RMD Open
Rheumatic & Musculoskeletal Diseases

REVIEW

Biological DMARDs and disease modification in axial spondyloarthritis: a review through the lens of causal inference

Alexandre Sepriano [1,2] Sofia Ramiro [2,3] Desirée van der Heijde [2] Robert Landewé [3,4]

To cite: Sepriano A, Ramiro S, van der Heijde D, *et al*. Biological DMARDs and disease modification in axial spondyloarthritis: a review through the lens of causal inference. *RMD Open* 2021;**7**:e001654. doi:10.1136/rmdopen-2021-001654

Received 1 May 2021
Accepted 30 June 2021

[1]Rheumatology, NOVA Medical School, Universidade Nova de Lisboa, Lisboa, Portugal
[2]Rheumatology, Leiden University Medical Center, Leiden, Netherlands
[3]Rheumatology, Zuyderland Medical Centre Heerlen, Heerlen, Netherlands
[4]Clinical Immunology and Rheumatology, Amsterdam University Medical Centres, Duivendrecht, Netherlands

Correspondence to
Dr Alexandre Sepriano; alexsepriano@gmail.com

ABSTRACT

Axial spondyloarthritis (axSpA) is a chronic rheumatic disease characterised by inflammation predominantly involving the spine and the sacroiliac joints. In some patients, axial inflammation leads to irreversible structural damage that in the spine is usually quantified by the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS). Available therapeutic options include biological disease-modifying antirheumatic drugs (bDMARDs), which have been proven effective in suppressing inflammation in several randomised controlled trials (RCT), the gold standard for evaluating causal treatment effects. RCTs are, however, unfeasible for testing structural effects in axSpA mainly due to the low sensitivity to change of the mSASSS. The available literature therefore mainly includes observational research, which poses serious challenges to the determination of causality. Here, we review the studies testing the effect of bDMARDs on spinal radiographic progression, making use of the principles of causal inference. By exploring the assumptions of causality under counterfactual reasoning (exchangeability, positivity and consistency), we distinguish between studies that likely have reported confounded treatment effects and studies that, on the basis of their design, have more likely reported causal treatment effects. We conclude that bDMARDs might, indirectly, interfere with spinal radiographic progression in axSpA by their effect on inflammation. Innovations in imaging are expected, so that placebo-controlled trials can in the future become a reality. In the meantime, causal inference analysis using observational data may contribute to a better understanding of whether disease modification is possible in axSpA.

Key messages

- The evaluation of the structural effect of biological disease-modifying antirheumatic drugs (bDMARDs) in axial spondyloarthritis (axSpA) by a randomised controlled trial is currently unfeasible.
- Several observational studies addressed this enduring and highly clinically relevant question.
- Observational research can yield causal treatment effects if key causal assumptions are met.
- Causal inference principles indicate that bDMARDs might slow spinal structural progression in axSpA by suppressing inflammation.

INTRODUCTION

Axial spondyloarthritis (axSpA) is characterised by inflammation and pathological new bone formation predominantly involving the spine and the sacroiliac joints (SIJ). Patients with axSpA and structural damage on pelvic radiographs, according to the modified New York criteria,[1] are referred to as radiographic axSpA (r-axSpA) and the others as non-radiographic axSpA (nr-axSpA).

The C reactive protein (CRP) and the Ankylosing Spondylitis Disease Activity Score (ASDAS) are measures of disease activity that quantify systemic inflammation.[2] Bone marrow oedema (BME) on MRI reflects local inflammation in vertebral corners and in the SIJ.[3] Several studies have consistently shown that inflammation (ASDAS, CRP and BME) may lead to new bone formation.[4–11] Bone biopsy studies and animal models have provided the necessary biological framework, by showing that when BME subsides, it is replaced by a repair tissue with new bone-forming capability.[12 13]

Non-steroidal anti-inflammatory drugs (NSAIDs) constitute the first-line pharmacological treatment in axSpA. Patients with axial involvement who do not respond to NSAIDs should be treated with biological disease-modifying antirheumatic drugs (bDMARDs).[14] Several randomised controlled trials (RCTs) have shown that tumour necrosis factor alpha inhibitors (TNFi) and interleukin 17 inhibitors (IL-17i) are effective in suppressing inflammation and

alleviating symptoms, both in r-axSpA and nr-axSpA.[15] Since inflammation may trigger structural damage, therapies that successfully suppress the former should, in theory, stop or at least retard the latter. However, after years of research, the structural effects of bDMARDs remain under debate.[16]

In some studies, bDMARDs did not appear to have structural effects, while in others, more positive results were found.[17–19] We will explore this inconsistency by reviewing the literature under the principles of causal inference. We here use counterfactual reasoning as proposed by Rubin,[20] Balke and Pearl[21] and as recently revised by Gvozdenović *et al*.[22] Treatment effects are considered causal, under the proviso of certain assumptions: *exchangeability, positivity and consistency*. We will start by defining causality under these assumptions. We will then use this definition as a benchmark to determine the likelihood of causality of the treatment effects from studies evaluating the structural effects of bDMARDs in axSpA. We conclude by anticipating the major advances expected in the field in the near future.

CAUSAL TREATMENT EFFECTS IN RCTS

Let us consider an individual patient with axSpA who starts a bDMARD. We quantify the patient's spinal damage before the start of treatment and after a certain period of time and then record the change (factual outcome). Now, let us consider the same patient in a 'counterfactual world', in which no treatment was given. We measure the initial damage and again the change (counterfactual outcome). Because the patient is the same, it is obvious that any difference between the factual and the counterfactual outcome must be *caused* by the treatment. Obviously, we will never observe the counterfactual outcome, so it is impossible to ascertain causality in an individual patient.

Let us now consider a population of patients with axSpA (figure 1A). In the hypothetical 'world 1', all patients receive a bDMARD. We follow them all and determine their mean progression (potential outcome 1). We then hypothetically follow the same population of patients but give them a different drug ('world 2') and determine their mean progression (potential outcome 2). Since the population is the same, it is obvious that any difference between outcomes 1 and 2 is *caused* by the treatment. Even though we cannot really *observe* the two potential outcomes, we can *estimate* their expected mean values, which is what we do in RCTs.

In an RCT, randomisation ensures that, at the group level, patients who actually get the bDMARD and those who get a comparator (usually placebo) have the same characteristics of their source population and, consequently, are similar to each other. We say that treatment allocation is ignorable (it does not matter who gets what) and that the two groups are *exchangeable* (have the same characteristics). Had the patients who actually got the

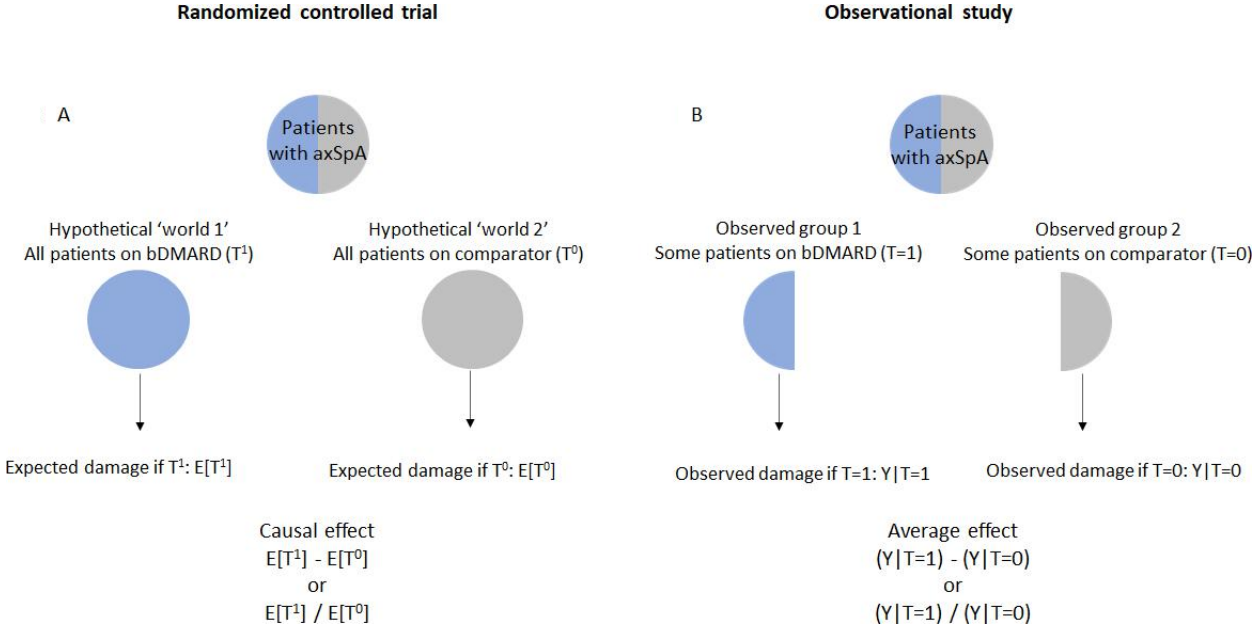

Figure 1 Causal effect versus association. Inspired by a figure from the following book: Hernán MA, Robins JM (2020). Causal inference: what if. Boca Raton: Chapman & Hall/CRC.[50] The notation below each figure indicates either a mean difference or a risk ratio as measures of the treatment effect. For instance, in an observational study, we determine the difference between the mean value of mSASSS among patients on bDMARDs (Y|T=1) and the mean value of mSASSS among the controls (Y|T=0). We might also determine the risk of developing a new syndesmophyte among those on bDMARD (Y|T=1), relative (divided) to the risk of a new syndesmophyte in controls (Y|T=0). Unlike randomised controlled trials, the treatment effect is biased because each group is not representative of the entire source population and groups are therefore not exchangeable. axSpA, axial spondyloarthritis; bDMARD, biological disease-modifying antirheumatic drug; mSASSS; modified Stoke Ankylosing Spondylitis Spinal Score.

bDMARD (factual world) hypothetically not got it (counterfactual world), they would have the same potential outcome as those on the comparator. Measuring the outcome in two groups formed by randomisation is the same as measuring the outcome in all patients under hypothetical 'world 1 and 2' conditions.

In addition to *exchangeability*, causal claims also imply the *positivity* assumption, which is met when all patients, irrespective of their characteristics, have a probability greater than zero to be allocated to the treatment or to the comparator. Finally, both the intervention and the comparator need to be well defined, and their definition must not change during the time in which the treatment effect is being evaluated (*consistency* assumption).

The three causal assumptions are usually (but not always) met in RCTs, and that is why they are the gold standard in causal inference. Of note, modern trials in rheumatology, such as treat-to-target trials and (other) strategy trials will meet the criteria of *exchangeability* and *positivity* but fail the *consistency* criterion since the content of the treatment may vary over time.

Disease modification in axSpA: RCTs

The assessment of causal treatment effects implies the use of valid outcome measures. Spinal radiography has been the imaging modality of choice to measure structural damage in patients with axSpA. The modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS) is the most often used score to measure spinal radiographic progression and performs better than any other score in terms of reliability, validity and sensitivity to change, both in patients with r-axSpA and nr-axSpA.[23 24]

Even though the mSASSS is the score with best sensitivity to change, it still takes ≥2 years for a change to be observable, at the group level, in patients with r-axSpA.[25] Progression is even slower in early axSpA.[26] Slow progression renders RCTs aiming at testing causal structural effects rather unfeasible. Patients in an RCT are all, by design, eligible to receive the treatment under study (eg, all have high levels of disease activity). As mentioned, bDMARDs reduce the signs and symptoms of the disease. It is therefore unethical to deprive patients from an effective therapy they would likely benefit from, for the time it takes to evaluate a potential structural effect. One alternative is to *observe* the effect of treatment as prescribed in clinical practice, a setting, however, in which the causality assumptions will less likely hold.

CAUSAL TREATMENT EFFECTS IN OBSERVATIONAL RESEARCH

Randomisation ensures that differences in outcomes between groups are fully explained by treatment. The same cannot be said if treatment prescription is not random but made by a clinician. Let us consider the population of patients with axSpA. In the factual world, only a fraction of patients, those who have failed—or are intolerant to—conventional treatment, are eligible to receive bDMARDs (figure 1B). That means treated patients usually have more severe disease than untreated patients. Treatment allocation is, therefore, not ignorable, and treated and untreated groups are *not exchangeable*. This problem is often referred to as 'confounding by indication'.

A confounder ('C') influences both the treatment decision ('T') and the outcome ('Y') and is not in the causal pathway between both (figure 2). For instance, patients with higher *pretreatment* levels of ASDAS ('C') are more likely to receive a bDMARD ('T') than those with lower levels. Also, higher ASDAS may lead to higher mSASSS ('Y'). Thus, in a non-randomised experiment, bDMARDs may affect the mSASSS due to (1) a causal effect of the drug on mSASSS ('front-door' path: T→Y) or (2) a spurious effect of the drug on mSASSS through ASDAS ('back-door' path: T←C→Y). In an RCT, randomisation 'closes' all measured and unmeasured 'back doors' (figure 2A). In an observational study, the 'back doors' keep open, which may lead to spurious effects (figure 2B).

In addition to ASDAS (also CRP and BME), the presence of damage at baseline, male gender, longer disease duration and smoking all associate with radiographic progression.[8 9 27] These characteristics are confounders if they also influence the decision to start a bDMARD. One possible solution is to 'fix' the values of confounders between treated and controls and estimate the so-called average marginal effect (AME) (figure 2C). Methods to estimate the AME, by 'back-door' adjustment, that is, methods used to take 'confounding by indication' into account, include matching, stratification, regression adjustment, propensity score (PS) adjustment and inverse probability of treatment weighting.

Under fixed values of all confounders, and assuming no unmeasured confounding, the treated and untreated are *fully exchangeable*. However, this alone does not suffice to guarantee causal treatment effects. In observational research, treatment groups are not necessarily *consistently* defined. Bias may also occur if patients under treatment who are included in studies, and are therefore compared with controls, have relevant prognostic dissimilarities with those who are not. If that happens, the *positivity* assumption is likely violated as patient's characteristics are constraining treatment allocation (eg, if patients with worse prognostic factors have zero probability to be treated).

Disease modification in axSpA: observational research

Studies with equal exposure to treatment and without a comparator

After the completion of an RCT, patients on placebo usually switch to the active drug and, together with those on treatment from the start, are followed in open-label extensions (OLE), provided they meet certain inclusion criteria. The structural effect of bDMARDs in axSpA have, thus far, been evaluated in OLE with patients continuously exposed to TNFi for up to 10 years. All studies included patients with r-axSpA, except for one

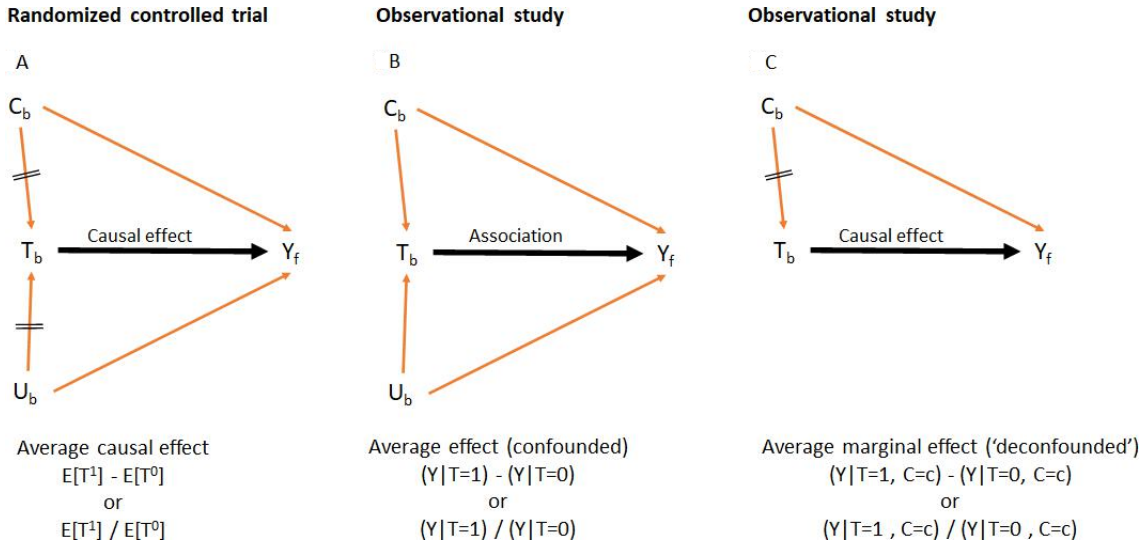

Figure 2 Directed acyclic graphs (DAGs) representing (A) a causal effect in a randomised controlled trial, (B) an association in an observational study and (2C) an average marginal effect in an observational study. Black arrow, association of interest ('front door'); red arrows, biasing paths ('back doors'); crossed red arrows, closed biasing paths (eg, by randomisation in A and adjustment in C). T_b, treatment status at baseline (yes/no); C_b, a set of observed pretreatment confounders at baseline; Ub, a set of unobserved pretreatment confounders at baseline; Y_f, outcome at follow-up. The notation below each DAG indicates either a mean difference or a risk ratio as measures of the treatment effect. The average marginal effect is the equivalent of a causal effect of a randomised controlled trial under the assumption that all confounders had been measured and adjusted for (no unmeasured confounding).

also including patients with nr-axSpA. The number of patients in the original RCTs ranged from 84 to 325 and those in the corresponding OLEs (ie, with complete follow-up and imaging data) from 17 to 93.[28–30] OLEs consistently report minimal change in the mSASSS. In one 4-year OLE, no meaningful change in mSASSS was observed in patients with nr-axSpA.[30] In the same study, the change in mSASSS was higher in the first 2 years than in the last 2 years in patients with r-axSpA (0.8 vs 0.4), suggesting a late-onset structural benefit.

In studies without a comparator, which is the case in OLE, it is impossible to address the *exchangeability* assumption. The following counterfactual question remains unanswered: had the patients not received the bDMARD, would their average change in mSASSS be different than the observed change? In addition, the few patients that continue on treatment and are therefore included in the OLE are not necessarily representative of the larger population of patients eligible for bDMARDs from the corresponding RCTs. The *positivity* assumption is most likely violated, since patients with milder disease are, arguably, more likely to stay on treatment in the OLE (right censoring bias). On the contrary, the *consistency* assumption is likely to hold, as all patients receive the same drug over the entire follow-up. Even if patients initially on placebo are included, this is usually for a well-defined and (very) short period of time.

Comparative studies with equal exposure to treatment

The large majority of studies evaluating the structural effect of bDMARDs included a comparator, and among these, most were done in a setting in which all patients had to be on bDMARD, or on the comparator, continuously over the entire study (time-fixed treatment). Confounders, when considered, were evaluated at baseline before the start of treatment (time-fixed confounders), and the outcome was assessed at the end (time-fixed outcome).

Studies with time-fixed treatment compared patients with r-axSpA on bDMARD to either ineligible patients or patients in whom bDMARDs were not an option (eg, historical cohorts). Table 1 summarises the main findings of studies reporting the mean change in mSASSS and table 2 the studies reporting binary definitions of progression (with some overlap). The effect size (Cohen's d in table 2 and OR in table 3) was calculated (when not reported). In each table, the methods used for handling confounding are shown.

In three studies, patients on TNFi from OLEs were compared with patients not receiving TNFi from the Outcome in AS International Study (OASIS) historical cohort.[31–33] As expected, patients on TNFi had worse prognostic factors (eg, higher levels of disease activity and damage) than patients from OASIS. Thus, patients on TNFi were compared with patients from OASIS who would have fulfilled the inclusion criteria of the OLE. The effect size was zero in the matched population (table 1) as well as in stratified and adjusted analyses. The absence of a structural effect was also reported in other studies comparing TNFi to no TNFi, as well as IL-17i to NSAIDs, up to 2 years of follow-up.[34–37]

One study comparing 22 patients from an 8-year OLE with a historical cohort has shown slower progression

Table 1 Effect of bDMARDs on the change in mSASSS: comparative studies with continuous exposure to treatment

Study	Study design	Follow-up (years)	Treatment	N	Handling of pretreatment confounders	Δ mSASSS Mean (SD)	Effect size* (95% CI)
Baraliakos *et al*, 2005[34]	OLE versus cohort (GESPIC)	2	INF	41	Adjust. (not clear which variables)	0.4 (2.7)	−0.1 (−0.5 to 0.3)†
			No TNFi	41		0.7 (2.8)	
van der Heijde *et al*, 2008[31]	OLE versus cohort (OASIS)	2	INF	165	Match. (inclusion criteria), adjust. (mSASSS) and strat. (gender, age, CRP, BASDAI, BASFI, BASMI, PGA and HLA-B27)	0.9 (2.6)	−0.1 (−0.4 to 0.2)‡
			No TNFi	61		1.2 (3.9)	
van der Heijde *et al*, 2008[32]	OLE versus cohort (OASIS)	Up to 2	ETA	257	Match. (inclusion criteria), adjust. (mSASSS) and strat. (NSAIDs)	0.9 (2.5)	−0.1 (−0.4 to 0.1)‡
			No TNFi	76		1.3 (3.6)	
van der Heijde *et al*, 2009[33]	OLE versus cohort (OASIS)	Up to 2	ADA	307	Match. (inclusion criteria) and adjust. (mSASSS)	0.8 (2.6)	0.0 (−0.3 to 0.2)‡
			No TNFi	77		0.9 (4.1)	
Pedersen *et al*, 2011[37]	Cohort	2	TNFi	23	Not addressed	1.4 (1.9)	0.0 (−0.6 to 0.5)
			No TNFi	27		1.5 (3.1)	
Kang *et al*, 2013[36]	Cohort	2	TNFi	26	Adjust. (mSASSS)	3.3 (4.2)**	0.2 (−0.3 to 0.7)†
			No TNFi	37		2.3 (5.1)**	
Braun *et al*, 2019[35]	OLE versus OLE (ENRADAS)	2	SEC 75/150	168	Adjust. (mSASSS)	0.6 (NR)††	NC
			NSAIDs	69		0.9 (NR)††	
Braun *et al*, 2014[51]	OLE (≠doses)	4	GOL 100	122	Strat. (CRP and syndesmophytes)	2.0 (5.6)	0.1 (−0.1 to 0.4)¶
			GOL 50	111		1.3 (4.1)	
Braun *et al*, 2019[52]	OLE (≠doses)	Up to 4	SEC 150	71	Not addressed	1.2 (3.9)	−0.1 (−0.6 to 0.4)
			SEC 75	61		1.8 (4.3)	0.0 (−0.4 to 0.5)
			SEC 75 to >150	23		1.6 (5.7)	REF
Baraliakos *et al*, 2014[38]	OLE versus cohort (Herne)	Up to 8	INF	22	Adjust. (mSASSS§)	7 (NR)	NC
			No TNFi	34		12 (NR)	

Only studies in which radiographs of both groups were scored simultaneously are included.
Cohen's d: standardised mean difference between treatment (t) and control (c)=mean$_t$ – mean$_c$/pooled SD. Pooled SD=square root((SD$_T^2$(N$_t$-1) + SD$_c^2$*(N$_c$-1))/(N$_t$ +N$_c$-2)).
†Calculated from the observed change in mSASSS (p value of adjusted analysis not significant and no adjusted effect size reported).
‡Calculated from observed change in mSASSS in the matched cohorts (adjusted and stratified analyses also not significant).
§Variables tested but not included (p>0.05): age, symptom duration, HLA-B27, BASDAI and BASFI.
¶Calculated from observed change in mSASSS in the entire study population (stratified analysis also non-significant). μ estimated from a model adjusted for mSASSS at baseline.
**Only lumbar spine used in the calculation of mSASSS.
††Estimated from a model adjusted for mSASSS at baseline.
ADA, adalimumab; adjust, adjustment; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; bDMARDs, biological disease-modifying antirheumatic drugs; CRP, C reactive protein; ETA, etanercept; GESPIC, German Spondyloarthritis Inception Cohort; GOL, golimumab; HLA, human leukocyte antigen; INF, infliximab; Match, matching; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; NC, not possible to calculate: SD of the change score per group not reported; NR, not reported; NSAIDs, non-steroidal anti-inflammatory drugs; OASIS, Outcome in AS International Study; OLE, open-label extension; PGA, patient global assessment; SEC, secukinumab; strat, stratification; TNFi, tumour necrosis factor alpha inhibitor.

Table 2 Effect of bDMARDs on the likelihood of progression: comparative studies with continuous exposure to treatment

Study	Study design	Follow-up (years)	Treatment	N	Handling of pretreatment confounders	Definition of progression	OR (95% CI)
Baraliakos *et al*, 2005[34]	OLE versus cohort (GESPIC)	2	INF No TNFi	41 41	Adjust. (not clear which variables)	≥1 mSASSS	1.5 (0.4 to 6.5)*
van der Heijde *et al*, 2008[31]	OLE versus cohort (OASIS)	2	INF No TNFi	165 61	Match. (inclusion criteria), adjust. (mSASSS) and strat. (gender, age, CRP, BASDAI, BASFI, BASMI, PGA and HLA-B27)	≥1 mSASSS	0.9 (0.6 to 1.5)†
Haroon *et al*, 2013[39]	Cohort	1.5–9	TNFi No TNFi	201 133	PS matching‡	≥1 mSASSS/year	**0.3 (0.1 to 0.8)**‡
Kim *et al*, 2016[40]	Cohort	2–5	TNFi No TNFi	269 341	PS matching§	≥2 mSASSS	0.7 (0.3 to 1.6)§

*Calculated from observed data (p value of adjusted analysis not significant and no adjusted effect size reported).
†Calculated from observed data in the matched cohorts (adjusted and stratified analyses also not significant).
‡Adjusted OR after PS matching (n=142); baseline variables used to estimate the PS: gender, HLA-B27, BASDAI, ESR, mSASSS, NSAID index, disease duration, smoking pack-years and age of onset of symptoms.
§adjusted OR after PS matching (n=166 in each group); baseline variables used to estimate the PS: age, gender and CRP.
adjust, adjustment; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; bDMARDs, biological disease-modifying antirheumatic drugs; CRP, C reactive protein; ESR, erythrocyte sedimentation rate; GESPIC, German Spondyloarthritis Inception Cohort; HLA, human leukocyte antigen; INF, infliximab; Match, matching; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; NSAIDs, non-steroidal anti-inflammatory drugs; OASIS, Outcome in AS International Study; OLE, open-label extension; PGA, patient global assessment; PS, propensity score; strat, stratification; TNFi, tumour necrosis factor alpha inhibitor.

with TNFi versus no TNFi after the fourth year of follow-up, adjusting for the mSASSS at baseline.[38] In two cohort studies, patients on TNFi were compared with those not on TNFi after PS matching, with one showing a positive structural effect and the other no effect.[39 40]

Not all studies evaluating the effect of 'time-fixed' exposure to bDMARDs addressed confounding, and those that did considered only the effect of baseline variables (mostly damage). In studies that span for several years, it is arguable whether handling baseline confounding suffices for full *exchangeability*. Under null (or partial) *exchangeability*, it is reasonable to expect worse prognostic factors to dominate in the treated and therefore for underestimated structural effects. In addition, right censoring bias is likely, since patients had to keep the drug for several years to qualify for inclusion, thus violating the *positivity* assumption. Overestimation of the treatment effect is likely, if patients with better prognosis are preferentially selected, since the comparison is made in a population most likely resembling the patients eligible to receive the control rather than the patients eligible for the treatment. Interpreting the direction of residual bias is difficult if neither assumption holds. Finally, the *consistency* assumption is likely not met when patients on bDMARD are compared with those not on bDMARD, as 'no bDMARD' is poorly defined and is likely to vary over time.

Comparative studies with variable exposure to treatment

In recent studies, patients were evaluated in regular intervals. In order to be included, they only had to be followed during one period. For instance, a patient could only start a bDMARD in the second interval and then keep it until the end of the study (eg, six intervals in total, five on bDMARD). Another patient could have been treated since baseline but was lost to follow-up somewhere in the end of the first interval, thus 'contributing' with data to only one of six possible intervals. In this type of study, with unequal exposure to bDMARDs ('time-varying'), the AME is the combined effect of treatment considering all available intervals, estimated with methods such as generalised estimating equations, that handle repeated observations per patient.

In studies with 'time-varying' treatment, treatment status is recorded at the start—and the outcome at the end—of each interval. Measures of disease activity (eg, ASDAS), damage (eg, mSASSS) and comedication (eg, NSAIDs), among other variables, are also recorded per interval. These features can be time-varying confounders, if they influence the prescription of bDMARDs at the start—and the outcome at the end—of each interval (figure 3A). ASDAS is also thought to mediate the structural effect of bDMARDs in axSpA. Importantly, a mediator, different to a confounder, resides in the causal pathway between the treatment and the outcome (figure 3B). Testing for mediation implies decomposing the total effect into (1) the direct effect of bDMARDs

Table 3 Effect of bDMARDs on the change in mSASSS and on the likelihood of progression: comparative studies with variable exposure to treatment

Study	Follow-up years	N patients/ intervals*	Comparison	Outcome	Handling of pretreatment time-varying confounders	Time-varying mediator	Total effect (95% CI)	Mediation analysis
Molnar *et al*, 2017[44]	Up to 10	432/616	TNFi versus no TNFi	≥2 mSASSS	Adjust. (mSASSS, **ASDAS** and NSAID+baseline variables)	ASDAS (start interval)	**OR: 0.5 (0.3 to 0.9)**	Indirect effect + Direct effect –
Gensler *et al*, 2018[42]	4	519/NR	TNFi versus no TNFi†	mSASSS	Adjust. (**ASDAS**+baseline variables) Strat. (NSAID index)	NA	**Δ: –3.3 (–4.0 to –2.6)**	NA
Park *et al*, 2019[45]	NR	215/518	TNFi versus NSAID	mSASSS	No TV confounders. Adjust. BL (age, **CRP,** smoking and syndesmophytes)	CRP (aver interval)	**β: –0.9 (–1.5 to –0.3)**	Indirect effect + Direct effect –
Koo *et al*, 2020[43]	Up to 18	338/2364	TNFi on versus TNFi off	mSASSS	IPTW (mSASSS, **ESR, CRP, BASDAI** and comedication+baseline variables)	NA	**β: –0.1 (–0.2 to 0.0)**	NA
Sepriano *et al*, 2021[46]	Up to 10	314/442	TNFi versus no TNFi	mSASSS	Adjust. (mSASSS)‡+PS. Adjust. BL (gender, symptom duration, **ASDAS**, HLA-B27, mSASSS, EMM and NSAIDs)	ASDAS (start interval)	**β: –0.8 (–1.4 to –0.2)**	Indirect effect + Direct effect +

Total effect: Effect of TNFi on radiographic progression at the end of each interval adjusting for pretreatment confounders but not for the mediator.
Direct effect: Effect of TNFi on radiographic progression at the end of each interval adjusting for pretreatment confounders and for the mediator.
Indirect effect: Effect of TNFi on radiographic progression at the end of each interval through the mediator (also adjusting for pretreatment confounders).
*2-year intervals in all studies.
†If NSAID index ≥50.
‡NSAIDs not kept in the final model because it did not prove to confound the association of interest.
adjust, adjustment; ASDAS, Ankylosing Spondylitis Disease Activity Score; aver, average; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; bDMARDs, biological disease-modifying antirheumatic drugs; BL, baseline; CRP, C reactive protein; EMM, extramusculoskeletal manifestations (uveitis, inflammatory bowel disease and psoriasis); ESR, erythrocyte sedimentation rate; HLA, human leukocyte antigen; IPTW, inverse probability treatment weighting; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; NR, not reported; NSAIDs, non-steroidal anti-inflammatory drugs; PS, propensity score; Strat, stratification; TNFi, tumour necrosis factor alpha inhibitor; TV, time-varying.

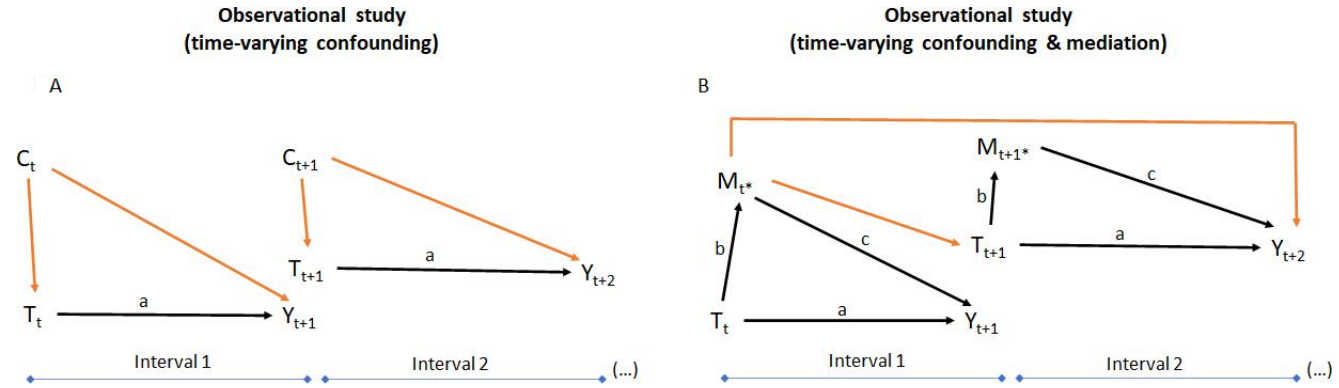

Figure 3 Directed acyclic graphs (DAGs) representing (A) a longitudinal study with time-varying confounding and (B) a longitudinal study with time-varying mediation. For the purpose of illustration, the DAGs are simplified (absent arrows) by making the following (unlikely to hold) assumptions: Y_{t+1} does not influence T_{t+1}; Y_{t+1} does not influence Y_{t+2}; C_t does not influence C_{t+1}. Black arrows, associations of interest ('front doors'); red arrows, biasing paths ('back doors'); T_t, treatment status (yes/no) at the start of the first interval; T_{t+1}, treatment status at the start of the second interval; C_t, a set of pretreatment confounders at the start of the first interval; C_{t+1}, a set of pretreatment confounders at the start of the second interval; Y_{t+1}, outcome at end of the first interval; Y_{t+2}, outcome at the end of the second interval; M_{t*}, mediator measured after treatment assignment and before outcome assessment on the first interval; M_{t+1*}, mediator measured after treatment assignment and before outcome assessment on the second interval.

on mSASSS adjusting for ASDAS ('path a') and (2) the indirect effect of bDMARDs through the reduction of ASDAS (path b), which in turn affects mSASSS (path c). Mediation occurs if the indirect effect drives part of the drug's total effect.[41] As illustrated in figure 3B, in theory, time-varying ASDAS can both confound and mediate the structural effects of bDMARDs.

The studies evaluating the effect of time-varying treatment with bDMARDs on spinal radiographic progression are summarised in table 3. All studies tested only TNFi in patients with r-axSpA.[39 42–46] Follow-up ranged from 4 to 18 years; however, most patients had only contributed to few intervals. Both baseline and time-varying pretreatment confounders were considered, including measures of disease activity, damage and comedication. The total effect, adjusting for pretreatment confounders, was significant in all studies (table 3). Two studies tested whether ASDAS at the start of the interval was a time-varying mediator.[44 46] Another study tested the mediating effect of the average value of CRP per interval.[45] All three studies have shown that bDMARDs inhibit structural progression *indirectly* by lowering the levels of ASDAS (or CRP). In two of these studies, the direct effect of bDMARDs was not statistically significant,[44 45] suggesting that the structural benefit was solely explained by the reduction of disease activity. In the third study, however, the direct effect of bDMARDs remained statistically significant after adjusting for ASDAS.[46]

Almost all studies reporting the time-varying structural effect of TNFi have handled time-varying confounding. Even though residual confounding is still a possibility, it can be argued that its likelihood is lower as compared with other types of studies discussed above. A claim of (full) *exchangeability* is therefore the only logical consequence. In addition, studies that allow a variable exposure to the treatment will minimise the risk of right censoring; both patients with a worse prognosis, who may be followed in fewer intervals, and those with a better prognosis, who may be followed in several intervals, can be included in the analysis. However, most studies had most patients followed for only a few intervals, which might render the *positivity* assumption less likely had a better balance been achieved. Finally, almost all studies compared treatment with TNFi with no TNFi, thus compromising the *consistency* assumption.

SUMMARY AND FUTURE PERSPECTIVES

The effect of bDMARDs on spinal radiographic progression in patients with axSpA has been extensively studied over the past 15 years. Studies without a comparator suggest that bDMARDs may slow progression, but a claim of causality is implausible in such a setting. The *exchangeability* assumption is not even possible to assess, and *positivity* is unlikely due to right censoring bias. The likelihood of causality increases in studies with a comparator. However, studies requiring all patients to stay on treatment during the entire study are also susceptible to right censoring bias (worst patients drop out). In these studies, confounding was only considered at baseline limiting the likelihood of *exchangeability*. It is therefore difficult to interpret both the negative, short-term (≤2 years), studies and the inconsistent results from studies with longer follow-up. Studies with unequal (time-varying) exposure to treatment are the most likely to yield causal structural effects. Their design protects, to some extent, against 'right censoring', thus making *positivity* more likely. In addition, these studies handle inherently time-varying confounders as such, thus increasing the chance of (full) *exchangeability* (table 4).

In all studies in the top of 'causal hierarchy', treatment with TNFi consistently reduced radiographic progression

Table 4 Causality of the effect of bDMARDs on spinal radiographic progression in the available literature

Study design	Assumption			
	Exchangeability	Positivity	Consistency	Causality
Equal exposure No comparator	–	–	+	Very unlikely
Equal exposure Comparator	+/–	–	–	Unlikely
Variable exposure Comparator	+	+/–	–	Likely

Assumption is very unlikely (–), unlikely (+/–) and likely (+) to be met.

as compared with no treatment. This effect was either partially or entirely mediated by their effect on inflammation. A causal inflammation-mediated effect is in line with the evidence that inflammation drives structural progression and strongly argues in favour of a treat-to-target strategy in axSpA. In one study, a direct effect, that means through other (unknown) mechanisms, was also found. Although not implausible, the fluctuating nature of inflammation in axSpA can also explain this finding. 'No detectable inflammation' (eg, no BME or ASDAS <1.3) is not necessarily the same as 'no inflammation present'. Despite consistent results, studies with 'time-varying' treatment are not without limitations. Future studies addressing their limitations, as exposed here, will likely contribute to a better understanding of the structural effect of bDMARDs in axSpA.

Recent data suggest that the CT Syndesmophyte Score (CTSS) is more sensitive to change than the mSASSS.[47] Low-dose CT, and other imaging innovations, may render RCTs testing structural effects feasible in the future, by decreasing the time needed to observe a treatment effect. Of note, observational studies using CTSS as a measure of structural damage face similar challenges, as studies using the mSASSS, in identifying causal treatment effects. Trials comparing TNFi to IL-17i are also expected but will only be informative if their structural effects really differ.[16] In the absence of an RCT, however, causal inference from observational research can still be informative.[48] New causal analyses done with well-defined comparators will likely clarify the effect of TNFi on mSASSS in r-axSpA as well as in nr-axSpA and also for drugs other than TNFi.

Directed acyclic graphs (DAGs) are powerful instruments in causal inference and will likely become more common in the rheumatological literature in the coming years. The model represented in a DAG (eg, figure 3) is causal, provided its underlying assumptions (arrows and nods) hold. This is why DAGs are also named structural causal models. In addition to 'back-door' adjustment, other methods such as 'front-door' adjustment with 'shielded mediators' and instrumental variables can be used to handle confounding in structural causal models.[49]

Is disease modification possible in axSpA? The definitive answer will likely be given in the next few years, by RCTs (when using a different structural damage assessment method, eg, with low-dose CT), but preceded by thorough theoretical causal analysis.

Collaborators None.

Contributors All authors drafted the text and approved the final version for publication.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests AS: Consulting/speaking fees from MSD, UCB, Novartis. SR: Research grant from MSD; Consultancy/speaking fees from AbbVie, Eli Lilly, MSD, Novartis, UCB, Sanofi. DvdH: Consulting fees AbbVie, Amgen, Astellas, AstraZeneca, Bayer, BMS, Boehringer Ingelheim, Celgene, Cyxone, Daiichi, Eisai, Eli-Lilly, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda, UCB Pharma Director of Imaging Rheumatology bv. RL: Consulting fees from AbbVie, BMS, Celgene, Eli-Lilly, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Merck, Novartis, Pfizer, Roche, UCB and is Director of Rheumatology Consultancy bv

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

ORCID iDs
Alexandre Sepriano http://orcid.org/0000-0003-1954-0229
Sofia Ramiro http://orcid.org/0000-0002-8899-9087
Desirée van der Heijde http://orcid.org/0000-0002-5781-158X
Robert Landewé http://orcid.org/0000-0002-0577-6620

REFERENCES

1 van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the new York criteria. *Arthritis Rheum* 1984;27:361–8.
2 Lukas C, Landewé R, Sieper J, *et al*. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:18–24.
3 Maksymowych WP. The role of imaging in the diagnosis and management of axial spondyloarthritis. *Nat Rev Rheumatol* 2019;15:657–72.
4 Baraliakos X, Heldmann F, Callhoff J, *et al*. Which spinal lesions are associated with new bone formation in patients with ankylosing spondylitis treated with anti-TNF agents? a long-term observational study using MRI and conventional radiography. *Ann Rheum Dis* 2014;73:1819–25.
5 Dougados M, Sepriano A, Molto A, *et al*. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. *Ann Rheum Dis* 2017;76:1823–8.
6 Machado PM, Baraliakos X, van der Heijde D, *et al*. Mri vertebral corner inflammation followed by fat deposition is the strongest contributor to the development of new bone at the same vertebral

corner: a multilevel longitudinal analysis in patients with ankylosing spondylitis. *Ann Rheum Dis* 2016;75:1486–93.
7 Poddubnyy D, Protopopov M, Haibel H, *et al*. High disease activity according to the ankylosing spondylitis disease activity score is associated with accelerated radiographic spinal progression in patients with early axial spondyloarthritis: results from the German spondyloarthritis inception cohort. *Ann Rheum Dis* 2016;75:2114–8.
8 Poddubnyy D, Rudwaleit M, Haibel H, *et al*. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. *Ann Rheum Dis* 2011;70:1369–74.
9 Ramiro S, van der Heijde D, van Tubergen A, *et al*. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. *Ann Rheum Dis* 2014;73:1455–61.
10 Sepriano A, Ramiro S, Wichuk S, *et al*. Disease activity is associated with spinal radiographic progression in axial spondyloarthritis independently of exposure to tumour necrosis factor inhibitors. *Rheumatology* 2021;60:461–2.
11 van der Heijde D, Machado P, Braun J, *et al*. Mri inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. *Ann Rheum Dis* 2012;71:369–73.
12 Bleil J, Maier R, Hempfing A, *et al*. Granulation tissue Eroding the Subchondral bone also promotes new bone formation in ankylosing spondylitis. *Arthritis Rheumatol* 2016;68:2456–65.
13 Tseng H-W, Pitt ME, Glant TT, *et al*. Inflammation-Driven bone formation in a mouse model of ankylosing spondylitis: sequential not parallel processes. *Arthritis Res Ther* 2016;18:35.
14 van der Heijde D, Ramiro S, Landewé R, *et al*. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;76:978–91.
15 Sepriano A, Regel A, van der Heijde D, *et al*. Efficacy and safety of biological and targeted-synthetic DMARDs: a systematic literature review Informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. *RMD Open* 2017;3:e000396.
16 van der Heijde D, Landewé R. Inhibition of spinal bone formation in AS: 10 years after comparing adalimumab to OASIS. *Arthritis Res Ther* 2019;21:225.
17 Boers N, Michielsens CAJ, van der Heijde D, *et al*. The effect of tumour necrosis factor inhibitors on radiographic progression in axial spondyloarthritis: a systematic literature review. *Rheumatology* 2019;58:1907–22.
18 Baraliakos X, Gensler LS, D'Angelo S, *et al*. Biologic therapy and spinal radiographic progression in patients with axial spondyloarthritis: a structured literature review. *Ther Adv Musculoskelet Dis* 2020;12:1759720X20906040.
19 Karmacharya P, Duarte-Garcia A, Dubreuil M, *et al*. Effect of therapy on radiographic progression in axial spondyloarthritis: a systematic review and meta-analysis. *Arthritis Rheumatol* 2020;72:733–49.
20 Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974;66:688–701.
21 Balke A, Pearl J. Counterfactual probabilities: computational methods, bounds and applications. UAI 1994.
22 Gvozdenović E, Malvisi L, Cinconze E, *et al*. Causal inference concepts applied to three observational studies in the context of vaccine development: from theory to practice. *BMC Med Res Methodol* 2021;21:35.
23 Ramiro S, Claudepierre P, Sepriano A, *et al*. Which scoring method depicts spinal radiographic damage in early axial spondyloarthritis best? five-year results from the DESIR cohort. *Rheumatology* 2018;57:1991–2000.
24 Wanders AJB, Landewé RBM, Spoorenberg A, *et al*. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the outcome measures in rheumatology clinical trials filter. *Arthritis Rheum* 2004;50:2622–32.
25 Ramiro S, Stolwijk C, van Tubergen A, *et al*. Evolution of radiographic damage in ankylosing spondylitis: a 12 year prospective follow-up of the OASIS study. *Ann Rheum Dis* 2015;74:52–9.
26 Ramiro S, van der Heijde D, Sepriano A, *et al*. Spinal radiographic progression in early axial spondyloarthritis: five-year results from the DESIR cohort. *Arthritis Care Res* 2019;71:1678–84.
27 Dougados M, Demattei C, van den Berg R, *et al*. Rate and predisposing factors for Sacroiliac joint radiographic progression after a two-year follow-up period in recent-onset spondyloarthritis. *Arthritis Rheumatol* 2016;68:1904–13.
28 Dijkmans B, Emery P, Hakala M, *et al*. Etanercept in the longterm treatment of patients with ankylosing spondylitis. *J Rheumatol* 2009;36:1256–64.
29 Poddubnyy D, Fedorova A, Listing J, *et al*. Physical function and spinal mobility remain stable despite radiographic spinal progression in patients with ankylosing spondylitis treated with TNF-α inhibitors for up to 10 years. *J Rheumatol* 2016;43:2142–8.
30 van der Heijde D, Baraliakos X, Hermann K-GA, *et al*. Limited radiographic progression and sustained reductions in MRI inflammation in patients with axial spondyloarthritis: 4-year imaging outcomes from the RAPID-axSpA phase III randomised trial. *Ann Rheum Dis* 2018;77:699–705.
31 van der Heijde D, Landewé R, Baraliakos X, *et al*. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. *Arthritis Rheum* 2008;58:3063–70.
32 van der Heijde D, Landewé R, Einstein S, *et al*. Radiographic progression of ankylosing spondylitis after up to two years of treatment with etanercept. *Arthritis Rheum* 2008;58:1324–31.
33 van der Heijde D, Salonen D, Weissman BN, *et al*. Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. *Arthritis Res Ther* 2009;11:R127.
34 Baraliakos X, Listing J, Rudwaleit M, *et al*. Radiographic progression in patients with ankylosing spondylitis after 2 years of treatment with the tumour necrosis factor alpha antibody infliximab. *Ann Rheum Dis* 2005;64:1462–6.
35 Braun J, Haibel H, de Hooge M, *et al*. Spinal radiographic progression over 2 years in ankylosing spondylitis patients treated with secukinumab: a historical cohort comparison. *Arthritis Res Ther* 2019;21:142.
36 Kang KY, Ju JH, Park S-H, *et al*. The paradoxical effects of TNF inhibitors on bone mineral density and radiographic progression in patients with ankylosing spondylitis. *Rheumatology* 2013;52:718–26.
37 Pedersen SJ, Chiowchanwisawakit P, Lambert RGW, *et al*. Resolution of inflammation following treatment of ankylosing spondylitis is associated with new bone formation. *J Rheumatol* 2011;38:1349–54.
38 Baraliakos X, Haibel H, Listing J, *et al*. Continuous long-term anti-TNF therapy does not lead to an increase in the rate of new bone formation over 8 years in patients with ankylosing spondylitis. *Ann Rheum Dis* 2014;73:710–5.
39 Haroon N, Inman RD, Learch TJ, *et al*. The impact of tumor necrosis factor α inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum* 2013;65:2645–54.
40 Kim T-J, Shin J-H, Kim S, *et al*. Radiographic progression in patients with ankylosing spondylitis according to tumor necrosis factor blocker exposure: observation study of Korean spondyloarthropathy registry (oskar) data. *Joint Bone Spine* 2016;83:569–72.
41 Pearl J. Interpretation and identification of causal mediation. *Psychol Methods* 2014;19:459–81.
42 Gensler LS, Gianfrancesco M, Weisman MH. OP0198 combined effects of tumour necrosis factor inhibitors and NSAIDs on radiographic progression in ankylosing spondylitis. *Annals of the Rheumatic Diseases* 2018;77:148.
43 Koo BS, Oh JS, Park SY, *et al*. Tumour necrosis factor inhibitors slow radiographic progression in patients with ankylosing spondylitis: 18-year real-world evidence. *Ann Rheum Dis* 2020;79:1327–32.
44 Molnar C, Scherer A, Baraliakos X, *et al*. Tnf blockers inhibit spinal radiographic progression in ankylosing spondylitis by reducing disease activity: results from the Swiss clinical quality management cohort. *Ann Rheum Dis* 2018;77:63–9.
45 Park JW, Kim MJ, Lee JS, *et al*. Impact of tumor necrosis factor inhibitor versus nonsteroidal antiinflammatory drug treatment on radiographic progression in early ankylosing spondylitis: its relationship to inflammation control during treatment. *Arthritis Rheumatol* 2019;71:82–90.
46 Sepriano A, Ramiro S, Wichuk S, *et al*. Tumor necrosis factor inhibitors reduce spinal radiographic progression in patients with radiographic axial spondyloarthritis: a longitudinal analysis from the Alberta prospective cohort. *Arthritis Rheumatol* 2021;73:1211–9.
47 de Koning A, de Bruin F, van den Berg R, *et al*. Low-Dose CT detects more progression of bone formation in comparison to conventional radiography in patients with ankylosing spondylitis: results from the SIAS cohort. *Ann Rheum Dis* 2018;77:293–9.
48 Landewé RBM, Ramiro S, Mostard RLM. COVID-19-induced hyperinflammation, immunosuppression, recovery and survival: how causal inference may help draw robust conclusions. *RMD Open* 2021;7:e001638.
49 Pearl J, MacKenzie D. *The book of why: the new science of cause and effect*. New York, USA, 2018.
50 Hernán MA, Robins JM. *Causal Inference: What If. Boca Raton: Chapman & Hall/CRC*, 2020. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

51 Braun J, Baraliakos X, Hermann K-GA, *et al*. The effect of two golimumab doses on radiographic progression in ankylosing spondylitis: results through 4 years of the GO-RAISE trial. *Ann Rheum Dis* 2014;73:1107–13.

52 Braun J, Baraliakos X, Deodhar A, *et al*. Secukinumab shows sustained efficacy and low structural progression in ankylosing spondylitis: 4-year results from the measure 1 study. *Rheumatology* 2019;58:859–68.