing 6-month response | 0.99 (0.28) | 0.26 to 5.35 |
| missing 12-month response | 0.99 (0.28) | 0.25 to 5.94 |
| IPCW, IP censoring weights; SD standard deviation.  |