

Original article

Depression and anxiety symptoms at TNF inhibitor initiation are associated with impaired treatment response in axial spondyloarthritis

Sizheng Steven Zhao ^{1,2}, Gareth T. Jones ³, David M. Hughes⁴, Robert J. Moots ^{2,5} and Nicola J Goodson²

Abstract

Objectives. Depression and anxiety are associated with more severe disease in cross-sectional studies of axial spondyloarthritis (axSpA). We examined the association between baseline symptoms of depression or anxiety and response to TNF inhibitors (TNFi) in axSpA.

Methods. Biologic naïve participants from a national axSpA register completed the Hospital Anxiety and Depression Scale (HADS) before initiating TNFi. Symptoms of anxiety and depression were each categorized as moderate–severe (≥ 11), mild (8–10) and ‘none’ (≤ 7), and compared against change in disease indices [BASDAI and AS Disease Activity Score (ASDAS)] over time and time to treatment discontinuation using marginal structural models. Inverse-probability weights balanced baseline age, gender, BMI, deprivation, education and baseline values of respective disease indices.

Results. Of the 742 participants (67% male, mean age 45 years), 176 (24%) had moderate–severe and 26% mild depression; 295 (40%) had moderate–severe and 23% mild anxiety. Baseline disease activity was higher in higher HADS symptom categories for both depression and anxiety. Participants with moderate–severe depression had significantly poorer response compared with those with ‘none’ throughout follow-up. At 6 months, the difference was approximately 2.2 BASDAI and 0.8 ASDAS units after balancing their baseline values. Equivalent comparisons for anxiety were 1.7 BASDAI and 0.7 ASDAS units. Treatment discontinuation was 1.59-fold higher (hazard ratio 95% CI: 1.12, 2.26) in participants with moderate–severe anxiety compared with ‘none’.

Conclusions. Symptoms of depression and anxiety at TNFi initiation are associated with poorer treatment outcomes. Targeted interventions to optimize mental health have potential to substantially improve treatment response and persistence.

Key words: axial spondyloarthritis, depression, anxiety, treatment response, mental health

Rheumatology key messages

- axSpA patients with moderate–severe depression have significantly poorer response to TNFi compared with those with none.
- Moderate–severe anxiety is associated with 59% increased treatment discontinuation than those with none.
- Seventy-one per cent of participants with moderate–severe depressive symptoms did not have a documented diagnosis of depression.

¹Musculoskeletal Biology, Institute of Life Course and Medical Sciences, University of Liverpool, ²Department of Rheumatology, Liverpool University Hospitals, Liverpool, ³Aberdeen Centre for Arthritis and Musculoskeletal Health (Epidemiology Group), School of Medicine Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, ⁴Department of Health Data Science, Institute of Population Health, University of Liverpool, Liverpool and ⁵Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, UK

Submitted 11 January 2021; accepted 23 February 2021

Introduction

Axial spondyloarthritis is characterized by severe inflammatory back pain and functional impairment. Symptom onset is commonly in early adulthood, which can be a

Correspondence to: Nicola J Goodson, Department of Rheumatology, Liverpool University Hospitals, Liverpool L9 7AL, UK.
E-mail: ngoodson@liverpool.ac.uk

critical time for education, career and relationships. The consequence of these disruptive symptoms on mental health is compounded by the often-significant delays to diagnosis and treatment [1]. The burden of mental health comorbidities is high, for example depression prevalence ranges from 15% to 38% depending on the screening tool and population [2].

Mental health disorders are well-known to influence the experience and reporting of symptoms [3]. This is particularly relevant for assessment of axSpA disease activity, since indices are mostly [e.g. AS Disease Activity Score (ASDAS)] or entirely (BASDAI and spinal pain) subjective. Prior studies have shown depression to be consistently and independently associated with disease activity and other indices [2], yet none have examined whether they influence longitudinal treatment outcomes. Unlike many other chronic comorbidities, symptoms of depression and anxiety are potentially modifiable, for example by pharmacological or talking therapies [4]. Finding modifiable factors to improve TNFi response is important as suboptimal response is observed in up to half of treated patients [5], and the number of therapeutic options remains relatively limited compared with rheumatoid arthritis. Estimating the potential impact of mental health interventions on TNFi response will inform future clinical trials or management guidelines.

Earlier exploratory analysis of the British Society for Rheumatology Biologics Register for Ankylosing Spondylitis (BSRBR-AS) identified, among others factors, poorer mental health as a predictor of TNFi response [6]. Macfarlane *et al.* used stepwise selection of predictors to show that, for each unit increase in the mental component summary of the Short Form Health Survey-12 (SF-12), odds of ASDAS clinically important response (reduction by ≥ 1.1) and ASDAS low disease activity (< 2.1) were significantly increased by 5%. Predictors do not necessarily have a causal interpretation (e.g. stepwise variable selection provides final models that may omit important confounders [7, 8]), while SF-12 is a quality of life instrument not validated to assess depression or anxiety. Nevertheless, these findings suggested a need for more detailed analysis, including a wider range of treatment response definitions. Thus, we sought to examine the association between baseline symptoms of depression or anxiety—using the Hospital Anxiety and Depression Scale (HADS)—on response to TNFi. Specifically, we aimed to estimate the potential benefits to treatment response—in terms of absolute change in disease indices, binary response criteria, and treatment discontinuation—if it were possible to reduce symptoms of depression and anxiety at or before TNFi initiation.

Methods

Patient population

The British Society for Rheumatology Biologics Register for Ankylosing Spondylitis (BSRBR-AS) is a UK-wide

prospective cohort study that recruited biologics-naïve patients fulfilling the ASAS criteria for axial SpA between December 2012 and December 2017 [9]. Biologics-naïve participants who started their first TNFi were eligible for this analysis. They were followed up at baseline, 3, 6 and 12 months and annually thereafter. Eligible participants were required to have a baseline questionnaire, including depression and anxiety symptoms, dated within a window from 1 year before to 7 days after the TNFi start date. This analysis used the study dataset of December 2018. Ethical approval was obtained from the National Research Ethics Service Committee (reference 11/NE/0374) and written informed consent was obtained from all participants.

Depression and anxiety

Symptoms of depression and anxiety were measured using the Hospital Anxiety and Depression Scale (HADS). HADS comprises 14 questions, seven each for the anxiety and depression subscores, to give each a score ranging of 0 (none) to 21 (indicating severe symptoms). Its original description used each subscale to describe 'case-ness' with 0–7 indicating 'non-cases', 8–10 'doubtful cases', and ≥ 11 'cases' [10, 11]. HADS depression ≥ 8 was reported to have a sensitivity of 82% and specificity of 74% for major depressive disorder [12]. For HADS anxiety ≥ 8 , sensitivity was reported as 90% and specificity 78% [13]. To categorize symptom levels, we used ≥ 11 to indicate moderate to severe symptoms and 8–10 mild. Snaitch described the 0–7 category as 'being in the normal range' [10]. For convenience, this category was referred to as 'none'.

Outcomes

Response to TNFi was assessed using three complementary types of outcomes. First, we studied change in (continuous) disease indices over follow-up time. Disease activity was assessed using BASDAI, ASDAS and the spinal pain numerical rating scale; other aspects of disease severity and life impact were measured using BASFI, AS quality of life questionnaire (ASQoL, which has a range of 0–18 with higher scores indicating poorer quality of life) and the Chalder Fatigue Scale, which has a range of 0–33 with higher scores indicating greater levels of fatigue [14]. To facilitate comparison of these indices with different ranges, we also standardized all to a 0–10 scale [i.e. (observed – minimum)/(maximum – minimum) $\times 10$]. Change in disease indices was assessed over the first 3 years because few participants had longer follow-up.

We also examined common binary response definitions at 6 months: BASDAI50/2 (50% or 2-unit reduction), ASDAS major improvement (ASDAS-MI, ≥ 2 -unit reduction), and two 'low disease activity' states (BASDAI < 4 and ASDAS < 2.1). Individuals who remained on drug but had a missing 6-month assessment were considered as responders if they demonstrated response at 3 or 12 months (participants were unlikely to remain on drug if they did not have or lost

response, as per UK prescribing guidelines [15]. Participants who discontinued treatment within 6 months for any reason were considered non-responders.

Lastly, we examined time to treatment discontinuation. There were insufficient data to allow examination of cause-specific discontinuation. Censoring was defined by the last study contact (visit or questionnaire) for those who did not discontinue.

Covariates

Covariates were determined *a priori* and supported by direct acyclic graphs, including: age, gender, BMI, socioeconomic status (Index of Multiple Deprivation [16] as a continuous variable) and educational attainment (as dummy variables). Baseline values of each disease index were included in respective models.

Statistics

Descriptive statistics were used to compare participant characteristics according to HADS categories. To estimate group (rather than individual) level effects analogous to those given by randomized trials, we used marginal methods throughout [i.e. using inverse probability (IP) weights [17] rather than conditioning on covariates]. The HADS ≤ 7 ('none') group was used as the reference in all comparisons.

IP weighted generalized estimating equations were used to assess for absolute change in disease activity. The model included each disease index in turn as the dependent variable, and HADS-D category, time and their interaction as the independent variable. Time was modelled as linear splines with knots at 3 and 6 months based on known response trajectories (i.e. improvement mostly occurs in the first 6 months then plateaus thereafter). IP weighted logistic regression was used for binary response definitions. IP weighted pooled logistic regression (i.e. marginal structural Cox models) was used for time-to-treatment discontinuation, with time modelled in quadratic form. We further relaxed proportional hazards assumptions with time-by-exposure interactions. Each of these models was then repeated for HADS-A categories. Missing follow-up data (proportion shown in [Supplementary Data](#), available at *Rheumatology* online) were not imputed.

IP weights balance all covariates in the weighted model to allow unconfounded descriptive comparisons [12]. The numerator was the predicted probability from a multinomial logistic model with three-level categorical HADS as the only variable, and the denominator was the same model conditioned on all covariates. Additional details on derivation of IP weights are shown in [Supplementary Data](#), available at *Rheumatology* online.

Sensitivity analyses

Participants with baseline BASDAI < 4 would not ordinarily be eligible for TNFi according to UK prescribing guidance. These individuals were excluded in the first set of

sensitivity analyses. Using the above definitions of binary response, individuals who stayed on treatment but did not have assessments recorded at 3, 6 or 12 months would have missing response values. In the second sensitivity analysis, these individuals were assumed to have responded at 6 months if they remained on treatment beyond 1 year. Third, adequate overlap in participant characteristics between HADS groups is required for valid causal inference, for which we used a weighting analogue of propensity score matching [18].

Results

Among 2687 participants in the BSRBR-AS, 1145 started on biologics; 742 completed the HADS and were eligible for this analysis (flow chart in [Supplementary Fig. S1](#), available at *Rheumatology* online). TNFi initiators included and excluded from the analysis set were similar in characteristics except the former were older (45 vs 43 years) (see [Supplementary Table S1](#)). The analysis cohort was predominantly (67%) male with a mean age of 45 years.

Participants in each category of depressive symptoms were similar in age and gender ([Table 1](#)). Those with no depression had lower BMI and higher educational attainment than mild to severe. Anxiety severity was associated with younger mean age, but not BMI or education ([Table 2](#)). More severe symptoms of depression and anxiety were each associated with greater deprivation, disease activity, fatigue and impairment to function and quality of life.

Only 29% of those with moderate–severe depressive symptoms had a documented depression diagnosis ([Table 1](#)). HADS depression and anxiety subscores were highly correlated ([Supplementary Fig. S2](#) and [Table S2](#), available at *Rheumatology* online).

Absolute improvement in continuous outcomes

All baseline covariates were adequately balanced ([Supplementary Fig. S3](#), available at *Rheumatology* online); IP weights are described in [Supplementary Table S3](#). The number of individuals included for analysis ranged from 673 to 741 depending on missing outcome data ([Supplementary Table S4](#)). Those with 'no' depressive symptoms had superior TNFi response across all six indices, compared with mild or moderate–severe groups. For example, those with 'no' depressive symptoms had approximately 2.2 units greater response in BASDAI, 0.8 units in ASDAS, and 2.3 units in spinal pain at 6 months than the moderate–severe depression group ([Fig. 1](#)). This persisted throughout follow-up. Improvement in quality of life was slow, unlike other indices that showed sharp improvements after TNFi initiation.

Results were similar for anxiety categories ([Fig. 2](#)). At 6 months, response in moderate–severe and 'none' groups differed by 1.7 BASDAI units, 0.7 ASDAS and 1.7 spinal pain (full model coefficients and predicted values

TABLE 1 Baseline characteristics compared according to depression symptom categories

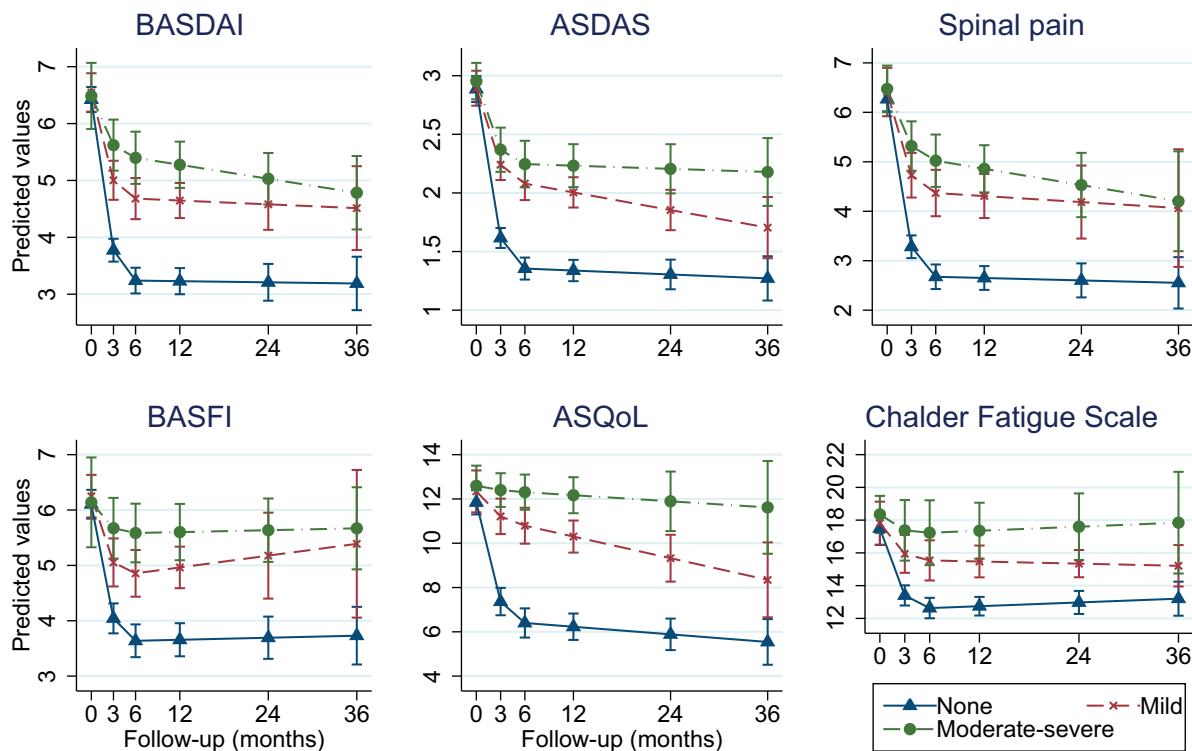
	'None' (n = 374)	Mild (n = 192)	Moderate–severe (n = 176)	P-value
Age, mean (s.d.), years	45.3 (14.6)	45.5 (13.6)	45.5 (12.8)	0.97
Males, n (%)	248 (66)	133 (69)	115 (65)	0.69
Meeting modified New York criteria, n (%)	221 (59)	118 (61)	96 (55)	0.40
Age at symptom onset, mean (s.d.), years	28.0 (10.7)	29.5 (12.2)	29.4 (12.3)	0.21
Symptom duration, mean (s.d.), years	17.3 (13.4)	16.0 (12.5)	16.1 (13.1)	0.43
HLA-B27 positive, n (%)	218 (78)	106 (75)	90 (70)	0.23
BMI, mean (s.d.), kg/m ²	27.2 (5.2)	29.0 (6.1)	28.8 (6.5)	0.002
Education, n (%)				
Secondary school	107 (29)	71 (37)	22 (50)	0.006
Apprenticeship	36 (10)	16 (8)	4 (9)	
Further education college	110 (29)	62 (33)	15 (34)	
University degree	93 (25)	34 (18)	2 (5)	
Further degree	28 (7)	7 (4)	1 (2)	
IMD, mean (s.d.)	2.9 (1.4)	3.1 (1.4)	3.3 (1.4)	0.004
NSAID use in past 6 months, n (%)	277 (75)	138 (74)	123 (71)	0.62
DMARD use in past 6 months, n (%)	44 (12)	25 (14)	38 (22)	0.009
ASDAS, mean (s.d.)	2.6 (0.8)	3.1 (0.7)	3.3 (0.7)	<0.001
BASDAI, median (IQR)	6.0 (4.5, 7.2)	7.0 (6.0, 7.8)	7.7 (6.9, 8.9)	<0.001
Spinal pain, median (IQR)	6.0 (4.0, 7.0)	7.0 (6.0, 8.0)	8.0 (6.0, 9.0)	<0.001
BASFI, median (IQR)	5.4 (3.3, 7.0)	6.9 (5.4, 8.3)	8.1 (6.8, 9.1)	<0.001
ASQoL, median (IQR)	10.0 (7.0, 13.0)	14.0 (12.0, 16.0)	16.0 (14.0, 17.0)	<0.001
Fatigue ^a , median (IQR)	15.0 (12.0, 19.0)	18.0 (16.0, 22.0)	22.0 (18.0, 26.5)	<0.001
History of physician diagnosed depression, n (%)	31 (8)	31 (16)	51 (29)	<0.001

^aThe Chalder Fatigue Scale ranges from 0 (low) to 33 (high). ASDAS: AS disease activity score; ASQoL: AS quality of life questionnaire; BASDAI: Bath AS disease activity index; BASFI: Bath AS functional index; IMD: index of multiple deprivation with 1 (least deprived) to 5 (most); IQR: interquartile range.

TABLE 2 Baseline characteristics compared according to anxiety symptom categories

	None (n = 278)	Mild (n = 169)	Moderate–severe (n = 295)	P-value
Age, mean (s.d.), years	47.0 (14.7)	45.5 (13.9)	43.8 (13.0)	0.022
Males, n (%)	188 (68)	120 (71)	188 (64)	0.27
Meeting modified New York criteria, n (%)	165 (59)	106 (63)	164 (56)	0.32
Age at symptom onset, mean (s.d.), years	28.3 (11.6)	29.3 (11.1)	28.8 (11.8)	0.66
Symptom duration, mean (s.d.), years	18.7 (13.4)	16.2 (13.1)	15.0 (12.5)	0.003
HLA-B27 positive, n (%)	174 (81)	83 (70)	157 (73)	0.047
BMI, mean (s.d.), kg/m ²	27.8 (5.5)	27.8 (5.1)	28.5 (6.4)	0.39
Education, n (%)				
Secondary school	89 (32)	60 (36)	102 (35)	0.067
Apprenticeship	29 (10)	14 (8)	27 (9)	
Further education college	81 (29)	42 (25)	104 (36)	
University degree	62 (22)	45 (27)	42 (14)	
Further degree	16 (6)	8 (5)	15 (5)	
IMD, mean (s.d.)	2.8 (1.4)	3.0 (1.3)	3.2 (1.4)	0.011
NSAID use in past 6 months, n (%)	205 (75)	125 (75)	208 (72)	0.56
DMARD use in past 6 months, n (%)	43 (16)	18 (11)	46 (16)	0.30
ASDAS, mean (s.d.)	2.6 (0.9)	2.9 (0.7)	3.1 (0.8)	<0.001
BASDAI, median (IQR)	5.9 (4.4, 7.2)	6.6 (5.5, 7.4)	7.4 (6.2, 8.5)	<0.001
Spinal pain, median (IQR)	6.0 (4.0, 8.0)	7.0 (5.0, 8.0)	7.0 (6.0, 8.0)	<0.001
BASFI, median (IQR)	5.6 (3.3, 7.1)	6.3 (4.6, 8.1)	7.2 (5.7, 8.6)	<0.001
ASQoL, median (IQR)	10.0 (6.0, 13.0)	12.0 (10.0, 15.0)	15.0 (12.0, 17.0)	<0.001
Fatigue ^a , median (IQR)	15.0 (12.0, 19.0)	17.0 (14.0, 21.0)	20.0 (17.0, 24.0)	<0.001

^aThe Chalder Fatigue Scale ranges from 0 (low) to 33 (high). ASDAS: AS disease activity score; ASQoL: AS quality of life questionnaire; BASDAI: Bath AS disease activity index; BASFI: Bath AS functional index; IMD: index of multiple deprivation with 1 (least deprived) to 5 (most); IQR: interquartile range.

Fig. 1 Change in disease indices according to baseline depression symptom categories

Full model output and marginal predictions are shown in [Supplementary Tables S4 and S5](#), available at *Rheumatology* online. ASDAS: AS disease activity score; ASQoL: AS quality of life questionnaire; BASDAI: Bath AS disease activity index; BASFI: Bath AS functional index.

are shown in online [Supplementary Tables S4–S7](#), available at *Rheumatology* online). All effect sizes were numerically smaller than the above comparisons for depression. For both depression and anxiety, difference between 'none' and moderate-severe groups was largest for ASQoL (standardized scales shown in [Supplementary Figs S4 and S5](#)).

Binary treatment response

The number of individuals included for analysis was 542 for BASDAI-based and 492 for ASDAS-based outcomes. BASDAI50/2 was achieved by 304 (56%), BASDAI < 4 by 308 (57%), ASDAS-MI by 129 (26%) and ASDAS < 2.1 by 167 (34%). Odds of achieving binary response reduced with increasing severity of baseline depression symptoms ([Fig. 3](#)). Compared with those with 'no' depression, participants with moderate-severe symptoms had around half the odds of achieving response at 6 months after accounting for all covariates including differences in baseline BASDAS or ASDAS. Groups with mild depression had 39–48% lower odds of response.

Treatment discontinuation

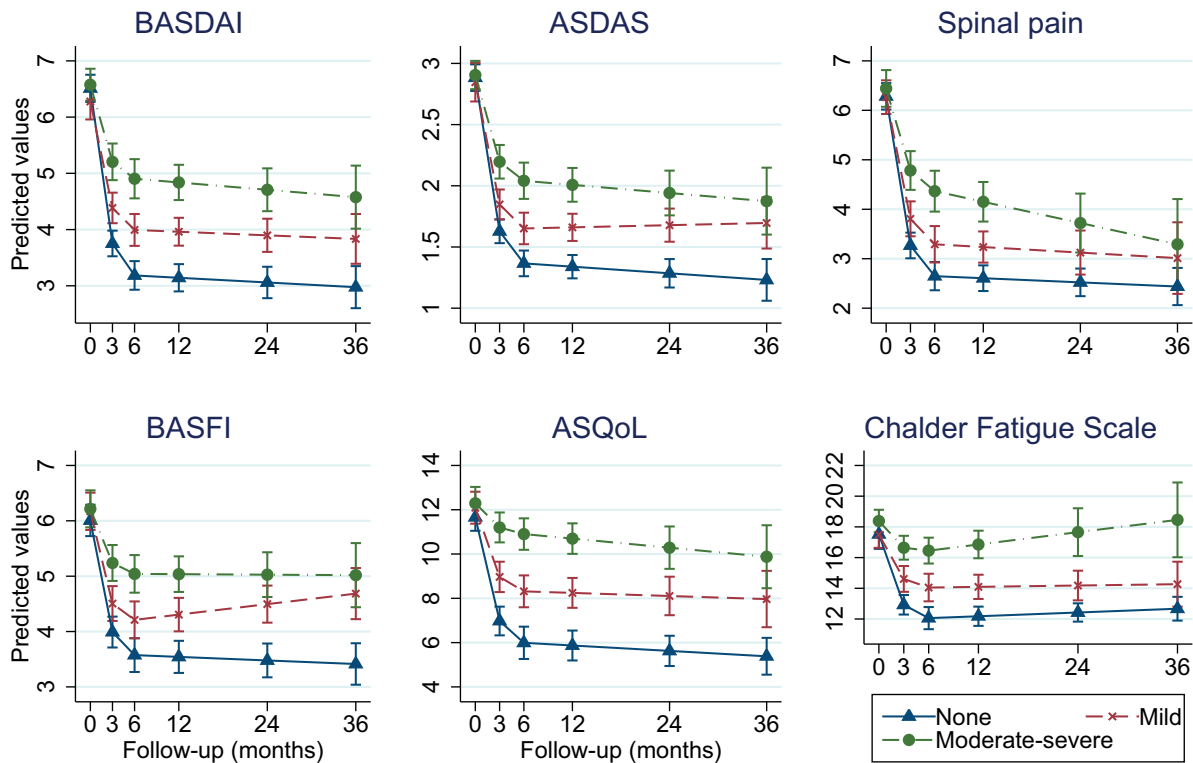
Analyses included 742 patients and 1036 person-years of follow-up, with median of 12 (interquartile range 5–25)

months. Very few individuals had follow-up beyond 4 years. Thirty-one per cent of the cohort stopped treatment over the study; 26% of participants in the 'none' group for depression discontinued, 35% in mild and 38% in the moderate-severe group; 25% in the 'none' group for anxiety discontinued, 31% in mild and 37% in the moderate-severe anxiety group. Drug survival according to HADS symptom categories are shown in [Fig. 4](#).

In marginal structural Cox models, symptoms of depression or anxiety were associated with greater hazard of treatment discontinuation (i.e. greater number of individuals discontinued at any one time, assuming rates are proportional). Compared with those with 'no' depression, the mild group had 32% higher (95% CI: 0.93, 1.87), and the moderate-severe group 45% higher (95% CI: 0.99, 2.12), hazard rate of treatment discontinuation. Compared with those with 'no' anxiety, the group with mild anxiety at baseline had 36% higher (95% CI: 0.91, 2.05), and the moderate-severe group 59% higher (95% CI: 1.12, 2.26), hazard of TNFi discontinuation.

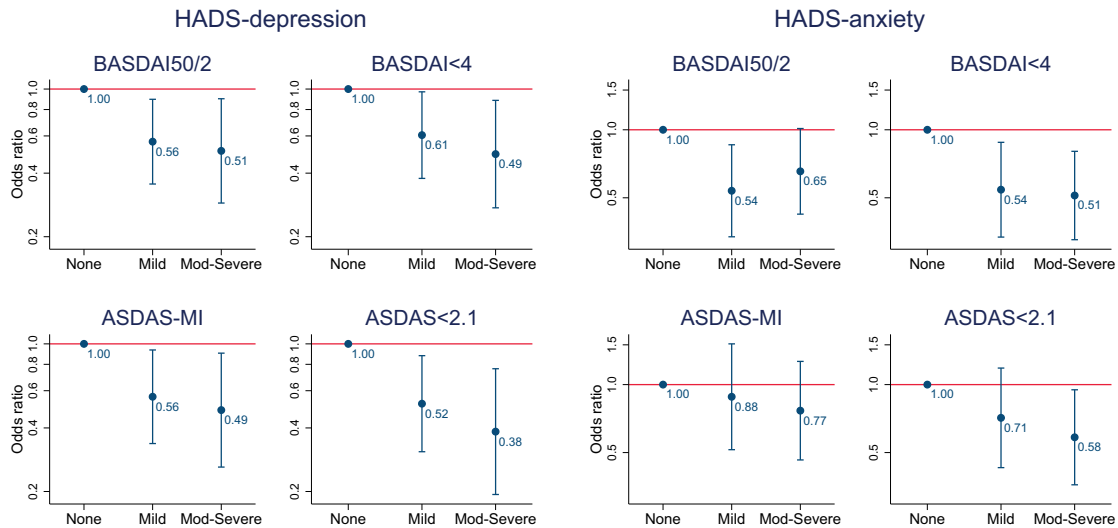
Kaplan–Meier estimators suggested potential violation of the proportional hazards assumption. Comparison between none and moderate-severe groups was approximately proportional, as shown by marginal structural Cox

Fig. 2 Change in disease indices according to baseline anxiety symptom categories



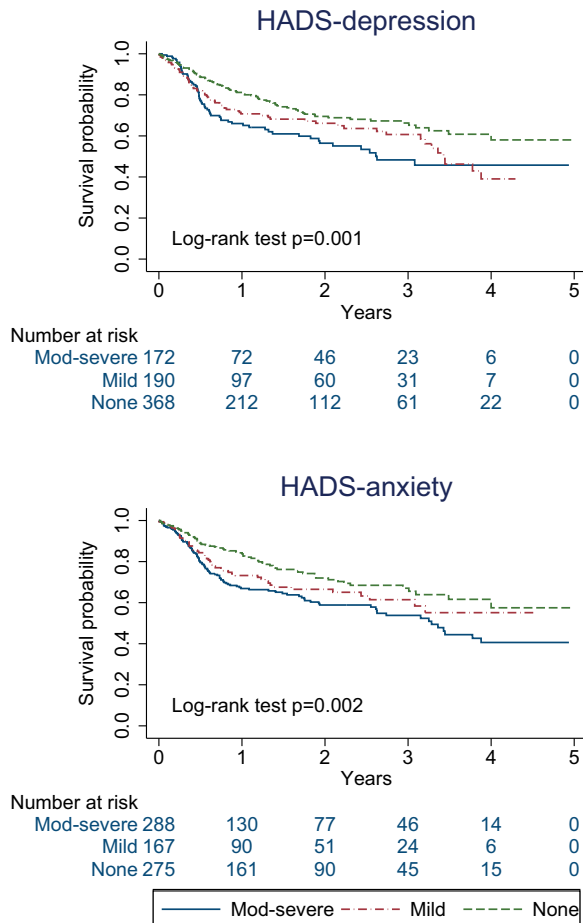
Full model output and marginal predictions are shown in [Supplementary Tables S6 and S7](#), available at *Rheumatology* online. ASDAS: AS disease activity score; ASQoL: AS quality of life questionnaire; BASDAI: Bath AS disease activity index; BASFI: Bath AS functional index.

Fig. 3 Associations between depression and anxiety symptom categories and binary responses at 6 months



Full model output shown in [Supplementary Table S8](#), available at *Rheumatology* online. ASDAS: AS disease activity score; BASDAI: Bath AS disease activity index; HADS: Hospital Anxiety and Depression Scale; MI: major improvement.

Fig. 4 Kaplan–Meier curves comparing drug survival between participant groups with different categories of baseline depression and anxiety symptom



HADS: Hospital Anxiety and Depression Scale.

models using flexible baseline hazards ([Supplementary Fig. S6](#), available at *Rheumatology* online).

Sensitivity analyses

Analyses excluding participants with baseline BASDAI < 4, imputing response at 6 months for participants who remained on treatment beyond one year and using matching weights to improve covariate overlap (thus causal inference) did not yield meaningfully different results (data not shown).

Discussion

In this large national cohort, axSpA patients with mild to severe symptoms of depression or anxiety had markedly poorer response to their first TNFi compared with those with less than mild symptoms. Interventions to optimize mental health at or before TNFi initiation may dramatically improve treatment response.

A key strength of this study is its large sample size, recruited from a broad range of rheumatology centres.

We used three response definitions that each lend unique strengths to the overall conclusion. Diagnoses correlate poorly with dynamic and often under-recognized symptoms; therefore studying symptoms of depression/anxiety, rather than documented diagnosis, provides effect estimates that have greater relevance to clinical practice and potential interventions. There were also limitations. Studying the causal effect of baseline mental health symptoms (a 'prevalent exposure') has conceptual difficulty; the implied hypothetical intervention would need to successfully improve baseline symptoms, but also symptoms pre-baseline of unknown duration. This might be considered as an intervention administered before (rather than at) TNFi initiation, but true causal effect sizes are likely smaller. Even if more realistic effect estimates were half the size, they still remain larger and more amenable to intervention than other 'modifiable risk factors': smoking status does not convincingly impact treatment outcomes [19]; and BMI is associated with treatment outcomes [20] but causal effects are conceptually problematic to estimate and intervention practically difficult to implement [21]. Categorizing HADS subscores will have reduced statistical power. Using IP weights for continuous HADS requires strong assumptions of its distribution; it would also assume a linear relationship between HADS and treatment outcomes, which the above results showed not to hold. Results from weighted generalized estimating equations should be interpreted with the limitation that participants who did not respond by the first assessment (usually after 3 months) or those who lost response would have had their treatment stopped under NICE guidance; therefore, record of such high disease activity would be censored. This informative censoring should not affect data within the first 3 months. Our results may be affected by unmeasured confounders (e.g. illness beliefs or attitudes to health). To estimate the potential effect of unmeasured confounding, take for example the comparison of 6-month BASDAI50/2 response between moderate–severe depression vs 'none' (odds ratio 0.51) and treatment discontinuation between moderate–severe anxiety vs 'none' (hazard ratio 1.59). These point estimates could be explained away by an unmeasured confounder that was associated with both the exposure and outcome by over 2-fold each, above and beyond measured confounders; weaker confounding could not do so [22]. An unmeasured confounder of this effect size is unlikely.

Mental health disorders are under-recognized and underdiagnosed in axSpA, despite their high prevalence and association with many other important health factors such as alcohol/drug abuse and suicide [23]. In this data, 71% of participants with HADS subscores ≥ 8 (which has high sensitivity and specificity for respective disorders, see Methods) did not have a documented diagnosis. Mental health symptoms are dynamic and should be assessed as such in routine clinical practice.

Numerous cross-sectional studies have shown higher disease activity in axSpA patients with depression [2]. To

our knowledge, only one prior study—which used an early version of the same BSRBR-AS dataset—described mental health (using SF-12) as part of overall predictors of treatment response in axSpA [6]. By contrast, the current analysis was designed to provide improved causal interpretation, and extended outcome measures beyond binary definitions. Binary definitions have inherent limitations when using observational data [24]; for example, high baseline DAS28 is a predictor that simultaneously increases odds of ACR70 response and decreases odds of DAS remission in RA [25]. The current analysis using HADS also has advantages. Unlike the SF-12, HADS is a validated tool to assess depression and anxiety, and excludes questions relating to somatic symptoms that may be confounded by concurrent axSpA. HADS is also an easier tool to adopt into clinical practice as it is free and its scores are easier to calculate.

Results of the current analysis are consistent with studies in RA, where symptoms of depression (measured using the SF-36) were associated with reduced good EULAR response [26]. The authors used mixed models to examine linear improvement through months 6 and 12, showing an adjusted difference in DAS28 of 0.01 units between those with and without depression symptoms. This and related effect sizes were orders of magnitude smaller than that deemed clinically meaningful. Contrast this to a difference of 2.2 BASDAI units and 0.8 ASDAS units between axSpA patients with moderate–severe and ‘no’ depression (clinically meaningful differences are around 1 unit for BASDAI and ASDAS [27]). This may be explained by limitations of SF-36 for assessing mental health or the uniquely subjective ways in which axSpA disease activity is assessed. Depression is known to influence the experience and reporting of symptoms [3], which is supported by larger effect sizes for BASDAI than (the more objective) ASDAS.

This study is also the first to assess the impact of anxiety on treatment outcomes. Anxiety is often assumed to exist in parallel with depression and thus overlooked in clinical practice and research. Although correlated, the two symptoms are not equivalent. Of those without depressive symptoms, nearly half had at least mild symptoms of anxiety and 1 in 5 had at least moderate. Conversely, 22% of those without anxiety had at least mild symptoms of depression (Supplementary Table S2, available at *Rheumatology* online). Baseline anxiety symptoms were significantly associated with treatment discontinuation, with effect sizes larger than equivalent analyses for depression. Assuming a valid causal model, reducing moderate–severe symptoms of anxiety may significantly improve treatment persistence.

These results suggest that symptoms of both depression and anxiety should be systematically screened in routine practice. This allows clinicians to predict treatment response better, but more importantly to highlight individuals who may benefit from mental health interventions. Approximately half of axSpA patients do not respond adequately to their first TNFi [5]. Optimizing mental health may offer substantial improvements in

axSpA treatment response. Conversely, neglecting mental health may lead those with severe mental health symptoms to ‘double jeopardy’, where apparent inadequate response in disease indices means their TNFi are withdrawn in healthcare systems like the UK. Randomized controlled trials of mental health interventions in axSpA (and indeed all chronic rheumatic disease) are needed. The number of pharmacological options is increasing but as yet not reliably effective in routine practice, while talking therapies are difficult to access. Improving access to the latter, for example using internet or telephone delivered cognitive behavioural therapy, may be one solution.

In conclusion, symptoms of depression and anxiety at TNFi initiation were each associated with adverse treatment outcomes. Assuming that marginal models provide an adequate approximation of real causal effects, reducing moderate–severe symptoms of depression to less than mild at TNFi initiation may improve absolute response by approximately 2 BASDAI and 1 ASDAS units, and binary response definitions by around 2-fold. Similarly, improving anxiety symptoms may reduce treatment discontinuation by up to a third. These findings highlight the importance of routinely screening and optimizing depression and anxiety in routine clinical practice. Randomized controlled trials are needed to identify efficacious mental health interventions for axSpA patients.

Acknowledgements

We are grateful to Professor Gary Macfarlane for commenting on the manuscript. We are also grateful to the staff of the BSRBR-AS register and to the recruiting staff at the clinical centres, details of which are available at: www.abdn.ac.uk/bsrbr-as. We thank Dr Lewis Carpenter for suggesting splines for modelling time. S.S.Z. analysed the data and wrote the manuscript with significant input from all co-authors. G.T.J. is the Deputy Chief Investigator on BSRBR-AS and designed the study and oversaw its conduct. In the current project they discussed results and provided input into drafts of the manuscript.

Funding: The BSRBR-AS is funded by the British Society for Rheumatology (BSR) who have received funding for this from Pfizer, AbbVie and UCB. These companies receive advance copies of manuscripts for comments. They have no input in determining the topics for analysis or work involved in undertaking it.

Disclosure statement: The authors have declared no conflicts of interest.

Data availability statement

Data from the British Society for Rheumatology Biologics Register for Ankylosing Spondylitis are available to external investigators, on reasonable request. For information on how to access data, see: www.rheumatology.org.uk.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

- Zhao SS, Pittam B, Harrison N *et al.* Diagnostic delay in axial spondyloarthritis: a systematic review and meta-analysis. *Rheumatology (Oxford)* 2021;60:1620–28
- Zhao S, Thong D, Miller N, Duffield SJ *et al.* The prevalence of depression in axial spondyloarthritis and its association with disease activity: a systematic review and meta-analysis. *Arthritis Res Ther* 2018;20:140.
- Strigo IA, Simmons AN, Matthews SC, Craig ADB, Paulus MP. Association of major depressive disorder with altered functional brain response during anticipation and processing of heat pain. *Arch Gen Psychiatry* 2008; 65:1275–84.
- Pilling S, Anderson I, Goldberg D, Meader N, Taylor C. Depression in adults, including those with a chronic physical health problem: summary of NICE guidance. *BMJ* 2009;339:b4108.
- Lord PA, Farragher TM, Lunt M *et al.*; BSR Biologics Register. Predictors of response to anti-TNF therapy in ankylosing spondylitis: results from the British Society for Rheumatology Biologics Register. *Rheumatology (Oxford)* 2010;49:563–70.
- Macfarlane GJ, Pathan E, Jones GT, Dean LE. Predicting response to anti-TNF α therapy among patients with axial spondyloarthritis (axSpA): results from BSRBR-AS. *Rheumatology* 2020;59:2481–90.
- VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol* 2019;34:211–9.
- Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance* 2019;32:42–9.
- Macfarlane GJ, Barnish MS, Jones EA *et al.* The British Society for Rheumatology Biologics Registers in Ankylosing Spondylitis (BSRBR-AS) study: protocol for a prospective cohort study of the long-term safety and quality of life outcomes of biologic treatment. *BMC Musculoskelet Disord* 2015;16:347.
- Snaith RP. The hospital anxiety and depression scale. *Health Qual Life Outcomes* 2003;1:29.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- Brennan C, Worrall-Davies A, McMillan D, Gilbody S, House A. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *J Psychosom Res* 2010;69:371–8.
- Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *J Psychosom Res* 2002;52:69–77.
- Jackson C. The Chalder Fatigue Scale (CFQ 11). *Occup Med Lond* 2015;65:86.
- National Institute for Health and Care Excellence (NICE). TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis. Technology appraisal guidance TA383. London: NICE, <https://www.nice.org.uk/guidance/ta383> (5 May 2020, date last accessed).
- Department of Communities and Local Government. The English indices of deprivation 2015. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015> (5 May 2020, date last accessed).
- Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;168:656–64.
- Yoshida K, Hernández-Díaz S, Solomon DH *et al.* Matching weights to simultaneously compare three treatment groups: comparison to three-way matching. *Epidemiology* 2017;28:387–95.
- Zhao S, Yoshida K, Jones GT *et al.* The impact of smoking on response to TNF inhibitors in axial spondyloarthritis: methodological considerations for longitudinal observational studies. *Arthritis Care Res* 2020;72:591–9.
- Liew JW, Huang IJ, Loudon DN, Singh N, Gensler LS. Association of body mass index on disease activity in axial spondyloarthritis: systematic review and meta-analysis. *RMD Open* 2020;6:e001225.
- Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes* 2008;32(Suppl 3):S8–14.
- VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-Value. *Ann Intern Med* 2017;167:268.
- Zhao SS, Solomon DH, Goodson NJ. Comment on: Comorbidity burden in axial spondyloarthritis: a cluster analysis: Reply. *Rheumatology* 2020;59:692–3.
- Zhao SS, Jones GT, Macfarlane GJ *et al.* Comorbidity and response to TNF inhibitors in axial spondyloarthritis: longitudinal analysis of the BSRBR-AS. *Rheumatology (Oxford)* 2021 (in press), doi: 10.1093/rheumatology/keaa900.
- Kristensen LE, Kapetanovic MC, Gülfe A *et al.* Predictors of response to anti-TNF therapy according to ACR and EULAR criteria in patients with established RA: results from the South Swedish Arthritis Treatment Group Register. *Rheumatology (Oxford)* 2007;47:495–9.
- Matcham F, Davies R, Hotopf M *et al.* The relationship between depression and biologic treatment response in rheumatoid arthritis: an analysis of the British Society for Rheumatology Biologics Register. *Rheumatology (Oxford)* 2018;57:835–43.
- Zochling J. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). *Arthritis Care Res (Hoboken)* 2011;63(Suppl 11):S47–58.