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 Desireé van der Heijde ▪,2 Robert Landewé ▪3,4

ABSTRACT

Axial spondyloarthritis (axSpA) is a chronic rheumatic disease characterised by inflammation predominantly involving the spine and the sacroiliac joints. 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.

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).

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

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.
alleviating symptoms, both in r-axSpA and nr-axSpA. 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. Here we use counterfactual reasoning as proposed by Rubin,20 Balke and Pearl21 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.

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 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).

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
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

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

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)=meant – meanc/pooled SD. Pooled SD=square root((SDT^2*(Nt-1) + SDc^2*(Nc-1))/(Nt +Nc-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.
with TNFi versus no TNFi after the fourth year of follow-up, adjusting for the mSASSS at baseline. 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.

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

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

**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.

‡NSAIDs not kept in the final model because it did not prove to confound the association of interest.

†If NSAID index ≥50.

adj, 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.
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. 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. 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 comorbidities. 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. Another study tested the mediating effect of the average value of CRP per interval. 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, 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.

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...
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. 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. In the absence of an RCT, however, causal inference from observational research can still be informative. 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 nodes) 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.

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.

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

<table>
<thead>
<tr>
<th>Study design</th>
<th>Exchangeability</th>
<th>Positivity</th>
<th>Consistency</th>
<th>Causality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal exposure</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>Very unlikely</td>
</tr>
<tr>
<td>No comparator</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal exposure</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>Unlikely</td>
</tr>
<tr>
<td>Comparator</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable exposure</td>
<td>+</td>
<td>+/-</td>
<td>–</td>
<td>Likely</td>
</tr>
<tr>
<td>Comparator</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

REFERENCES


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 column, and license their derivative works on different terms, provided the original work is available under the terms of the Creative Commons Attribution Non Commercial (CC BY-NC) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

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/0000-0002-6657-6620

RMD Open: first published as 10.1136/rmdopen-2021-001654 on 12 July 2021. Downloaded from http://rmdopen.bmj.com/ on July 22, 2023 by guest. Protected by copyright.


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*;21:35.


