ACR / SPARTAN / SAA

Ankylosing Spondylitis and Non-radiographic Axial Spondyloarthritis Treatment Guidelines

Evidence Report

This Evidence Report is comprised of data tables organized according to seven content domains, beginning with *Pharmacologic therapy*. Each table summarizes the data that are relevant to the previously developed "PICO" questions--i.e. clinical scenarios that specify a **P**atient population, a proposed **I**ntervention, a **C**omparison or alternate course of action, and an **O**utcome.

The tables include the number of studies that report the particular outcome, the study design employed in those studies (e.g. randomized trial), and the quality assessment of the data from these studies, according to ratings of members of the Literature Review Team. Four components were rated in this quality assessment: risk of bias in the studies, imprecision in the estimates of effect for the outcome, inconsistency among studies, indirectness (e.g., whether the study examined a similar, but distinct patient group or intervention), and publication bias. The table also reports on the number of patients in the Intervention and Comparison (Control) groups across all relevant studies, as well as the weighted relative and absolute effect size across all the relevant studies. This effect was often reported as the mean difference (MD) in the outcome between the intervention and comparison groups. All of these items were synthesized to produce an overall quality score (rated 1 to 4; ⊕OOO to ⊕⊕⊕⊕) for that particular outcome. Finally, an importance rating was applied to the outcome based on the priority assigned to the outcome in the outcome framework (manuscript, Table 1).

Table of Contents

Literature Review Team Members

Liron Caplan MD, PhD; Literature Review Lead
Andrew Lui, PT DPT
Joerg Ermann, MD
Lianne Gensler, MD
Judith Smith, MD, PhD
David Borenstein, MD
Jayme Hiratzka, MD
Pamela Weiss, MD, MSCE
Robert Inman, MD
Vikas Majithia, MD, MPH
Nigil Haroon, MBBS, MD, DM
Walter Maksymowych, MD, FRCPC
James Witter, MD, PhD
Robert Hart, MD
Siba Raychaudhuri, MD

Core Oversight Team

Michael Ward, MD, MPH; Principal Investigator
Atul Deodhar, MD, MRCP
John Reveille, MD
Elie Akl, MD, MPH, PhD

ACR Staff

Amy Miller; Senior Director, Quality
Regina Parker

Janet Joyce, MLS; Medical Librarian

Notes

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect; quality is rated 4 out of 4, represented as: ⊕⊕⊕⊕

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; quality is rated 3 out of 4, represented as: ⊕⊕⊕O

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; quality is rated 2 out of 4, represented as: ⊕⊕OO

Very low quality: We are very uncertain about the estimate; quality is rated 1 out of 4, represented as: ⊕OOO

Abbreviations

AE = Adverse Event
Amor = Amor criteria for the classification of spondylarthropathies. (Amor B, et al. Rev Rhum Mal Osteoartic. 1990;57:85–89)
APR = Acute Phase Reactant
ASDAS = Ankylosing Spondylitis Disease Activity Score
ASQOL = Ankylosing Spondylitis Quality of Life instrument
BASDAI = Bath Ankylosing Spondylitis Disease Activity Index
BASFI = Bath Ankylosing Spondylitis Functional Index
BASMI = Bath Ankylosing Spondylitis Metrology Index
CRP = C-reactive protein
DFI = Dougados Functional Index
ESSG = European Spondyloarthropathy Study Group
FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue
FEV = Forced expiratory volume
GI = gastrointestinal
IBD = inflammatory bowel disease
HHS = Harris Hip Score
MD = mean difference (the absolute difference between intervention and control groups or between baseline and final values for a measurement)
mNYCC = modified New York Classification Criteria for Ankylosing Spondylitis (van der Linden S, et al. Arthritis Rheum. 1984;27:361–368)
mSASSS = Modified Stoke Ankylosing Spondylitis Spinal Score
NC = Not Calculatable
NHP = Nottingham Health Profile
OR = Odds ratio
PICO = Patient/Intervention/Comparison/Outcome formatted question used in the GRADE system
PGART = Patient Global Assessment of Response to Therapy
RCTs = Randomized Clinical Trials
ROM = Range of Motion
SAARDs = Slow-acting antirheumatic drugs
SAE = Serious Adverse Event
SAPHO = synovitis, acne, pustulosis, hyperostosis, osteitis syndrome
SF-36 = Short form-36
URI = Upper respiratory infection
uSpA = undifferentiated spondyloarthritis
VAS = Visual analogue scale

PHARMACOLOGIC THERAPY

Non-steroidal Anti-Inflammatory Drugs (NSAIDS)

PICO 1. In adults with active or stable AS, is continuous treatment with NSAIDs more effective than on-demand treatment with NSAIDs in improving outcomes?

Summary: This PICO was directly addressed by 1 RCT, a 2-year open-label (unblinded) study (Wanders, 2005)[1]. There were no significant differences between groups in any clinical endpoint, with wide confidence intervals, and high risk of bias. The change in mSASSS was lower in the continuous treatment group. Hypertension and depression were more common in the continuous treatment group. One cohort study (Poddubnyy, 2012)[2] examined associations between high NSAID users (>50% on index of time and dose) versus low NSAID users for change in mSASSS but no clinical endpoints—data from that study are not included.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	continuous NSAIDs	Control: on-demand NSAID	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up mean 2 years; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	76	74	-	MD 6 lower (11.95 to 0.05 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Pain (follow-up mean 2 years; measured with: VAS; Better indicated by lower values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	76	74	-	MD 6 lower (12.59 lower to 0.59 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Fatigue (follow-up mean 2 years; measured with: Fatigue question on BASDAI; Better indicated by lower values)												
1	randomized trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	none	76	74	-	MD 5 lower (11.76 lower to 1.76 higher)	⊕OOO VERY LOW	CRITICAL
Health Status: Stiffness (follow-up mean 2 years; measured with: BASDAI 5+6; Better indicated by lower values)												
1	randomized trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	76	74	-	MD 5 lower (11.41 lower to 1.41 higher)	⊕⊕OO LOW	IMPORTANT
Health Status: Acute Phase Reactants (follow-up mean 2 years; measured with: CRP; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[9]	none	76	74	-	MD 3.7 lower (8.37 lower to 0.97 higher)	⊕⊕OO LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up mean 2 years; Better indicated by lower values)												
1	randomized trials	serious[10]	no serious inconsistency	no serious indirectness	serious[11]	none	76	74	-	MD 3 lower (9.76 lower to 3.76 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: ROM (follow-up mean 2 years; measured with: Chest expansion; Better indicated by lower values)												
1	randomized trials	serious[12]	no serious inconsistency	no serious indirectness	serious[13]	none	76	74	-	MD 0.2 lower (0.87 lower to 0.47 higher)	⊕⊕OO LOW	NOT IMPORTANT

Functional Status: Inflammation on Imaging (follow-up mean 2 years; measured with: mSASSS change; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[14]	none	76	74	-	MD 1.1 lower (1.79 to 0.41 lower)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Hypertension (follow-up mean 2 years)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[15]	none	10/111 (9%)	3/103 (2.9%)	OR 3.3 (0.88 to 12.35)	61 more per 1000 (from 3 fewer to 241 more)	⊕⊕⊕O MODERATE	IMPORTANT
Dyspepsia (follow-up mean 2 years)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	46/111 (41.4%)	39/103 (37.9%)	OR 1.16 (0.67 to 2.01)	35 more per 1000 (from 89 fewer to 172 more)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: Depression (follow-up mean 2 years)												
1	randomized trials	serious[16]	no serious inconsistency	no serious indirectness	serious[17]	none	15/111 (13.5%)	4/103 (3.9%)	OR 3.87 (1.24 to 12.07)	96 more per 1000 (from 9 more to 289 more)	⊕⊕OO LOW	IMPORTANT

[1-17]Wide confidence intervals and unblinded

PICO 1 includes RCT:	Wanders 2005[1]
PICO 1 includes Observational studies :	None

1. Wanders A, Heijde Dv, Landewe R, Behier JM, Calin A, Olivieri I, et al. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. Arthritis Rheum 2005;52:1756-65.
2. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Marker-Hermann E, Zeidler H, et al. Effect of non-steroidal anti-inflammatory drugs on radiographic spinal progression in patients with axial spondyloarthritis: results from the German Spondyloarthritis Inception Cohort. Ann Rheum Dis 2012;71:1616-22.

PICO 33. In adults with active or stable non-radiographic axial SpA, is continuous treatment with NSAIDs more effective than on-demand NSAID treatment in improving outcomes?

Summary: This PICO was directly addressed by 1 cohort study that examined associations between high NSAID users (>50% on index of time and dose) versus low NSAID users for change in mSASSS (Poddubnyy, 2012)[1]. No difference was found. No clinical endpoints were reported. 1 RCT (Wanders, 2005)[2], a 2-year open-label study addressed this question in patients with AS. There were no significant differences between groups in any clinical endpoint, with wide confidence intervals, and high risk of bias. The change in mSASSS was lower in the continuous treatment group. Hypertension and depression were more common in the continuous treatment group. .

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	continuous NSAIDs	Control: on demand NSAIDs	Relative (95% CI)	Absolute		
Functional Status: Inflammation on Imaging (follow-up mean 2 years; measured with: change in mSASSS; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	serious[3]	None	19	57	-	MD 0.2 higher (0.78 lower to 1.18 higher)	⊕OOO VERY LOW	NOT IMPORTANT

[1] Unblinded, not randomized
[2] NSAID index is not a measure of continuous use, only relative frequency of use
[3] Wide confidence intervals

PICO 33 includes RCT:	None
PICO 33 includes Observational studies :	Poddubnyy 2012[1]

1. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Marker-Hermann E, Zeidler H, et al. Effect of non-steroidal anti-inflammatory drugs on radiographic spinal progression in patients with axial spondyloarthritis: results from the German Spondyloarthritis Inception Cohort. Ann Rheum Dis 2012;71:1616-22.

2. Wanders A, Heijde Dv, Landewe R, Behier JM, Calin A, Olivieri I, et al. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. Arthritis Rheum 2005;52:1756-65.

PICO 2. In adults with active AS, are NSAIDs more effective than no treatment with NSAIDs in improving outcomes?

Summary: This PICO was directly addressed by 4 RCTs reported in 5 manuscripts and no relevant cohort studies or case control studies. Because endpoints were variably reported across studies, there were no more than 2-3 studies included for each endpoint with fewer studies included on the more relevant endpoints with only small to moderate effect sizes. There is good evidence NSAIDs improve symptoms of active AS (defined by pain, stiffness, or the BASDAI). All efficacy endpoints reported below favored the intervention.

Quality of Evidence Across All Critical Outcomes: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	NSAIDs	Placebo	Relative (95% CI)	Absolute		
Health Status: Pain (follow-up 6-12 weeks; measured with: VAS; 0-100; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	serious[2]	none	776	310	-	MD 17.06 lower (20.76 to 13.37 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Patient global (follow-up 6-12 weeks; measured with: VAS 0-100mm; range of scores: 0-100; Better indicated by lower values)												
3	randomized trials	serious[1]	no serious inconsistency (0%)	serious[3,4]	serious[2]	none	909	427	-	MD 18.36 lower (21.5 to 15.21 lower)	⊕OOO VERY LOW	IMPORTANT
Health Status: BASDAI (follow-up 6-12 weeks; range of scores: 0-100; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (0%)	serious[4]	no serious imprecision	none	529	240	-	MD 17.44 lower (20.72 to 14.16 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Spinal Pain (follow-up 6 weeks; measured with: VAS; 0-100; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[4]	no serious imprecision	none	97	93	-	MD 21.1 lower (27.47 to 14.73 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Stiffness (follow-up 6 weeks; measured with: VAS; 0-100; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	no serious inconsistency (0%)	serious[4]	serious[2,5]	none	495	290	-	MD 17.11 lower (22.93 to 11.28 lower)	⊕⊕OO LOW	IMPORTANT
Functional Status: ROM (follow-up 6 weeks; measured with Chest Expansion: cm; Better indicated by higher values)												
2	randomized trials	no serious risk of bias	no serious inconsistency (NC)	no serious indirectness	serious[2,5]	none	398	197	-	MD 0.44 higher (0.21 to 0.66 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: Sleep (follow-up 6 weeks; assessed with: 1-4 ordinal scale 3&4 = sleep disturbance)												
2	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	serious[2,5]	none	144/398 (36.2%)	197/250 (78.8%)	OR 0.01 (0 to 0.07)	752 fewer per 1000 (from 582 fewer to 788 fewer)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up 6-12 weeks; measured with: change in CRP; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (5%)	no serious indirectness	serious[2]	none	812	334	-	MD 3.63 lower (5.4 to 1.86 lower)	⊕⊕OO LOW	NOT IMPORTANT
Health Status: Night Pain (follow-up 6 weeks; measured with: VAS; 0-100; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	170	76	-	MD 16.61 lower (24.84 to 8.38 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: ROM (follow-up 6 weeks; measured with: cm (specific methodology not always specified - modified vs true Schober); Better indicated by higher values)												
3	randomized trials	no serious risk of bias	serious[6] (55%)	serious[4]	serious[2,5]	none	495	290	-	MD 0.34 higher (0.19 to 0.5 higher)	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up 6-12 weeks; range of scores: 0-100; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (NC)	serious[4]	no serious imprecision	none	681	306	-	MD 12.72 lower (15.61 to 9.83 lower)	⊕⊕OO LOW	CRITICAL

Functional Status: DFI (follow-up 6 weeks; measured with: 20 item scale using 1-3; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	228	121	-	MD 3.35 lower (4.75 to 1.95 lower)	⊕⊕⊕O MODERATE	CRITICAL
PGART (follow-up 6 weeks; assessed with: % responded scale 0-4)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[4]	no serious imprecision	none	55/97 (56.7%)	25/93 (26.9%)	OR 3.56 (1.94 to 6.55)	298 more per 1000 (from 147 more to 438 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Serious Adverse Event: GI bleeding (follow-up 6-52 weeks)												
2	randomized trials	serious[7]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	8/807 (0.99%)	0/277 (0%)	RR 2.86 (0.36 to 22.94)	-	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Event: all combined (follow-up 6-52 weeks)												
2	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	5/554 (0.9%)	2/249 (0.8%)	RR 0.86 (0.17 to 4.37)	1 fewer per 1000 (from 7 fewer to 27 more)	⊕⊕⊕⊕ HIGH	CRITICAL

[1] Benhamou 2010 is a post hoc analysis of 2 studies (1 of which was not otherwise included in this PICO because of missing SD/SE) [1]
[2] Dougados 1999 required splitting the placebo group to assess effect. [2]
[3] Patient global of disease activity is defined differently in studies.
[4] The main intervention comparison was etoricoxib, but a single arm was compared to naproxen and also to placebo. Therefore indirect comparison of naproxen vs placebo
[5] Required splitting the placebo group to assess effect
[6] Methodology of measurement not necessarily consistent across studies
[7] No prospective endoscopy
NC=not calculatable

PICO 2 includes RCT:	Dougados 1999[2]; Barkhuizen 2006[3]; Benhamou 2010[1]; Dougados 2001[4]; van der Heijde 2005[5]
PICO 2 includes Observational studies :	None

1. Benhamou M, Gossec L, Dougados M. Clinical relevance of C-reactive protein in ankylosing spondylitis and evaluation of the NSAIDs/coxibs' treatment effect on C-reactive protein. Rheumatology (Oxford) 2010;49:536-41.

2. Dougados M, Gueguen A, Nakache JP, Velicitat P, Veys EM, Zeidler H, et al. Ankylosing spondylitis: what is the optimum duration of a clinical study? A one year versus a 6 weeks non-steroidal anti-inflammatory drug trial. Rheumatology (Oxford) 1999;38:235-44.

3. Barkhuizen A, Steinfeld S, Robbins J, West C, Coombs J, Zwillich S. Celecoxib is efficacious and well tolerated in treating signs and symptoms of ankylosing spondylitis. J Rheumatol 2006;33:1805-12.

4. Dougados M, Behier JM, Jolchine I, Calin A, van der Heijde D, Olivieri I, et al. Efficacy of celecoxib, a cyclooxygenase 2-specific inhibitor, in the treatment of ankylosing spondylitis: a six-week controlled study with comparison against placebo and against a conventional nonsteroidal antiinflammatory drug. Arthritis Rheum 2001;44:180-5.

5. van der Heijde D, Baraf HS, Ramos-Remus C, Calin A, Weaver AL, Schiff M, et al. Evaluation of the efficacy of etoricoxib in ankylosing spondylitis: results of a fifty-two-week, randomized, controlled study. Arthritis Rheum 2005;52:1205-15.

PICO 34. In adults with active non-radiographic axial SpA, is treatment with NSAIDs more effective than no treatment with NSAIDs in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 3. In adults with active AS, are certain NSAIDs more effective than other NSAIDS in improving outcomes?

Summary: This PICO was directly addressed by 20 RCTs and no observational studies of comparative effectiveness. The comparator drug was **phenylbutazone** in 5 studies, however, it is no longer available in North America and, therefore, was excluded. **Indomethacin** was evaluated in 12 studies (versus naproxen, fenoprofen, aspirin, tolmetin, ibuprofen, etodolac, diclofenac, nabumetone, piroxicam, meclofenamate, sulindac, and flubiprofen). **Celecoxib** was compared to ketoprofen or diclofenac in 2 studies, with the diclofenac study having a non-inferiority design. One study compared flurbiprofen and **naproxen**. An evidence profile for each of these 3 agents (in bold) appears below. Small samples resulted in imprecise estimates of efficacy and reporting of blinding was poor. There was no evidence to suggest that indomethacin had different effects on pain or stiffness compared to other NSAIDs, nor of differences in efficacy between celecoxib and either diclofenac or ketoprofen. Celecoxib had lower rates of gastrointestinal side effects than its comparators, but rates were similar between other drugs in other trials.

Quality of Evidence Across All Critical Outcomes for **indomethacin**: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Indomethacin	Other NSAIDs	Relative (95% CI)	Absolute		
Health Status: Pain (follow-up median 6 weeks; measured with VAS or other; Better indicated by lower values)												
8	randomized trials	serious[1]	serious (55%)	no serious indirectness	no serious imprecision	none	444	434	-	MD 0.36 lower (1.06 lower to 0.34 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Stiffness (follow-up median 6 weeks; measured with: duration of stiffness; Better indicated by lower values)												
8	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	444	434	-	MD 0.16 lower (0.56 lower to 0.23 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up median 6 weeks; measured with: ESR; Better indicated by lower values)												
3	randomized trials	serious[1]	no serious inconsistency (NC)	no serious indirectness	serious[2]	reporting bias[3]	60	62	-	MD 1 lower (13.46 lower to 11.46 higher)	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: ROM (follow-up median 6 weeks; measured with: Schober's test; Better indicated by higher values)												
8	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	443	429	-	MD 0.38 higher (0.22 to 0.55 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Serious Adverse Event: all combined (follow-up median 6 weeks; assessed with: physician report)												
2	randomized trials	no serious risk of bias	no serious inconsistency (22%)	no serious indirectness	serious[2]	none	4/121 (3.3%)	7/121 (5.8%)	RR 0.65 (0.14 to 3.16)	23 fewer per 1000 (from 47 fewer to 48 more)	⊕⊕⊕O MODERATE	CRITICAL
Any GI side effects (follow-up median 6 weeks; assessed with: physician report)												
10	randomized trials	no serious risk of bias	no serious inconsistency (4%)	no serious indirectness	serious[2]	none	129/477 (27%)	90/357 (25.2%)	RR 0.95 (0.74 to 1.23)	25 fewer per 1000 (from 79 fewer to 41 more)	⊕⊕⊕O MODERATE	IMPORTANT

[1] randomization not completely described
[2] wide confidence interval
[3] data collected and not reported

Quality of Evidence Across All Critical Outcomes for **celecoxib**: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	celecoxib	Other NSAIDs	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	303	310	-	MD 0.31 higher (0.01 to 0.63 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain (follow-up median 12 weeks; measured with VAS; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (22%)	no serious indirectness	no serious imprecision	none	383	400	-	MD 0.07 higher (4.34 lower to 4.48 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness (follow-up median 12 weeks; measured with: duration of stiffness; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	80	90	-	MD 1 lower (36.71 lower to 34.71 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up median 12 weeks; measured with: CRP; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	383	400	-	MD 0.34 higher (1.13 lower to 1.82 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Functional Status: BASFI (follow-up median 12 weeks; Better indicated by lower values)												
2	randomized trials	serious[1]	serious (41%)	no serious indirectness	no serious imprecision	none	383	400	-	MD 0.02 higher (0.41 lower to 0.45 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: BASMI (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	303	310	-	MD 0.2 higher (0.01 lower to 0.41 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Serious Adverse Event: myocardial infarction (follow-up median 12 weeks; assessed with: physician reported)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	0/303 (0%)	2/310 (0.6%)	RR 0.34 (0.04 to 3.26)	4 fewer per 1000 (from 6 fewer to 14 more)	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Event: all combined (follow-up median 12 weeks; assessed with: physician reported)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	5/303 (1.7%)	4/310 (1.3%)	RR 1.28 (0.34 to 4.74)	4 more per 1000 (from 8 fewer to 46 more)	⊕⊕⊕O MODERATE	CRITICAL
Any GI side effect (follow-up median 12 weeks; assessed with: Physician reported)												
2	randomized trials	no serious risk of bias	serious (73%)	no serious indirectness	no serious imprecision	none	48/383 (12.5%)	85/400 (21.3%)	RR 0.56 (0.26 to 1.18)	87 fewer per 1000 (from 37 fewer to 124 fewer)	⊕⊕⊕O MODERATE	NOT IMPORTANT

[1] randomization not described
[2] wide confidence interval

Quality of Evidence Across All Critical Outcomes for **naproxen**: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Naproxen	Other NSAIDs	Relative (95% CI)	Absolute		
Health Status: Pain (follow-up median 2 weeks; measured with ordinal scale; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	29	29	-	not pooled	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness (follow-up median 2 weeks; measured with: duration of stiffness; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	29	29	-	not pooled	⊕⊕⊕O MODERATE	IMPORTANT
Any GI side effect (follow-up median 2 weeks; assessed with: physician report)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	5/29 (17.2%)	1/29 (3.4%)	OR 5.00 (0.62 to 40.20)	138 more per 1000 (from 12 fewer to 622 more)	⊕⊕OO LOW	NOT IMPORTANT

[1] randomization not described
[2] wide confidence interval

PICO 3 includes RCT:	Sieper 2008[1]; Dougados 2001[2]; Mena 1977[3]; Calin 1979[4]; Sydnes 1981[5]; Wasner 1981[6] Palferman 1991[7]; Calabro 1986[8]; Ebner 1983[9] Tannenbaum 1984[10]; Bacon 1990[11]; Burry 1980[12]; Franssen 1986[13]; Wordsworth 1980[14]; Ansell 1978[15]; Mena 1977[16]; Van Gerwen 1978[17]; Sturrock 1974[18]; Shipley 1980[19]; Gibson 1980[20]
PICO 3 includes Observational studies :	None

1. Sieper J, Klopsch T, Richter M, Kapelle A, Rudwaleit M, Schwank S, et al. Comparison of two different dosages of celecoxib with diclofenac for the treatment of active ankylosing spondylitis: results of a 12-week randomised, double-blind, controlled study. Ann Rheum Dis 2008;67:323-9.

2. Dougados M, Behier JM, Jolchine I, Calin A, van der Heijde D, Olivieri I, et al. Efficacy of celecoxib, a cyclooxygenase 2-specific inhibitor, in the treatment of ankylosing spondylitis: a six-week controlled study with comparison against placebo and against a conventional nonsteroidal antiinflammatory drug. Arthritis Rheum 2001;44:180-5.

3. Mena HR, Good AE. Management of ankylosing spondylitis with flurbiprofen or indomethacin. South Med J 1977;70:945-7.

4. Calin A, Britton M. Sulindac in ankylosing spondylitis. Double-blind evaluation of sulindac and indomethacin. JAMA 1979;242:1885-6.

5. Sydnes OA. Comparison of piroxicam with indomethacin in ankylosing spondylitis: a double-blind crossover trial. Br J Clin Pract 1981;35:40-4.

6. Wasner C, Britton MC, Kraines RG, Kaye RL, Bobrove AM, Fries JF. Nonsteroidal anti-inflammatory agents in rheumatoid arthritis and ankylosing spondylitis. JAMA 1981;246:2168-72.

7. Palferman TG, Webley M. A comparative study of nabumetone and indomethacin in ankylosing spondylitis. Eur J Rheumatol Inflamm 1991;11:23-9.

8. Calabro JJ. Efficacy of diclofenac in ankylosing spondylitis. Am J Med 1986;80:58-63.

9. Ebner W, Poal Ballarin JM, Boussina I. Meclofenamate sodium in the treatment of ankylosing spondylitis. Report of a European double-blind controlled multicenter study. Arzneimittelforschung 1983;33:660-3.

10. Tannenbaum H, DeCoteau WE, Esdaile JM. A double blind multicenter trial comparing piroxicam and indomethacin in ankylosing spondylitis with long-term follow-up. Curr Ther Res Clin Exp 1984;36:426-35.

11. Bacon PA. An overview of the efficacy of etodolac in arthritic disorders. Eur J Rheumatol Inflamm 1990;10:22-34.

12. Burry HC, Siebers R. A comparison of flurbiprofen with naproxen in ankylosing spondylitis. N Z Med J 1980;92:309-11.

13. Franssen MJ, Gribnau FW, van de Putte LB. A comparison of diflunisal and phenylbutazone in the treatment of ankylosing spondylitis. Clin Rheumatol 1986;5:210-20.

14. Wordsworth BP, Ebringer RW, Coggins E, Smith S. A double-blind cross-over trial of fenoprofen and phenylbutazone in ankylosing spondylitis. Rheumatol Rehabil 1980;19:260-3.

15. Ansell BM, Major G, Liyanage SP, Gumpel JM, Seifert MH, Mathews JA, et al. A comparative study of Butacote and Naprosyn in ankylosing spondylitis. Ann Rheum Dis 1978;37:436-9.

16. Mena HR, Willkens RF. Treatment of ankylosing spondylitis with flurbiprofen or phenylbutazone. Eur J Clin Pharmacol 1977;11:263-6.

17. Van Gerwen F, Van der Korst JK, Gribnau FW. Double-blind trial of naproxen and phenylbutazone in ankylosing spondylitis. Ann Rheum Dis 1978;37:85-8.

18. Sturrock RD, Hart FD. Double-blind cross-over comparison of indomethacin, flurbiprofen, and placebo in ankylosing spondylitis. Ann Rheum Dis 1974;33:129-31.

19. Shipley M, Berry H, Bloom B. A double-blind cross-over trial of indomethacin, fenoprofen and placebo in ankylosing spondylitis, with comments on patient assessment. Rheumatol Rehabil 1980;19:122-5.

20. Gibson T, Laurent R. Sulindac and indomethacin in the treatment of ankylosing spondylitis: a double-blind cross-over study. Rheumatol Rehabil 1980;19:189-92.

PICO 35. In adults with active non-radiographic axial SpA, are certain NSAIDs more effective than other NSAIDs in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Glucocorticoids

PICO 4. In adults with active AS, are systemic glucocorticoids more effective than no treatment with systemic glucocorticoids in improving outcomes?

Summary: This PICO was only assessed by one RCT of very short (2 week) duration that compared placebo to prednisolone 20 mg and prednisolone 50 mg (Haibel, 2014)[1]. It was also addressed by 3 case series with small numbers of subjects (between 12-15 patients for each outcome) and a median of 6 month follow-up. In the RCT, only 2 of 10 outcomes favored prednisolone 20 mg over placebo; 5 of 10 favored prednisolone 50 mg over placebo (4 of which are represented in the table below). For the case series, there were very modest differences (see footnote below table) attributed to intravenous systemic glucocorticoids and the studies demonstrated high risk of bias. An additional single RCT with small sample size compared low versus high dose glucocorticoids and was therefore not directly relevant.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Systemic Glucocoricoids	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up 2 weeks; range of scores: 0-10; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	13	-	MD 2.39 lower (1.38 to 3.4 lower)[1]	⊕⊕⊕⊕ HIGH	CRITICAL
Health Status: Pain (follow-up mean 4.5 months; range of scores: 0-100; Better indicated by lower values)												
2	observational studies	very serious[2]	no serious inconsistency	no serious indirectness	very serious[2]	none	15	-	-	MD 34 lower (unable to calculate CI)[3]	⊕OOO VERY LOW	CRITICAL
Health Status: Acute Phase Reactants (follow-up 2 weeks; measured with: CRP mg/L; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	13	-	MD 15.6 lower (8.1 to 23.1 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Functional Status: BASFI (follow-up 2 weeks; range of scores: 0-10; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	13	-	MD 1.76 lower (0.51 to 3.01 lower)[4]	⊕⊕⊕⊕ HIGH	CRITICAL
Functional Status: ROM – Schober's (follow-up median 6 months; measured with: modified Schober's; Better indicated by higher values)												
2	observational studies	serious[2]	no serious inconsistency	no serious indirectness	very serious[5]	none	12	-	-	mean 0.9 higher (unable to calculate CI)[3]	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: ROM – Finger-to-floor[3] (follow-up median 6 months; Better indicated by lower values)												
2	observational studies	serious[2]	no serious inconsistency	no serious indirectness	very serious[2]	none	12	-	-	MD 0.9 lower (unable to calculate CI)[3]	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: BASMI (follow-up 2 weeks; range of scores: 0-10; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	13	-	not pooled	⊕⊕⊕⊕ HIGH	NOT IMPORTANT

[1] Results significantly different between 50mg prednisolone and controls (p<0.03)
[2] Case series (I^2 not calculatable)
[3] Schober's improved by 0.9 cm; finger to floor improved by 13 cm; pain improved 34mm
[4] Results NOT significantly different between 50 mg prednisolone and controls (p=0.2)
[5] Unclear what method of measuring Schober's

PICO 4 includes RCT:	Haibel 2014[1]
PICO 4 includes Observational studies :	Ejstrup 1985[2], Richter 1983[3], Mintz 1981[4]

1. Haibel H, Fendler C, Listing J, Callhoff J, Braun J, Sieper J. Efficacy of oral prednisolone in active ankylosing spondylitis: results of a double-blind, randomised, placebo-controlled short-term trial. Ann Rheum Dis 2014;73:243-6.

2. Ejstrup L, Peters ND. Intravenous methylprednisolone pulse therapy in ankylosing spondylitis. Dan Med Bull 1985;32:231-3.

3. Richter MB, Woo P, Panayi GS, Trull A, Unger A, Shepherd P. The effects of intravenous pulse methylprednisolone on immunological and inflammatory processes in ankylosing spondylitis. Clin Exp Immunol 1983;53:51-9.

4. Mintz G, Enriquez RD, Mercado U. Intravenous methylprednisolone pulse therapy in severe ankylosing spondylitis. Arthritis Rheum 1981;24:734-6.

PICO 36. In adults with active non-radiographic axial SpA, are systemic glucocorticoids more effective than no treatment with systemic glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 13. In adults with AS and isolated active sacroiliitis despite treatment with NSAIDs, is treatment with locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was directly addressed by two small RCTs of poor quality. The RCTs used non-standardized outcomes and one was not blinded. The PICO was also addressed by 2 observational pre/post studies (n=34 total) with 18 month follow-up that consistently showed improvement of about 40 mm in a 0-100 mm pain scale lasting 9 months. Three additional observational studies included 51 AS patients and 44 uSpA patients. Results (which were not reported separately for AS) were very similar to the results of the RCTs (references not provided).

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	local GCC for sacroiliitis	Control: No GCC	Relative (95% CI)	Absolute		
Health Status: Pain (follow-up mean 1.5 months; range of scores: 0-100; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (NC)	serious[2]	serious[3]	None	11	13	-	MD 20 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: Pain at 9mo (follow-up mean 18 months; range of scores: 0-100; Better indicated by lower values)												
4	observational studies	very serious[1]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	None	85	-	-	mean 45 lower (unable to calculate CI)	⊕OOO VERY LOW	IMPORTANT

[1] small numbers; not blinded
[2] Met ESSG + Amor and specifies that pts have AS, but not clear that all patients meet mNYCC. Individuals with SAPHO excluded.
[3] Measure is non standardized

PICO 13 includes RCT:	Maugars 1996[1] Luukkainen 1999[2]
PICO 13 includes Observational studies :	Gunaydin 2006[3]; Migliore 2009[4]

1. Maugars Y, Mathis C, Berthelot J-M, Charlier C, Prost A. Assessment of the efficacy of sacroiliac corticosteroid injections in spondylarthropathies: A double-blind study. Br J Rheumatol 1996;35:767-70.

2. Luukkainen R, Nissila M, Asikainen E, Sanila M, Lehtinen K, Alanaatu A, et al. Periarticular corticosteroid treatment of the sacroiliac joint in patients with seronegative spondylarthropathy. Clin Exp Rheumatol 1999;17:88-90.

3. Gunaydin I, Pereira PL, Fritz J, Konig C, Kotter I. Magnetic resonance imaging guided corticosteroid injection of sacroiliac joints in patients with spondylarthropathy. Are multiple injections more beneficial? Rheumatol Int 2006;26:396-400.

4. Migliore A, Bizzi E, Massafra U, Vacca F, Martin-Martin LS, Granata M, et al. A new technical contribution for ultrasound-guided injections of sacro-iliac

joints. Eur Rev Med Pharmacol Sci 2010;14:465-9.

PICO 45. In adults with non-radiographic axial SpA and isolated active sacroiliitis despite treatment with NSAIDs, is treatment with locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by RCTs. Four observational studies included data on 44 uSpA, but results were only reported in aggregate with AS patients (see PICO 13 above), so the efficacy in nr-axSpA is not discernable.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 14. In adults with AS with stable axial disease and active enthesitis despite treatment with NSAIDs, are locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 46. In adults with non-radiographic axial SpA and active enthesitis despite treatment with NSAIDs, are locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 15. In adults with AS with stable axial disease and active peripheral arthritis despite treatment with NSAIDs, are locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 47. In adults with non-radiographic axial SpA and active peripheral arthritis despite treatment with NSAIDs, are locally administered parenteral glucocorticoids more effective than no treatment with local glucocorticoids in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Tumor Necrosis Factor Inhibitors (TNFi) and non-TNFi Biologics

PICO 5. In adults with active AS, are certain TNFi more effective than other TNFi in improving outcomes?

Summary: This PICO was directly addressed by 1 RCT (Giardina, 2010)[1], a 2-year open-label study of 50 patients randomized to either infliximab or etanercept. There were no differences between groups in point estimates of BASDAI or BASFI at 2 years; no confidence intervals were reported. Several observational studies report comparable short-term clinical effects with infliximab, etanercept, adalimumab, and similar drug survivals (data not shown). There were few comparisons that included golimumab and none included certolizumab. Four meta-analyses tested indirect comparisons using data from short-term RCTs; none found any TNFi to have higher ASAS20 responses than any other TNFi. Outcomes other than ASAS20 were not analyzed. Infliximab use was associated with lower rates of IBD flares (see PICO 32) but higher rates of tuberculosis than etanercept and adalimumab.

Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Infliximab	Control: Etanercept	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up mean 104 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1,2]	none	25	25	-	mean 0 higher (unable to calculate CI)[1]	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI (follow-up mean 104 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	25	25	-	MD 0 higher (unable to calculate CI)[1]	⊕⊕⊕O MODERATE	CRITICAL

[1] no confidence intervals provided
[2] small sample size

PICO 5 includes Meta-analysis:	Migliore 2012[2]; McLeod 2007[3]; Machado 2013[4]; Ren 2013[5]
PICO 5 includes RCT:	Giardina 2010[1]
PICO 5 includes Observational studies :	None

1. Giardina AR, Ferrante A, Ciccia F, Impastato R, Miceli MC, Principato A, et al. A 2-year comparative open label randomized study of efficacy and safety of etanercept and infliximab in patients with ankylosing spondylitis. Rheumatol Int 2010;30:1437-40.

2. Migliore A, Broccoli S, Bizzi E, Lagana B. Indirect comparison of the effects of anti-TNF biological agents in patients with ankylosing spondylitis by means of a mixed treatment comparison performed on efficacy data from published randomised, controlled trials. J Med Econ 2012;15:473-80.

3. McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, et al. Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation. Health Technol Assess 2007;11:1-158.

4. Machado MA, Barbosa MM, Almeida AM, de Araujo VE, Kakehasi AM, Andrade EI, et al. Treatment of ankylosing spondylitis with TNF blockers: a meta-analysis. Rheumatol Int 2013;33:2199-213.

5. Ren L, Li J, Luo R, Tang R, Zhu S, Wan L. Efficacy of antitumor necrosis factor(alpha) agents on patients with ankylosing spondylitis. Am J Med Sci 2013;346:455-61.

PICO 37. In adults with active non-radiographic axial SpA, are certain TNFi more effective than other TNFi in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 6. In adults with active AS despite treatment with NSAIDs, are TNFi more effective than no treatment with TNFi in improving outcomes?

Summary: This PICO was directly addressed by 13 RCTs; therefore we did not rely upon observational data. We were able to aggregate efficacy data for four TNFi's (adalimumab, etanercept, infliximab, and golimumab) using data from 10 RCTs. Sample sizes (including both intervention and control arms) ranged from 44 to 566 subjects. The efficacy outcomes consistently favored TNFi over placebo and traditional slow acting anti-rheumatic drugs (SAARDs) across outcomes. In 9 RCTs with available data, adverse events did not differ between TNFi exposed and placebo-treated patients for short term outcomes. There was limited opportunity to aggregate data beyond 6 months because of cross-over designs and variability in reporting. Data quality for efficacy was very high, but only moderate for safety data.

Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	TNFi	Control: No TNFi	Relative (95% CI)	Absolute		
Mortality												
2	randomized trials	no serious risk of bias	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	0/608 (0%)	0/302 (0%)	-	-	⊕⊕⊕⊕ HIGH	CRITICAL
Health Status: BASDAI (follow-up 12 weeks; Better indicated by lower values)												
6	randomized trials	no serious risk of bias	serious (31%)	no serious indirectness	no serious imprecision	none	552	338	-	MD 1.35 lower (1.72 to 0.98 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: BASDAI (follow-up 24 weeks; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	serious[2] (82%)	no serious indirectness	no serious imprecision	none	270	166	-	MD 0.88 lower (2.11 lower to 0.35 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain (follow-up 12 weeks; Better indicated by lower values)												
4	randomized trials	no serious risk of bias	serious (99%)	no serious indirectness	no serious imprecision	none	954	548	-	MD 20.73 lower (29.75 to 11.71 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: ASDAS (follow-up 12 weeks; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	serious (39%)	no serious indirectness	no serious imprecision	none	292	442	-	MD 1.28 lower (1.55 to 1.01 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up 12 weeks; Measured by CRP; Better indicated by lower values)												
7	randomized trials	no serious risk of bias	serious (63%)	no serious indirectness	no serious imprecision	none	938	529	-	MD 11.1 lower (13.94 to 8.26 lower)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Functional Status: BASFI (follow-up 12 weeks; Better indicated by lower values)												
5	randomized trials	no serious risk of bias	serious[2] (75%)	no serious indirectness	no serious imprecision	none	394	285	-	MD 1.33 lower (2.27 to 0.38 lower)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI (follow-up 24 weeks; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	serious[2] (82%)	no serious indirectness	no serious imprecision	none	270	166	-	MD 0.88 lower (2.11 lower to 0.35 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASMI at 12 weeks (Better indicated by lower values)												
7	randomized trials	no serious risk of bias	no serious inconsistency (18%)	no serious indirectness	no serious imprecision	none	1036	629	-	MD 0.35 lower (0.46 to 0.23 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Continued on next page												

Serious Adverse Event: myocardial infarction												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	1/45 (2.2%)	0/39 (0%)	RR 2.61 (0.11 to 62.26)	-	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Event: serious infections												
5	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	serious[1]	none	5/1296 (0.39%)	2/561 (0.36%)	RR 0.71 (0.17 to 2.99)	1 fewer per 1000 (from 3 fewer to 7 more)	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Event: life threatening cancer												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	1/38 (2.6%)	0/39 (0%)	RR 3.08 (0.13 to 73.26)	-	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Event: serious neurologic disease												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/379 (0%)	0/187 (0%)	not pooled	not pooled	⊕⊕⊕⊕ HIGH	CRITICAL
Serious Adverse Event: all combined												
6	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	36/1312 (2.7%)	15/572 (2.6%)	RR 0.91 (0.5 to 1.66)	2 fewer per 1000 (from 13 fewer to 17 more)	⊕⊕⊕⊕ HIGH	CRITICAL

[1] Definition of serious infection unclear
[2] high variation in this outcome for control group
[3] wide confidence intervals

PICO 6 includes RCT:	Hu 2012[1]; Huang 2013[2]; Braun 2002[3]; van der Heijde 2006[4]; Inman 2010[5]; Brandt 2003[6]; Dougados 2011[7]; Gorman 2002[8]; Braun 2011[9]; Calin 2004; [10];van der Heijde 2005[11]; Barkham 2010[12]; Bao 2014[13]
PICO 6 includes Observational studies :	None included

1. Hu Z, Xu M, Li Q, Lin Z, Liao Z, Cao S, et al. Adalimumab significantly reduces inflammation and serum DKK-1 level but increases fatty deposition in lumbar spine in active ankylosing spondylitis. Int J Rheum Dis 2012;15:358-65.

2. Huang F, Gu J, Zhu P, Bao C, Xu J, Xu H, et al. Efficacy and safety of adalimumab in Chinese adults with active ankylosing spondylitis: results of a randomised, controlled trial. Ann Rheum Dis 2014;73:587-94.

3. Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W, et al. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 2002;359:1187-93.

4. van der Heijde D, Kivitz A, Schiff MH, Sieper J, Dijkmans BA, Braun J, et al. Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2006;54:2136-46.

5. Inman RD, Maksymowych WP. A double-blind, placebo-controlled trial of low dose infliximab in ankylosing spondylitis. J Rheumatol 2010;37:1203-10.

6. Brandt J, Khariouzov A, Listing J, Haibel H, Sorensen H, Grassnickel L, et al. Six-month results of a double-blind, placebo-controlled trial of etanercept treatment in patients with active ankylosing spondylitis. Arthritis Rheum 2003;48:1667-75.

7. Dougados M, Braun J, Szanto S, Combe B, Elbaz M, Geher P, et al. Efficacy of etanercept on rheumatic signs and pulmonary function tests in advanced ankylosing spondylitis: results of a randomised double-blind placebo-controlled study (SPINE). Ann Rheum Dis 2011;70:799-804.

8. Gorman JD, Sack KE, Davis JC, Jr. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. N Engl J Med 2002;346:1349-56.

9. Braun J, van der Horst-Bruinsma IE, Huang F, Burgos-Vargas R, Vlahos B, Koenig AS, et al. Clinical efficacy and safety of etanercept versus sulfasalazine in patients with ankylosing spondylitis: a randomized, double-blind trial. Arthritis Rheum 2011;63:1543-51.

10. Calin A, Dijkmans BA, Emery P, Hakala M, Kalden J, Leirisalo-Repo M, et al. Outcomes of a multicentre randomised clinical trial of etanercept to treat ankylosing spondylitis. Ann Rheum Dis 2004;63:1594-600.

11. van der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K, Williamson P, et al. Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT). Arthritis Rheum 2005;52:582-91.

12. Barkham N, Coates LC, Keen H, Hensor E, Fraser A, Redmond A, et al. Double-blind placebo-controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. Ann Rheum Dis 2010;69:1926-8.

13. Bao C, Huang F, Khan MA, Fei K, Wu Z, Han C, et al. Safety and efficacy of golimumab in Chinese patients with active ankylosing spondylitis: 1-year results of a multicentre, randomized, double-blind, placebo-controlled phase III trial. Rheumatology (Oxford) 2014;53:1654-63.

PICO 38. In adults with active non-radiographic axial SpA despite treatment with NSAIDs, are TNFi more effective than no treatment with TNFi in improving outcomes?

Summary: This PICO was directly addressed by 5 RCTs and 2 observational studies. The RCTs examined the efficacy of TNFi (adalimumab, infliximab, certolizumab, and etanercept) in patients with non-radiographic axial SpA with samples ranging from 40 to 185 patients and results were reported at 12 to 48 weeks. The results were consistently in favor of TNFi over placebo or sulfasalazine for clinical outcomes and imaging (MRI) results. The magnitude of effect was imprecise for most outcomes, and 1 trial included an unknown number of patients with AS. Two short term observational studies also reported improvement in clinical outcomes with treatment.
Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	TNFi	Control: No TNFi	Relative (95% CI)	Absolute		
Health Status: BASDAI (measured with: 0-10 scale; Better indicated by lower values)												
5	randomized trials	no serious risk of bias	no serious inconsistency (88%)	serious[1]	no serious imprecision	none	336	297	-	MD 1.36 lower (2.5 to 0.21 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain (measured with: 0-10 NRS scale; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	219	227	-	MD 0.9 lower (0.98 to 0.82 lower)	⊕⊕⊕⊕ HIGH	CRITICAL
Health Status: Stiffness (Better indicated by lower values)												
2	randomized trials	no serious risk of bias	serious (68%)	no serious indirectness	serious[2]	none	128	133	-	MD 1.34 lower (2.55 to 0.13 lower)	⊕⊕OO LOW	IMPORTANT
Health Status: Joint Counts (measured with: Swollen Joint Count; Better indicated by lower values)												
2	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	128	133	-	MD 0.2 higher (0.16 to 0.24 higher)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: ASDAS (Better indicated by lower values)												
2	randomized trials	no serious risk of bias	no serious inconsistency (NC)	serious[1]	no serious imprecision	none	188	144	-	MD 5 lower (5.28 to 4.72 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (measured with: CRP (mg/L); Better indicated by lower values)												
4	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	239	247	-	MD 3.1 lower (3.38 to 2.82 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Health Status: Inflammation on Imaging (measured with: Sacroiliac joint MRI score; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	217	223	-	MD 3.8 lower (3.97 to 3.63 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Health Status: SF-36_mental (measured with: SF36 MCS; Better indicated by higher values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	22	24	-	MD 0.7 higher (6.4 lower to 7.8 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: ASQOL (Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	20	20	-	not pooled	⊕⊕⊕O MODERATE	CRITICAL

Health Status: BASFI (measured with: 0-10 scale; Better indicated by lower values)												
5	randomized trials	no serious risk of bias	no serious inconsistency (89%)	serious[1]	no serious imprecision	none	336	297	-	MD 1.31 lower (2.56 to 0.07 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: HAQ-S (measured with: HAQ or HAQ-S; Better indicated by lower values)												
2	randomized trials	no serious risk of bias	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	111	114	-	not pooled	⊕⊕⊕⊕ HIGH	CRITICAL
Functional Status: SF-36_physical (measured with: SF36 PCS; Better indicated by higher values)												
2	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	113	118	-	MD 3.9 higher (2.35 lower to 10.15 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: ROM (measured with: Chest Expansion; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	20	20	-	not pooled	⊕⊕⊕O MODERATE	NOT IMPORTANT
Functional Status: BASMI (Better indicated by lower values)												
4	randomized trials	no serious risk of bias	serious[3] (90%)	serious[1]	no serious imprecision	none	316	277	-	MD 0.33 lower (0.86 lower to 0.21 higher)	⊕⊕OO LOW	NOT IMPORTANT
Serious Adverse Events: serious infection												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	0/111 (0%)	1/113 (0.9%)	RR 0.34 (0.01 to 8.24)	6 fewer per 1000 (from 9 fewer to 64 more)	⊕⊕⊕O MODERATE	CRITICAL
Serious Adverse Events: all serious combined												
4	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	serious[2]	none	5/248 (2%)	3/254 (1.2%)	RR 1.63 (0.38 to 7.09)	7 more per 1000 (from 7 fewer to 72 more)	⊕⊕⊕O MODERATE	CRITICAL

[1] Some studies included patients with AS
[2] Confidence intervals wide
[3] Effect differs among studies

PICO 38 includes RCT:	Sieper 2013[1]; Landewe 2014[2]; Barkham 2009[3]; Haibel 2008[4]; Dougados 2014[5]
PICO 38 includes Observational studies :	Brandt 2004[6]; Gerard 2008[7]

1. Sieper J, van der Heijde D, Dougados M, Mease PJ, Maksymowych WP, Brown MA, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis 2013;72:815-22.

2. Landewe R, Braun J, Deodhar A, Dougados M, Maksymowych WP, Mease PJ, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. Ann Rheum Dis 2014;73:39-47.

3. Barkham N, Keen HI, Coates LC, O'Connor P, Hensor E, Fraser AD, et al. Clinical and imaging efficacy of infliximab in HLA-B27-Positive patients with magnetic resonance imaging-determined early sacroiliitis. Arthritis Rheum 2009;60:946-54.

4. Haibel H, Rudwaleit M, Listing J, Heldmann F, Wong RL, Kupper H, et al. Efficacy of adalimumab in the treatment of axial spondylarthritis without radiographically defined sacroiliitis: results of a twelve-week randomized, double-blind, placebo-controlled trial followed by an open-label extension up to week fifty-two. Arthritis Rheum 2008;58:1981-91.

5. Dougados M, van der Heijde D, Sieper J, Braun J, Maksymowych WP, Citera G, et al. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol 2014;66:2091-102.

6. Brandt J, Khariouzov A, Listing J, Haibel H, Sorensen H, Rudwaleit M, et al. Successful short term treatment of patients with severe undifferentiated spondyloarthritis with the anti-tumor necrosis factor-alpha fusion receptor protein etanercept. J Rheumatol 2004;31:531-8.

7. Gerard S, Le Goff B, Maugars Y, Berthelot JM. Six-month response to anti-TNF drugs in axial spondylarthropathy according to the fulfillment or not of New-York criteria for ankylosing spondylitis or French recommendations for anti-TNF use. A "real life" retrospective study on 175 patients. Joint Bone Spine 2008;75:680-7.

PICO 8. In adults with active AS despite treatment with NSAIDs and who have contraindications to TNFi, is treatment with a non-TNFi biologic more effective than treatment with SAARDs in improving outcomes?

Summary: This PICO was not directly addressed by any RCTs or observational studies comparing non-TNFi biologics to SAARDs in patients with TNFi contraindications (including TNFi non-responders). The only way to compare non-TNFi biologics to SAARDs would be to qualitatively compare the outcomes of trials of non-TNFi biologics to the outcomes for PICO 7, which is a fairly indirect comparison. We report the outcomes below for two pre/post studies of abatacept (Lekpa 2012[1]; Song 2011[2]), which failed to show any obvious benefit across outcomes at 24 weeks. We also report results for an ustekinumab pre/post study (Poddubnyy 2014[3]); however patients were TNFi naïve), a tocilizumab RCT and pre/post study (Sieper 2014[4] and Lekpa 2012[5]), and a rituximab pre/post study (Song 2010[6]). None of these studies demonstrated significant benefits across endpoints.

Quality of Evidence Across All Critical Outcomes for **abatacept**: Very Low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Abatacept	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up 3 months; range of scores: 0-10; Better indicated by lower values)												
2	observational studies	very serious[1]	no serious inconsistency (NC)	very serious[2]	no serious imprecision	none	20	-	-	MD .3 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: Pain (follow-up 3 months; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	5	-	-	MD 0.02 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: ROM – Schober's test (cm) (follow-up 3 months; Better indicated by higher values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	5	-	-	not pooled	⊕OOO VERY LOW	IMPORTANT
Health Status: ASDAS (follow-up 3 months; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	15	-	-	MD 0.1 higher (unable to calculate CI)	⊕OOO VERY LOW	IMPORTANT
Health Status: Acute Phase Reactants - CRP (mg/L) (follow-up 3 months; Better indicated by lower values)												
2	observational studies	very serious[1]	no serious inconsistency (NC)	very serious[2]	no serious imprecision	none	20	-	-	not pooled	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up 3 months; range of scores: 0-10; Better indicated by lower values)												
2	observational studies	very serious[1]	no serious inconsistency (NC)	very serious[2]	no serious imprecision	none	20	-	-	not pooled	⊕OOO VERY LOW	CRITICAL
Functional Status: BASMI (follow-up 3 months; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	15	-	-	not pooled	⊕OOO VERY LOW	NOT IMPORTANT

[1] Small sample size; no control
[2] Indirect comparison: does not directly address non-TNFi versus SAARD

Quality of Evidence Across All Critical Outcomes for **ustekinumab**: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ustekinumab	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	20	-	-	MD 2.3 lower (5.3 lower to 1.3 higher)[3]	⊕⊕OO LOW	CRITICAL
Health Status: Pain (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	20	-	-	MD 3.2 lower (5.6 to 0.8 lower)	⊕⊕OO LOW	CRITICAL
Health Status: ASDAS (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	20	-	-	MD 1 lower (3 lower to 1.2 higher)	⊕⊕OO LOW	IMPORTANT
Health Status: Acute Phase Reactants - CRP (mg/L) (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	20	-	-	MD 0.5 higher (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT
Health Status: Inflammation on Imaging (follow-up 24 weeks; measured with: MRI-sacroiliac osteitis score; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	17	-	-	MD 2.2 lower (5.4 lower to 4.6 higher)	⊕OOO VERY LOW	NOT IMPORTANT
Health Status: ASQOL (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	20	-	-	MD 4.3 lower (9.4 lower to 3.7 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: BASFI (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	20	-	-	MD 2.3 lower (5.3 lower to 2.3 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: BASMI (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20	-	-	MD 0.4 lower (1.6 lower to 2.2 higher)	⊕OOO VERY LOW	NOT IMPORTANT
Health Status: ASAS40 (follow-up 24 weeks)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	13/20 (65%)	-	41 to 85	-	⊕⊕OO LOW	NOT IMPORTANT
Health Status: BASDAI50 (follow-up 24 weeks)												
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	strong association[2]	11/20 (55%)	-	32 to 77	-	⊕⊕OO LOW	NOT IMPORTANT

[1] Observational study of 20 subjects with 3 dropouts for lack of effect may indicate bias
[2] Large effect seen or p<0.001
[3] 95% CI not available. Rough estimate: 2xSD to give range
[4] Large SD
[5] large SD, p=0.026

Quality of Evidence Across All Critical Outcomes for **rituximab**: Very Low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Rituximab	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI in TNFi_naïve (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	mean 2.0 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: BASDAI in TNFi exposed (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	mean 0.9 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: Acute Phase Reactants - CRP (mg/L)_TNFi_naïve (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 5.5 lower (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT
Health Status: Acute Phase Reactants - CRP (mg/L) _TNFi_exposed (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 1.4 lower (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT
Health Status: ASQOL_TNFi_naïve (follow-up 24 weeks; range of scores: 0-18; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 3.3 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: ASQOL_TNFi_exposed (follow-up 24 weeks; range of scores: 0-18; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 3.1 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Functional Status: BASFI_TNFi_naïve (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 1.3 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Functional Status: BASFI_TNFi_exposed (follow-up 24 weeks; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD 0.5 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Functional Status: BASMI_TNFi_naïve (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD .4 lower (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: BASMI_TNFi_exposed (follow-up 24 weeks; Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	10	-	-	MD .3 lower (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT

[1] No control
[2] Indirect comparison: does not directly address non-TNFi versus SAARD

Quality of Evidence Across All Critical Outcomes for **tocilizumab**: Very Low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Tocilizumab	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up 3 months; range of scores: 0-10; Better indicated by lower values)												
1	observational studies	very serious[3]	serious[1]	no serious indirectness	no serious imprecision	none	8	-	-	MD 0.3 lower (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: Pain (follow-up 3 months; range of scores: 0-100; Better indicated by lower values)												
1	observational studies	very serious	no serious inconsistency	no serious indirectness	no serious imprecision	none	8	-	-	MD 2.9 higher (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL
Health Status: ASDAS (follow-up 12 weeks; range of scores: 0-10; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	51	51	-	mean 0.9 lower (unable to calculate CI)[2]	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants - CRP (mg/L)_ (follow-up 3 months; Better indicated by lower values)												
1	observational studies	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7	-	-	mean 0 higher (unable to calculate CI)	⊕OOO VERY LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up 3 months; Better indicated by lower values)												
1	observational studies	very serious[3]	serious[1]	no serious indirectness	no serious imprecision	none	5	-	-	MD 0.4 higher (unable to calculate CI)	⊕OOO VERY LOW	CRITICAL

[1] highly variable patient responses
[2] SD for both controls and intervention were ~0.9 but no formal test for significance reported
[3] No controls

PICO 8 includes RCT:	Sieper 2014[4]
PICO 8 includes Observational studies :	Poddubnyy 2014[3]; Lekpa 2012[1]; Song 2010[6]; Lekpa 2012[5]; Song 2011[2]

1. Lekpa FK, Farrenq V, Canoui-Poitrine F, Paul M, Chevalier X, Bruckert R, et al. Lack of efficacy of abatacept in axial spondylarthropathies refractory to tumor-necrosis-factor inhibition. Joint Bone Spine 2012;79:47-50.

2. Song IH, Heldmann F, Rudwaleit M, Haibel H, Weiss A, Braun J, et al. Treatment of active ankylosing spondylitis with abatacept: an open-label, 24-week pilot study. Ann Rheum Dis 2011;70:1108-10.

3. Poddubnyy D, Hermann KG, Callhoff J, Listing J, Sieper J. Ustekinumab for the treatment of patients with active ankylosing spondylitis: results of a 28-week, prospective, open-label, proof-of-concept study (TOPAS). Ann Rheum Dis 2014;73:817-23.

4. Sieper J, Porter-Brown B, Thompson L, Harari O, Dougados M. Assessment of short-term symptomatic efficacy of tocilizumab in ankylosing spondylitis: results of randomised, placebo-controlled trials. Ann Rheum Dis 2014;73:95-100.

5. Lekpa FK, Poulain C, Wendling D, Soubrier M, De BM, Berthelot JM, et al. Is IL-6 an appropriate target to treat spondyloarthritis patients refractory to anti-TNF therapy? A multicentre retrospective observational study. Arthritis Res Ther 2012;14:R53.

6. Song IH, Heldmann F, Rudwaleit M, Listing J, Appel H, Braun J, et al. Different response to rituximab in tumor necrosis factor blocker-naive patients with active ankylosing spondylitis and in patients in whom tumor necrosis factor blockers have failed: a twenty-four-week clinical trial. Arthritis Rheum 2010;62:1290-.

PICO 40. In adults with active non-radiographic axial SpA despite treatment with NSAIDs and who have contraindications to TNFi, is treatment with a non-TNFi biologic more effective than treatment with SAARDs in improving outcomes?

Summary: This PICO was not directly addressed by any RCTs or observational studies comparing non-TNFi biologics to SAARDs in patients with TNFi contraindications (including TNFi non-responders). The only way to compare non-TNFi biologics to SAARDs would be to qualitatively compare the outcomes for PICO 39, which is a fairly indirect comparison. Data were only available for 5 patients (Lepka 2012[1]; n=3 treated with tocilizumab and Lekpa 2012[2]; n=2 treated with abatacept). Results are not presented in an evidence profile, as results were highly variable for these 5 patients.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

1. Lekpa FK, Poulain C, Wendling D, Soubrier M, De BM, Berthelot JM, et al. Is IL-6 an appropriate target to treat spondyloarthritis patients refractory to anti-TNF therapy? A multicentre retrospective observational study. Arthritis Res Ther 2012;14:R53.

2. Lekpa FK, Farrenq V, Canoui-Poitrine F, Paul M, Chevalier X, Bruckert R, et al. Lack of efficacy of abatacept in axial spondylarthropathies refractory to tumor-necrosis-factor inhibition. Joint Bone Spine 2012;79:47-50.

PICO 9. In adults with active AS despite treatment with the first TNFi agent used, is switching to a different TNFi more effective than adding a SAARD in improving outcomes?

Summary: This PICO was not directly addressed by any RCTs; 3 observational studies provided some relevant data regarding the efficacy of TNFi switching compared with baseline or those remaining on their initial TNFi, but these were not compared directly to patients who switched to SAARDs. The results showed some improvement versus baseline, but outcomes were not as positive compared to patients who did not switch.

Quality of Evidence Across All Critical Outcomes: Very Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Switching TNFi	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up mean 3 months; range of scores: 0-10; Better indicated by lower values)												
3	observational studies	very serious[1]	serious[2] (NC)	no serious indirectness	no serious imprecision	none	532	1441	-	not pooled	⊕OOO VERY LOW	CRITICAL
Health Status: Pain (follow-up mean 3 months; range of scores: 0-10; Better indicated by lower values)												
3	observational studies	very serious[1]	serious[2] (NC)	no serious indirectness	no serious imprecision	none	531	1441	-	not pooled	⊕OOO VERY LOW	CRITICAL
Health Status: Acute Phase Reactants - CRP (follow-up mean 9 months; Better indicated by lower values)												
2	observational studies	very serious[1]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	532	1441	-	not pooled[4]	⊕⊕OO LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up mean 3 months; range of scores: 0-10; Better indicated by lower values)												
2	observational studies	very serious[3]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	100	437	-	not pooled	⊕⊕OO LOW	CRITICAL
Functional Status: BASMI (follow-up mean 3 months; range of scores: 0-10; Better indicated by lower values)												
2	observational studies	very serious[1]	very serious[4] (NC)	no serious indirectness	no serious imprecision	none	455	1004	-	not pooled	⊕OOO VERY LOW	NOT IMPORTANT

[1] 2 studies compare with non-switchers and initial TNFi response; 1 study compares to baseline (time of switch); not clear that these pts would have taken SAARDs
[2] compared with non-switchers and initial TNFi switchers are worse; compared with baseline (time of switch), switchers are better
[3] 1 study compares with non-switchers; 1 study compares to baseline (time of switch); not clear that these pts would have taken SAARDs
[4] wide variation in outcomes

PICO 9 includes RCT:	None
PICO 9 includes Observational studies :	Lie 2011[1]; Cantini 2006[2]; Glintborg 2013[3]

1. Lie E, van der Heijde D, Uhlig T, Mikkelsen K, Rodevand E, Koldingsnes W, et al. Effectiveness of switching between TNF inhibitors in ankylosing spondylitis: data from the NOR-DMARD register. Ann Rheum Dis 2011;70:157-63.

2. Cantini F, Niccoli L, Benucci M, Chindamo D, Nannini C, Olivieri I, et al. Switching from infliximab to once-weekly administration of 50 mg etanercept in resistant or intolerant patients with ankylosing spondylitis: results of a fifty-four-week study. Arthritis Rheum 2006;55:812-6.

3. Glintborg B, Ostergaard M, Krogh NS, Tarp U, Manilo N, Loft AG, et al. Clinical response, drug survival and predictors thereof in 432 ankylosing spondylitis patients after switching tumour necrosis factor alpha inhibitor therapy: results from the Danish nationwide DANBIO registry. Ann Rheum Dis 2013;72:1149-55.

PICO 41. In adults with active non-radiographic axial SpA despite treatment with the first TNFi agent used, is switching to a different TNFi more effective than adding a SAARD in improving outcomes?

Summary: This PICO was not directly addressed by any RCTs; 1 small observational study (Delaunay 2005)[1] reported on 6 patients who switched from etanercept to infliximab). All 6 responded, however the magnitude of change was very imprecise.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

1. Delaunay C, Farrenq V, Marini-Portugal A, Cohen JD, Chevalier X, Claudepierre P. Infliximab to etanercept switch in patients with spondyloarthropathies and psoriatic arthritis: preliminary data. J Rheumatol 2005;32:2183-5.

PICO 10. In adults with active AS despite treatment with the first TNFi agent used, is switching to a different TNFi more effective than switching to non-TNFi biologics in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 42. In adults with active non-radiographic axial SpA despite treatment with the first TNFi agent used, is switching to a different TNFi more effective than switching to non-TNFi biologics in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Slow-Acting Anti-Rheumatic Drugs (SAARDS)

PICO 7. In adults with active AS despite treatment with NSAIDs, are SAARDs more effective than no treatment with SAARDs in improving outcomes?

Summary: This PICO was addressed by 15 RCTs. Comparator drugs were sulfasalazine in 8 trials, methotrexate in 3 trials, leflunomide in 1 trial, pamidronate in 1 trial, thalidomide in 1 trial, and apremilast in 1 trial.

- The trials that examined the effect of **sulfasalazine** were all performed before 1996 and thus before the development of contemporary composite scores. Outcome measures were diverse, which precluded the pooling of data for meta-analysis in many instances. Sulfasalazine had a weak beneficial effect on spinal pain but not on other critical outcome measures, other than poorly defined "episodes of joint symptoms (arthritis or peri-arthritis)" and ad-hoc "composite peripheral joint scores" (per Kirwan 1993[1] and Clegg 1996[2]). These peripheral joint scores favored sulfasalazine despite no difference in actual tender/swollen joint counts.
- The three studies that compared **methotrexate** with placebo used weekly doses of 10 mg or less. There was no benefit over placebo for any critical outcomes. A dose of 10 mg weekly is likely suboptimal. However, a cohort study (Haibel et al. Ann Rheum Dis 2007[3]) analyzed the efficacy of 20 mg weekly in AS and similarly failed to detect significant benefit.
- The **pamidronate** study compared two doses of drug without a placebo group. Patients treated with the higher dose had better BASDAI, BASFI, and BASMI responses. There was a statistically non-significant higher rate of arthralgias and myalgias after the first infusion.
- There was no benefit of **leflunomide** on any outcome measures in one study.
- The phosphodiesterase inhibitor **apremilast** demonstrated improvement in BASFI with trends toward benefit for other outcome measures but these were not statistically significant.
- The **thalidomide** study was an unblinded randomized trial that compared the effect of thalidomide with naproxen (and sulfasalazine in a third group) on maintenance of TNF inhibitor-induced treatment responses. Patients on thalidomide had a lower relapse rate. At the same time, significantly more patients in the thalidomide group withdrew due to adverse reactions or were lost to follow-up suggesting significant drug side effects.

Overall, there is a lack of evidence that treatment with SAARDs improves outcomes in AS. However, the small number of trials and of patients included in these studies represent important caveats. Furthermore, methotrexate was used at a dose that is considered subtherapeutic for the treatment of rheumatoid arthritis.

PICO continued on next page

Quality of Evidence Across All Critical Outcomes for **sulfasalazine**: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Sulfasalazine	Control	Relative (95% CI)	Absolute		
Health Status: Pain (axial) (follow-up median 31 weeks; measured with: VAS; Better indicated by lower values)												
6	randomized trials	serious[1]	no serious inconsistency (1%)	no serious indirectness	no serious imprecision	none	264	262	-	MD 1.84 lower (3.44 to 0.24 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Stiffness (follow-up median 36 weeks; measured with: duration (hours) or VAS; Better indicated by lower values)												
5	randomized trials	serious[1]	serious (70%)	no serious indirectness	no serious imprecision	none	241	238	-	MD 0.65 lower (1.73 lower to 0.42 higher)	⊕⊕OO LOW	IMPORTANT
Health Status: Physical Exam/Joint Counts (follow-up median 30 months; measured with: joint score; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	154	157	-	MD 0.9 lower (2.95 lower to 1.14 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up median 25 weeks; measured with: CRP or ESR; Better indicated by lower values)												
6	randomized trials	serious[1]	serious (65%)	no serious indirectness	no serious imprecision	none	257	259	-	MD 0.07 lower (0.36 lower to 0.23 higher)	⊕⊕OO LOW	NOT IMPORTANT
Functional Status: DFI (follow-up median 30 weeks; Better indicated by lower values)												
2	randomized trials	serious[2]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	154	157	-	MD 0.21 lower (1.21 lower to 0.8 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: ROM (follow-up median 36 weeks; measured with: Schober's test; Better indicated by higher values)												
5	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	249	249	-	MD 0.01 lower (0.2 lower to 0.18 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: Sleep disturbance (follow-up median 32 weeks)												
2	randomized trials	serious[2]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	11/35 (31.4%)	13/33 (39.4%)	OR 0.71 (0.26 to 1.93)	78 fewer per 1000 (from 249 fewer to 163 more)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Tender Joint Count (follow-up median 26 weeks; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	42	-	MD 0.4 lower (1.04 lower to 0.24 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Swollen Joint Count (follow-up median 26 weeks; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	42	-	MD 0 higher (0.28 lower to 0.28 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Overall responders % (follow-up median 36 weeks; assessed with: improvement in 2/4 domains)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	50/131 (38.2%)	48/133 (36.1%)	OR 1.09 (0.66 to 1.8)	20 more per 1000 (from 89 fewer to 143 more)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Physician global % responders (follow-up median 26 weeks; assessed with: 5-point rating scale)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70/131 (53.4%)	74/133 (55.6%)	OR 0.91 (0.56 to 1.49)	23 fewer per 1000 (from 144 fewer to 95 more)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Patient global % responders (follow-up median 26 weeks; assessed with: 5-point rating scale)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	53/131 (40.5%)	56/133 (42.1%)	OR 0.93 (0.57 to 1.53)	18 fewer per 1000 (from 128 fewer to 106 more)	⊕⊕⊕O MODERATE	IMPORTANT

Continued on next page

Health Status: Morning stiffness % responders (follow-up median 26 weeks; assessed with: VAS)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	64/131 (48.9%)	59/133 (44.4%)	OR 1.2 (0.74 to 1.94)	45 more per 1000 (from 73 fewer to 164 more)	⊕⊕⊕O MODERATE	IMPORTANT
								44.4%		45 more per 1000 (from 73 fewer to 164 more)		
Health Status: Back pain % responders (follow-up median 26 weeks)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	31/131 (23.7%)	36/133 (27.1%)	OR 0.84 (0.48 to 1.46)	33 fewer per 1000 (from 119 fewer to 81 more)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Joint pain (follow-up median 48 weeks; measured with: VAS; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	30	-	MD 0 higher (unable to calculate CI)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Joint swelling (follow-up median 48 weeks; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	131	133	-	MD 0.3 higher (1.05 lower to 1.65 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Dactylitis score (follow-up median 48 weeks; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	131	133	-	MD 0.1 higher (0.04 lower to 0.24 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Enthesitis score (follow-up median 48 weeks; Better indicated by lower values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	131	133	-	MD 0.3 higher (0.94 lower to 1.54 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Serious Adverse Event: all combined (study discontinuation) (follow-up median 36 weeks)												
7	randomized trials	serious[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	42/306 (13.7%)	30/309 (9.7%)	OR 1.52 (0.91 to 2.55)	43 more per 1000 (from 8 fewer to 118 more)	⊕⊕⊕O MODERATE	CRITICAL
Adverse Event: GI (follow-up median 36 weeks)												
7	randomized trials	serious[1]	serious (25%)	no serious indirectness	no serious imprecision	none	18/306 (5.9%)	16/309 (5.2%)	OR 1.52 (0.91 to 2.55)	25 more per 1000 (from 4 fewer to 70 more)	⊕⊕OO LOW	IMPORTANT

[1] Randomization and blinding poorly described in several studies.
[2] Randomization poorly described.

PICO continued on next page

Quality of Evidence Across All Critical Outcomes for **methotrexate**: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Methotrexate	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up median 24 weeks; Better indicated by lower values)												
2	randomized trials	no serious risk of bias	serious (35%)	serious[1]	no serious imprecision	none	29	34	-	MD 0.39 higher (0.69 lower to 1.47 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Pain (follow-up median 38 weeks; measured with: VAS; Better indicated by lower values)												
2	randomized trials	serious[2]	no serious inconsistency (0%)	serious[1]	no serious imprecision	none	43	43	-	MD 0.76 lower (2.02 lower to 0.49 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Stiffness (follow-up median 24 weeks; measured with: VAS; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	17	18	-	MD 6 higher (12.35 lower to 24.35 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up median 38 weeks; measured with: CRP; Better indicated by lower values)												
2	randomized trials	serious[2]	no serious inconsistency (0%)	serious[1]	no serious imprecision	none	38	41	-	MD 0.13 higher (0.27 lower to 0.54 higher)	⊕⊕OO LOW	NOT IMPORTANT
Functional Status: BASFI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	17	18	-	MD 0.3 higher (1.03 lower to 1.63 higher)	⊕⊕⊕O MODERATE	CRITICAL
fs-HAQ-S (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	17	18	-	MD 0 higher (0.3 lower to 0.3 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: DFI (follow-up median 52 weeks; Better indicated by lower values)												
1	randomized trials	Serious[3,4]	no serious inconsistency	serious[1]	no serious imprecision	none	26	25	-	MD 4.41 higher (0.27 lower to 9.09 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: BASMI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	serious[3]	no serious inconsistency	serious[1]	no serious imprecision	none	12	16	-	MD 0.25 higher (0.91 lower to 1.41 higher)	⊕⊕OO LOW	NOT IMPORTANT
Health Status: Composite score (follow-up median 24 weeks; assessed with: non-validated composite score, improvement of 20% or more in 5/7 domains)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	9/17 (52.9%)	3/18 (16.7%)	OR 5.62 (1.18 to 26.85)	363 more per 1000 (from 24 more to 676 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: Patient global (follow-up median 38 weeks; measured with: VAS or 5-point rating scale; Better indicated by lower values)												
2	randomized trials	serious[2]	no serious inconsistency (0%)	serious[1]	no serious imprecision	none	43	43	-	MD 0.31 higher (0.41 lower to 1.02 higher)	⊕⊕OO LOW	IMPORTANT
Health Status: Physician global (follow-up median 38 weeks; measured with: VAS or 5-point rating scale; Better indicated by lower values)												
2	randomized trials	serious[2]	serious (70%)	serious[1]	no serious imprecision	none	43	43	-	MD 4.95 lower (16.95 to 6.60 higher)	⊕OOO VERY LOW	IMPORTANT
Health Status: Enthesis index (follow-up median 52 weeks; Better indicated by lower values)												
1	randomized trials	serious[4]	no serious inconsistency	serious[1]	no serious imprecision	none	26	25	-	MD 1.27 lower (4.6 lower to 2.06 higher)	⊕⊕OO LOW	IMPORTANT
Continued on next page												

fs Spondylitis index (follow-up median 52 weeks; Better indicated by lower values)												
1	randomized trials	serious[4]	no serious inconsistency	serious[1]	no serious imprecision	none	26	25	-	MD 0.07 lower (1.51 lower to 1.37 higher)	⊕⊕OO LOW	IMPORTANT

[1] Lower dose than used in clinical practice.
[2] One of two studies not blinded.
[3] Randomization not explained.
[4] Study not blinded.

Quality of Evidence Across All Critical Outcomes for **pamidronate**: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Pamidronate	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up median 6 months; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	41	43	-	MD 1.27 lower (2.05 to 0.49 lower)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI (follow-up median 6 months; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	41	43	-	MD 1.52 lower (2.09 to 0.95 lower)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASMI (follow-up median 6 months; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	41	43	-	MD 0.48 lower (0.9 to 0.06 lower)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: arthralgia/myalgia (follow-up median 6 months)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	28/41 (68.3%)	20/43 (46.5%)	OR 2.48 (1.02 to 6.03)	218 more per 1000 (from 5 more to 375 more)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: BAS-G (follow-up median 6 months; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	41	43	-	MD 1.06 lower (1.86 to 0.26 lower)	⊕⊕⊕O MODERATE	IMPORTANT

[1] No placebo group.

PICO continued on next page

Quality of Evidence Across All Critical Outcomes for **leflunomide**: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Leflunomide	Control	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.8 lower (2 lower to 0.5 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain (follow-up median 24 weeks; measured with: VAS; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.9 lower (2.8 lower to 0.9 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Acute Phase Reactants (follow-up median 24 weeks; measured with: CRP; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 12.6 higher (5.8 lower to 30.9 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Functional Status: BASFI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.4 higher (0.5 lower to 1.3 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASMI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.3 lower (0.8 lower to 0.1 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: ASAS20 (follow-up median 24 weeks)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8/30 (26.7%)	3/15 (20%)	OR 1.45 (0.32 to 6.53)	66 more per 1000 (from 126 fewer to 420 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: BAS-G (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	-	-	MD 0.7 lower (2.4 lower to 0.9 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Swollen Joint Count (follow-up median 24 weeks; measured with: 44 joint count; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.4 higher (0.1 lower to 0.9 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Physician global (follow-up median 24 weeks; measured with: vas; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30	15	-	MD 0.2 higher (0.8 lower to 1.1 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Adverse Event: GI												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	17/30 (56.7%)	5/15 (33.3%)	OR 2.62 (0.72 to 9.54)	234 more per 1000 (from 69 fewer to 493 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: URI												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	5/30 (16.7%)	4/15 (26.7%)	OR 0.55 (0.12 to 2.45)	100 fewer per 1000 (from 225 fewer to 204 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT

Continued on next page

Adverse Event: dermatitis/prurigo											
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	4/30 (13.3%)	2/15 (13.3%)	OR 1 (0.16 to 6.19)	0 fewer per 1000 (from 109 fewer to 354 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: DVT											
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/30 (0%)	1/15 (6.7%)	OR 0.16 (0.01 to 4.13)	55 fewer per 1000 (from 66 fewer to 161 more)	⊕⊕⊕O MODERATE	IMPORTANT
Adverse Event: LFT elevation											
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1/30 (3.3%)	0/15 (0%)	OR 1.58 (0.06 to 41.03)	-	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: HTN											
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1/30 (3.3%)	0/15 (0%)	OR 1.58 (0.06 to 41.03)	-	⊕⊕⊕O MODERATE	NOT IMPORTANT

[1] Randomization not explained.

Quality of Evidence Across All Critical Outcomes for **apremilast**: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency (I^2)**	**Indirectness**	**Imprecision**	**Other considerations**	**Apremilast**	**Control**	**Relative (95% CI)**	**Absolute**		
Health Status: BASDAI (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	17	19	-	MD 0.82 lower (1.79 lower to 0.15 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: ASDAS (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	17	19	-	MD 0.31 higher (0.14 lower to 0.76 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants (follow-up median 12 weeks; measured with: CRP; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	17	19	-	MD 3.61 lower (18.33 lower to 11.11 higher)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Functional Status: BASFI (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	17	19	-	MD 1.46 lower (2.62 to 0.3 lower)	⊕⊕⊕⊕ HIGH	CRITICAL
Functional Status: BASMI (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	17	19	-	MD 0.3 lower (0.87 lower to 0.27 higher)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Health Status: BAS-G (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	17	19	-	MD 1.19 lower (2.88 lower to 0.5 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: FACIT-F (follow-up median 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	17	19	-	MD 4.31 higher (4.26 lower to 12.88 higher)	⊕⊕⊕⊕ HIGH	CRITICAL
Continued on next page												

Health Status: ASAS20 (follow-up median 12 weeks)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	6/17 (35.3%)	3/19 (15.8%)	OR 2.91 (0.6 to 14.18)	195 more per 1000 (from 57 fewer to 569 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: ASAS40 (follow-up median 12 weeks)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	4/17 (23.5%)	1/19 (5.3%)	OR 5.54 (0.55 to 55.49)	183 more per 1000 (from 23 fewer to 702 more)	⊕⊕OO LOW	NOT IMPORTANT
Health Status: ASAS5/6 (follow-up median 12 weeks)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	3/17 (17.6%)	1/19 (5.3%)	OR 3.86 (0.36 to 41.2)	124 more per 1000 (from 33 fewer to 643 more)	⊕⊕OO LOW	NOT IMPORTANT
Health Status: Night pain (Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	17	19	-	MD 0.58 lower (2.47 lower to 1.31 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Adverse Event: headache												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	8/19 (42.1%)	5/19 (26.3%)	OR 2.04 (0.52 to 8)	158 more per 1000 (from 107 fewer to 478 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: loose stools												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	5/19 (26.3%)	2/19 (10.5%)	OR 3.04 (0.51 to 18.11)	158 more per 1000 (from 49 fewer to 575 more)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Adverse Event: elevated serum amylase												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	2/19 (10.5%)	0/19 (0%)	OR 5.57 (0.25 to 124.19)	-	⊕⊕OO LOW	NOT IMPORTANT

[1] Wide CI.

Quality of Evidence Across All Critical Outcomes for **thalidomide**: Very Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Thalidomide	Control	Relative (95% CI)	Absolute		
Recurrence rate (follow-up median 1 years)												
1	randomized trials	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	none	25/37 (67.6%)	33/37 (89.2%)	OR 0.25 (0.07 to 0.88)	218 fewer per 1000 (from 13 fewer to 526 fewer)	⊕OOO VERY LOW	IMPORTANT
Adverse Event: Discontinuation or lost to follow-up (follow-up median 1 years)												
1	randomized trials	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	none	7/37 (18.9%)	0/37 (0%)	OR 18.44 (1.01 to 335.96)	-	⊕OOO VERY LOW	IMPORTANT

[1] Study not blinded.
[2] Randomization not explained.
[3] Maintenance of clinical benefit after prior TNF inhibitor therapy.

PICO 7 includes RCT:	Clegg 1996[2]; Corkill 1990[4]; Davis 1989[5]; Dougados 1986[6]; Feltelius 1986[7]; Kirwan 1993[1]; Nissila 1988[8]; Taylor 1991[9]; Altan 2001[10]; Roychowdhury 2002[11]; Gonzalez-Lopez 2004[12]; Maksymowych 2002[13]; van Denderen 2005[14]; Pathan 2013[15]; Deng 2013[16]
PICO 7 includes Observational studies :	None

1. Kirwan J, Edwards A, Huitfeldt B, Thompson P, Currey H. The course of established ankylosing spondylitis and the effects of sulphasalazine over 3 years. Br J Rheumatol 1993;32:729-33.

2. Clegg DO, Reda DJ, Weisman MH, Blackburn WD, Cush JJ, Cannon GW, et al. Comparison of sulfasalazine and placebo in the treatment of ankylosing spondylitis. A Department of Veterans Affairs Cooperative Study. Arthritis Rheum 1996;39:2004-12.

3. Haibel H, Brandt HC, Song IH, Brandt A, Listing J, Rudwaleit M, et al. No efficacy of subcutaneous methotrexate in active ankylosing spondylitis: a 16-week open-label trial. Ann Rheum Dis 2007;66:419-21.

4. Corkill MM, Jobanputra P, Gibson T, Macfarlane DG. A controlled trial of sulphasalazine treatment of chronic ankylosing spondylitis: failure to demonstrate a clinical effect. Br J Rheumatol 1990;29:41-5.

5. Davis MJ, Dawes PT, Beswick E, Lewin IV, Stanworth DR. Sulphasalazine therapy in ankylosing spondylitis: its effect on disease activity, immunoglobulin A and the complex immunoglobulin A-alpha-1-antitrypsin. Br J Rheumatol 1989;28:410-3.

6. Dougados M, Boumier P, Amor B. Sulphasalazine in ankylosing spondylitis: a double blind controlled study in 60 patients. Br Med J (Clin Res Ed) 1986;293:911-4.

7. Feltelius N, Hallgren R. Sulphasalazine in ankylosing spondylitis. Ann Rheum Dis 1986;45:396-9.

8. Nissila M, Lehtinen K, Leirisalo-Repo M, Luukkainen R, Mutru O, Yli-Kerttula U. Sulfasalazine in the treatment of ankylosing spondylitis. A twenty-six-week, placebo-controlled clinical trial. Arthritis Rheum 1988;31:1111-6.

9. Taylor HG, Beswick EJ, Dawes PT. Sulphasalazine in ankylosing spondylitis. A radiological, clinical and laboratory assessment. Clin Rheumatol 1991;10:43-8.

10. Altan L, Bingol U, Karakoc Y, Aydiner S, Yurtkuran M, Yurtkuran M. Clinical investigation of methotrexate in the treatment of ankylosing spondylitis. Scand J Rheumatol 2001;30:255-9.

11. Roychowdhury B, Bintley-Bagot S, Bulgen DY, Thompson RN, Tunn EJ, Moots RJ. Is methotrexate effective in ankylosing spondylitis? Rheumatology (Oxford) 2002;41:1330-2.

12. Gonzalez-Lopez L, Garcia-Gonzalez A, Vazquez-Del-Mercado M, Munoz-Valle JF, Gamez-Nava JI. Efficacy of methotrexate in ankylosing spondylitis: a randomized, double blind, placebo controlled trial. J Rheumatol 2004;31:1568-74.

13. Maksymowych WP, Jhangri GS, Fitzgerald AA, LeClercq S, Chiu P, Yan A, et al. A six-month randomized, controlled, double-blind, dose-response comparison of intravenous pamidronate (60 mg versus 10 mg) in the treatment of nonsteroidal antiinflammatory drug-refractory ankylosing spondylitis. Arthritis Rheum 2002;46:766-73.

14. van Denderen JC, van der Paardt M, Nurmohamed MT, de Ryck YM, Dijkmans BA, van der Horst-Bruinsma IE. Double blind, randomised, placebo controlled study of leflunomide in the treatment of active ankylosing spondylitis. Ann Rheum Dis 2005;64:1761-4.

15. Pathan E, Abraham S, Van Rossen E, Withrington R, Keat A, Charles PJ, et al. Efficacy and safety of apremilast, an oral phosphodiesterase 4 inhibitor, in ankylosing spondylitis. Ann Rheum Dis 2013;72:1475-80.

16. Deng X, Zhang J, Zhang J, Huang F. Thalidomide reduces recurrence of ankylosing spondylitis in patients following discontinuation of etanercept. Rheumatol Int 2013;33:1409-13.

PICO 39. In adults with active non-radiographic axial SpA despite treatment with NSAIDs, are SAARDs more effective than no treatment with SAARDs in improving outcomes?

Summary: This PICO was addressed by a single RCT that compared the effectiveness of sulfasalazine (SSZ) with placebo in patients with inflammatory back pain, spondyloarthritis according to ESSG criteria, and disease duration <5 years (Braun, 2006) [1]. While this study did not use the ASAS 2009 classification criteria for axial spondyloarthritis, only 13% of subjects had radiographic sacroiliitis. The study population therefore largely reflects patients with non-radiographic axial SpA. There was no evidence that SSZ improved critical outcomes compared to placebo.

Quality of Evidence Across All Critical Outcomes: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Sulfasalazine	placebo	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	112	118	-	MD 0.24 lower (0.82 lower to 0.33 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Pain (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	112	118	-	MD 0.01 higher (0.62 lower to 0.64 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	112	118	-	MD 0.08 higher (0.66 lower to 0.83 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Acute Phase Reactants CRP (mg/dL) (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	35	25	-	MD 4.79 higher (3.3 lower to 12.69 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: Function (WOMAC, physical function) (follow-up median 234 weeks; measured with: WOMAC index physical function; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	112	118	-	MD 0.18 lower (0.68 lower to 0.31 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI (follow-up median 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	112	118	-	MD 0.18 lower (0.67 lower to 0.31 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: ROM – Schober's (follow-up median 24 weeks; measured with: Schober's test; Better indicated by higher values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	112	118	-	MD 0.06 lower (0.4 lower to 0.28 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Serious Adverse Event: all combined (follow-up median 24 weeks)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	6/112 (5.4%)	10/118 (8.5%)	RR 0.63 (0.23 to 1.64)	31 fewer per 1000 (from 65 fewer to 54 more)	⊕⊕⊕O MODERATE	CRITICAL

[1] included 13% with sacroiliitis on plain films
[2] CI calculated in paper (represented) not consistent with results through RevMan

PICO 39 includes RCT:	Braun 2006 [1]
PICO 39 includes Observational studies :	None

1. Braun J, Zochling J, Baraliakos X, Alten R, Burmester G, Grasedyck K, et al. Efficacy of sulfasalazine in patients with inflammatory back pain due to undifferentiated spondyloarthritis and early ankylosing spondylitis: a multicentre randomised controlled trial. Ann Rheum Dis 2006;65:1147-53.

Treatment of Stable Disease

PICO 11. In adults with stable AS on treatment with TNFi and NSAIDs, is continuing both medications more effective in improving outcomes than continuing treatment with TNFi alone?

Summary: This PICO was not directly addressed by any studies. The INFAST Part 2 (Sieper 2014)[1] RCT examined patients who had achieved remission on infliximab and compared treatment with an NSAID to no treatment (all patients had discontinued infliximab). The study was comprised of 60% AS patients and 40% nr-axSpA patients, but results were not reported separately and therefore are not included.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

1. Sieper J, Lenaerts J, Wollenhaupt J, Rudwaleit M, Mazurov VI, Myasoutova L, et al. Maintenance of biologic-free remission with naproxen or no treatment in patients with early, active axial spondyloarthritis: results from a 6-month, randomised, open-label follow-up study, INFAST Part 2. Ann Rheum Dis 2014;73:108-13.

PICO 43. In adults with stable non-radiographic axial SpA on treatment with TNFi and NSAIDs, is continuation of both medications more effective in improving outcomes than continuing treatment with TNFi alone?

Summary: This PICO was not directly addressed by any studies.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 12. In adults with stable AS on treatment with TNFi and SAARD, is continuing both medications more effective in improving outcomes than withdrawing one treatment and continuing either TNFi or SAARD alone?

Summary: This PICO was not directly addressed by any RCTs; one observational study only indirectly addressed the question. An open-label RCT extension study suggested that >90% of stable patients previously on infliximab monotherapy flared by 48 weeks (Baraliakos 2005)[1].

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 12 includes RCT:	None
PICO 12 includes Observational studies :	Baraliakos 2005[1]

1. Baraliakos X, Listing J, Brandt J, Zink A, Alten R, Burmester G *et al.* Clinical response to discontinuation of anti-TNF therapy in patients with ankylosing spondylitis after 3 years of continuous treatment with infliximab. Arthritis Res Ther 2005;7:R439-R444.

PICO 44. In adults with stable non-radiographic axial SpA on treatment with TNFi and SAARD, is continuation of both medications more effective in improving outcomes than withdrawing one treatment and continuing either TNFi or SAARD alone?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

REHABILITATION/PHYSICAL THERAPY

PICO 16. In adults with active AS, is any form of PT more effective than no PT in improving health status and functional status?

Summary: This PICO was directly addressed by 2 small RCTs. There were some significant differences between groups, with 3 of the 6 endpoints favoring the intervention, with wide confidence intervals, and high risk of bias.

Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Any PT	Control: No PT	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up mean 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	46	49	-	MD 14.3 lower (22.64 to 5.96 lower)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI (follow-up mean 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	46	49	-	MD 6 lower (12.82 lower to 0.82 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: SF-36 physical - role physical (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	46	49	-	MD 11.8 higher (2.02 to 21.58 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: SF-36 physical -physical function (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	46	49	-	MD 2.2 higher (4.48 lower to 8.88 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: ROM: hip External Rotation mean difference (follow-up mean 3 weeks; Better indicated by higher values)												
1	randomized trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	12	-	MD 5.6 higher (1.86 to 9.34 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: BAS-G (follow-up mean 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	46	49	-	MD 6.4 lower (14.8 lower to 2 higher)	⊕⊕⊕O MODERATE	IMPORTANT

[1] Small samples
[2] Unclear allocation concealment, small sample of convenience

PICO 16 includes RCT:	Kjeken [1]; Bulstrode[2]
PICO 16 includes Observational studies :	None

1. Kjeken I, Bo I, Ronningen A, Spada C, Mowinckel P, Hagen KB, et al. A three-week multidisciplinary in-patient rehabilitation programme had positive long-term effects in patients with ankylosing spondylitis: randomized controlled trial. J Rehabil Med 2013;45:260-7.

2. Bulstrode SJ, Barefoot J, Harrison RA, Clarke AK. The role of passive stretching in the treatment of ankylosing spondylitis. Br J Rheumatol 1987;26:40-2.

PICO 17. In adults with active AS, are active PT interventions (supervised exercise) more effective than passive PT interventions (massage, ultrasound, heat) in improving health status and functional status?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 18. In adults with active AS, are aquatic PT interventions more effective than land-based PT interventions in improving health status and functional status?

Summary: This PICO was directly addressed by 5 RCTs. There were some significant differences between groups, with 7 of the 16 examined endpoints favoring the intervention, with narrow/wide confidence intervals, and serious risk of bias complicating 5 of the 16 reported outcomes.
Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Aquatic	Control: land	Relative (95% CI)	Absolute		
Health Status: BASDAI (follow-up mean 4 weeks; Better indicated by lower values)												
5	randomized trials	no serious risk of bias	no serious inconsistency (12%)	serious[1]	no serious imprecision	none	185	140	-	MD 0.37 lower (0.69 to 0.04 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain - (follow-up mean 3.5 weeks; Better indicated by lower values)												
2	randomized trials	no serious risk of bias	serious (64%)	no serious indirectness[1]	no serious imprecision	none	108	66	-	MD 0.51 lower (1.52 lower to 0.49 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	28	26	-	MD 0.2 higher (0.12 lower to 0.52 higher)	⊕⊕⊕⊕ HIGH	IMPORTANT
Functional Status: ROM - modified Schober's test (follow-up mean 4.5 weeks; Better indicated by higher values)												
3	randomized trials	serious[2]	serious (40%)	no serious indirectness	no serious imprecision	none	76	72	-	MD 0.19 lower (0.75 lower to 0.38 higher)	⊕⊕OO LOW	NOT IMPORTANT
Continued on next page												

Health Status: Depression BDI (follow-up mean 6 weeks; Better indicated by lower values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13	12	-	MD 0.74 higher (5.6 lower to 7.08 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: ASQOL - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	29	28	-	MD 2.07 lower (3 to 1.14 lower)	⊕⊕⊕⊕ HIGH	CRITICAL
Health Status: Patient global disease activity - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	28	26	-	MD 0.54 lower (1 to 0.08 lower)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: Patient global well-being (follow-up mean 4 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	very serious[1]	no serious imprecision	none	80	40	-	MD 0.93 lower (1.78 to 0.08 lower)	⊕⊕OO LOW	IMPORTANT
Health Status: NHP - Health Status: NHP total - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	28	26	-	MD 51.07 lower (81.45 to 20.69 lower)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: VO2 (follow-up mean 6 weeks; Better indicated by higher values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13	12	-	MD 3.74 higher (1.32 lower to 8.8 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Pulmonary function FEV1 (follow-up mean 6 weeks; Better indicated by lower values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13	12	-	MD 0.57 lower (1.43 lower to 0.29 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: BASFI (follow-up mean 4 weeks; Better indicated by lower values)												
5	randomized trials	no serious risk of bias	no serious inconsistency (0%)	serious[1]	no serious imprecision	none	185	140	-	MD 0.22 lower (0.51 lower to 0.07 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: HAQ-S - mean difference (follow-up mean 4 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	80	80	-	MD 0.24 lower (0.33 to 0.15 lower)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: DFI - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	28	26	-	MD 2.6 lower (5.1 to 0.1 lower)	⊕⊕⊕⊕ HIGH	CRITICAL
Functional Status: BASMI (follow-up mean 4.5 weeks; Better indicated by lower values)												
3	randomized trials	serious[2]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	77	74	-	MD 0.08 higher (0.75 lower to 0.92 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Health Status: 6 minute walk test (follow-up mean 6 weeks; Better indicated by higher values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13	12	-	MD 87.17 higher (55.79 to 118.55 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT

[1] 1 study was a 3 arm study comparing 2 types of spa therapy vs. control group
[2] For 1 of 2 studies, assessor was not blinded to group assignment
[3] Assessor not blinded to group assignment

PICO 18 includes RCT:	Karapolat[1]; Gurcay[2]; Altan[3]; Van Tubergen[4]; Dundar[5]
PICO 18 includes Observational studies :	None

1. Karapolat H, Eyigor S, Zoghi M, Akkoc Y, Kirazli Y, Keser G. Are swimming or aerobic exercise better than conventional exercise in ankylosing spondylitis patients? A randomized controlled study. Eur J Phys Rehabil Med 2009;45:449-57.

2. Gurcay E, Yuzer S, Eksioglu E, Bal A, Cakci A. Stanger bath therapy for ankylosing spondylitis: illusion or reality? Clin Rheumatol 2008;27:913-7.

3. Altan L, Bingol U, Aslan M, Yurtkuran M. The effect of balneotherapy on patients with ankylosing spondylitis. Scand J Rheumatol 2006;35:283-9.

4. van Tubergen A, Landewe R, van der Heijde D, Hidding A, Wolter N, Asscher M, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.

5. Dundar U, Solak O, Toktas H, Demirdal US, Subasi V, Kavuncu V, et al. Effect of aquatic exercise on ankylosing spondylitis: a randomized controlled trial. Rheumatol Int 2014.

PICO 19. In adults with stable AS, is any form of PT more effective than no PT in improving health status and functional status?

Summary: This PICO was directly addressed by 10 RCTs reported in 11 articles. There were some significant differences between groups, with 9 of the 14 examined endpoints favoring the intervention, with narrow confidence intervals, and substantial variation in the quality of studies.

Quality of Evidence Across All Critical Outcomes: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	any PT	Control: No PT	Relative (95% CI)	Absolute		
Health Status: BASDAI - mean difference (follow-up mean 13.5 weeks; Better indicated by lower values)												
4	randomized trials	serious[1]	serious (76%)	no serious risk of bias	no serious imprecision	none	450	433	-	MD 0.67 lower (1.2 to 0.14 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Pain - mean difference (follow-up mean 13.75 weeks; Better indicated by lower values)												
4	randomized trials	serious[1]	serious (86%)	no serious indirectness	no serious imprecision	none	432	424	-	MD 1.26 lower (2.8 lower to 0.28 higher)	⊕⊕OO LOW	CRITICAL
Health Status: Fatigue/MAF (follow-up mean 9 weeks; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	46	39	-	MD 0.4 lower (0.9 lower to 0.11 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness (follow-up mean 4.5 weeks; measured with: VAS; Better indicated by lower values)												
2	randomized trials	serious[3]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	31	31	-	MD 1.72 lower (3.51 lower to 0.08 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: ROM – Schober's test - at follow up (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	15	15	-	MD 1.35 higher (0.14 to 2.56 higher)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Health Status: Depression - BDI (follow-up mean 12 weeks; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 3.89 lower (5.31 to 2.47 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: ASQOL - mean difference (follow-up mean 24 weeks; Better indicated by lower values)												
2	randomized trials	no serious risk of bias	serious (64%)	no serious indirectness	no serious imprecision	none	410	399	-	MD 0.31 lower (1.48 lower to 0.86 higher)	⊕⊕⊕O MODERATE	CRITICAL
Functional Status: BASFI - mean difference (follow-up 13.5 weeks; Better indicated by lower values)												
4	randomized trials	serious[1]	serious (68%)	no serious indirectness	no serious imprecision	none	503	486	-	MD 0.47 lower (0.90 to 0.04 lower)	⊕⊕OO LOW	CRITICAL
Functional Status: SF-36 physical function (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 0.18 higher (0.07 to 0.29 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: BASMI - mean difference (follow-up mean 7 weeks; Better indicated by lower values)												
3	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	103	98	-	MD 0.26 lower (0.36 to 0.15 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT
Continued on next page												

Health Status: Patient global disease activity - mean difference (follow-up mean 24 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	381	375	-	MD 0.39 lower (0.71 to 0.07 lower)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: Physical work capacity - Health Status: physical work capacity 170 (follow-up mean 12 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	15	15	-	MD 0.69 higher (0.26 to 1.12 higher)	⊕⊕⊕⊕ HIGH	IMPORTANT
Health Status: Pulmonary function tests - Health Status: VC - mean difference (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	19	13	-	MD 9.64 higher (5.53 to 13.75 higher)	⊕⊕OO LOW	IMPORTANT
Functional Status: 6 min walk test (follow-up mean 12 weeks; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	19	13	-	MD 62.81 higher (45.6 to 80.02 higher)	⊕⊕OO LOW	NOT IMPORTANT

[1] 1 study with unclear randomization of assignment and blinding of assessor
[2] 1 study used 3 arm design comparing 2 interventions to a control group
[3] Small sample size, short follow up

PICO 19 includes RCT :	Altan[1]; Durmus[2]; Durmus[3]; Rodriguez-Lozano[4]; Kraag[5]; Gemignani[6]; Ince 2006[7];Niedermann[8]; Widberg 2009[9]; Masiero [10]; Masiero[11]
PICO 19 includes Observational studies :	none

1. Altan L, Korkmaz N, Dizdar M, Yurtkuran M. Effect of Pilates training on people with ankylosing spondylitis. Rheumatol Int 2012;32:2093-9.

2. Durmus D, Alayli G, Cil E, Canturk F. Effects of a home-based exercise program on quality of life, fatigue, and depression in patients with ankylosing spondylitis. Rheumatol Int 2009;29:673-7.

3. Durmus D, Alayli G, Uzun O, Tander B, Canturk F, Bek Y, et al. Effects of two exercise interventions on pulmonary functions in the patients with ankylosing spondylitis. Joint Bone Spine 2009;76:150-5.

4. Rodriguez-Lozano C, Juanola X, Cruz-Martinez J, Pena-Arrebola A, Mulero J, Gratacos J, et al. Outcome of an education and home-based exercise programme for patients with ankylosing spondylitis: a nationwide randomized study. Clin Exp Rheumatol 2013;31:739-48.

5. Kraag G, Stokes B, Groh J, Helewa A, Goldsmith C. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis--a randomized controlled trial. J Rheumatol 1990;17:228-33.

6. Gemignani G, Olivieri I, Ruju G, Pasero G. Transcutaneous electrical nerve stimulation in ankylosing spondylitis: a double-blind study. Arthritis Rheum 1991;34:788-9.

7. Ince G, Sarpel T, Durgun B, Erdogan S. Effects of a multimodal exercise program for people with ankylosing spondylitis. Phys Ther 2006;86:924-35.

8. Niedermann K, Sidelnikov E, Muggli C, Dagfinrud H, Hermann M, Tamborrini G, et al. Effect of cardiovascular training on fitness and perceived disease activity in people with ankylosing spondylitis. Arthritis Care Res (Hoboken) 2013;65:1844-52.

9. Widberg K, Karimi H, Hafstrom I. Self- and manual mobilization improves spine mobility in men with ankylosing spondylitis--a randomized study. Clin Rehabil 2009;23:599-608.

10. Masiero S, Bonaldo L, Pigatto M, Lo NA, Ramonda R, Punzi L. Rehabilitation treatment in patients with ankylosing spondylitis stabilized with tumor necrosis factor inhibitor therapy: a randomized controlled trial. J Rheumatol 2011;38:1335-42.

11. Masiero S, Poli P, Bonaldo L, Pigatto M, Ramonda R, Lubrano E, et al. Supervised training and home-based rehabilitation in patients with stabilized ankylosing spondylitis on TNF inhibitor treatment: a controlled clinical trial with a 12-month follow-up. Clin Rehabil 2013;28:562-72.

PICO 20. In adults with active or stable AS, are unsupervised back exercises more effective than no exercise in improving health status and functional status?

Summary: This PICO was directly addressed by 2 RCTs. Only 1 outcome out of 5 favored the intervention (mailed educational materials), and it was a measure of self-efficacy, rather than a true clinical outcome. Confidence intervals were narrow and risk of bias was assessed as serious.

Quality of Evidence Across All Critical Outcomes: Moderate ⊕⊕⊕O

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	unsupervised ex	Control: No ex	Relative (95% CI)	Absolute		
Health Status: BASDAI - mean difference (follow-up mean 4.5 weeks; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	104	111	-	MD 0.33 higher (0.09 lower to 0.74 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Self Efficacy Scale Pain - mean difference (follow-up mean 6 months; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	75	80	-	MD 0.1 higher (0.38 lower to 0.58 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT
Functional Status: BASFI - mean difference (follow-up mean 13.5 weeks; Better indicated by lower values)												
2	randomized trials	serious[1]	no serious inconsistency (NC)	no serious indirectness	no serious imprecision	none	104	111	-	MD 0.58 higher (1.17 lower to 2.33 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: BAS-G - mean difference (follow-up mean 6 months; Better indicated by lower values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	75	80	-	MD 0.14 higher (0.72 lower to 1 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Exercise Self Efficacy - mean difference (follow-up mean 6 months; Better indicated by higher values)												
1	randomized trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	75	80	-	MD 1.96 higher (0.57 to 3.35 higher)	⊕⊕⊕O MODERATE	NOT IMPORTANT

[1] Randomization and allocation concealment not addressed; blinding not discussed

PICO 20 includes RCT :	Sweeney[1]; Ayhan[2]
PICO 20 includes Observational studies :	None

1. Sweeney S, Taylor G, Calin A. The effect of a home based exercise intervention package on outcome in ankylosing spondylitis: a randomized controlled trial. J Rheumatol 2002;29:763-6.

2. Ayhan F, Gecene M, Gunduz R, Borman P, Yorgancioglu R. Long-term effects of comprehensive inpatient rehabilitation on function and disease activity in patients with chronic rheumatoid arthritis and ankylosing spondylitis. Turkish Journal of Rheumatology 2011;26:135-44.

PICO 21. In adults with active or stable AS and spinal fusion or advance spinal osteoporosis, is spinal manipulation (chiropractic or osteopathic) more effective than no spinal manipulation in improving health status and functional status?

Summary: This PICO was not directly addressed by any RCTs. A number of systematic reviews have identified adverse events associated with spinal manipulation and a few case series have reported untoward events resulting from spinal manipulation in AS patients.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 21 includes systematic reviews :	Ernst 2007[1]; Hebert 2013[2]; Carnes 2010[3]
PICO 21 includes Observational studies :	Rinsky 1976[4]; Liao 2007[5]

1. Ernst E. Adverse effects of spinal manipulation: a systematic review. J R Soc Med 2007;100:330-8.
2. Hebert JJ, Stomski NJ, French SD, Rubinstein SM. Serious adverse events and spinal manipulative therapy of the low back region: a systematic review of cases. J Manipulative Physiol Ther 2013:Jun 17.
3. Carnes D, Mars TS, Mullinger B, Froud R, Underwood M. Adverse events and manual therapy: a systematic review. Man Ther 2010;15:355-63.
4. Rinsky LA, Reynolds GG, Jameson RM, Hamilton RD. A cervical spinal cord injury following chiropractic manipulation. Paraplegia 1976;13:223-7.
5. Liao CC, Chen LR. Anterior and posterior fixation of a cervical fracture induced by chiropractic spinal manipulation in ankylosing spondylitis: a case report. J Trauma 2007;63:E90-E94.

PICO 22. In adults with active non-radiographic axial SpA, is any form of PT more effective than no PT in improving health status and functional status?

Summary: This PICO was directly addressed by 2 small RCTs. There some significant differences between groups with 5 of 7 outcomes favoring the intervention, wide confidence intervals, but no serious risk of bias.

Quality of Evidence Across All Critical Outcomes: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Any PT	Control: no PT	Relative (95% CI)	Absolute		
Health Status: BASDAI - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 18 lower (26.65 to 9.35 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Pain (follow-up mean 2 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	12	12	-	MD 22.7 lower (38.48 to 6.92 lower)	⊕⊕OO LOW	CRITICAL
Health Status: Stiffness - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 17 lower (30.38 to 3.62 lower)	⊕⊕OO LOW	IMPORTANT
Functional Status: BASFI - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 1 lower (1.76 to 0.24 lower)	⊕⊕OO LOW	CRITICAL
Functional Status: HAQ-S - mean difference (follow-up 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 0.11 lower (0.46 lower to 0.24 higher)	⊕⊕OO LOW	CRITICAL
Functional Status: DFI - mean difference (follow-up mean 3 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 1.1 lower (3.47 lower to 1.27 higher)	⊕⊕OO LOW	CRITICAL
Health Status: BAS-G mean difference (follow-up mean 2 weeks; Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	25	18	-	MD 17 lower (28.42 to 5.58 lower)	⊕⊕OO LOW	IMPORTANT

[1] Sample size limited

PICO 22 includes RCT :	Cozzi[1]; Viitanen[2]
PICO 22 includes Observational studies :	None

1. Cozzi F, Podswiadek M, Cardinale G, Oliviero F, Dani L, Sfriso P *et al.* Mud-bath treatment in spondylitis associated with inflammatory bowel disease--a pilot randomised clinical trial. Joint Bone Spine 2007;74:436-9.

2. Viitanen JV, Heikkila S. Functional changes in patients with spondylarthropathy. A controlled trial of the effects of short-term rehabilitation and 3-year follow-up. Rheumatol Int 2001;20:211-4.

PICO 23. In adults with active non-radiographic axial SpA, are active PT interventions (supervised exercise) more effective than passive PT interventions (massage, ultrasound, heat) in improving health status and functional status?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 24. In adults with active non-radiographic axial SpA, are aquatic PT interventions more effective than land-based PT interventions in improving health status and functional status?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

SURGICAL TREATMENTS

PICO 25. In adults with AS and advanced hip arthritis, is total hip arthroplasty (THA) more effective than no surgery in improving outcomes?

Summary: Studies prior to 1996 were not included due to changes in surgical technique. Even so, only 4 of the 8 studies employed modern techniques and implants; the remaining 4 described obsolete implants and techniques. This PICO was not directly addressed by any RCTs. It was indirectly addressed by 1 observational study (Li 2009),[1] and 7 case-series. The observational study addressed total ROM in metal on metal resurfacing (n=38; currently done less commonly in the US) compared with THA; no true placebo or non-surgical control groups were included. The THA group (n=25 patients, 41 hips) followed for a mean of 2.9 years demonstrated a mean Harris Hip Score (HHS) improvement of 39.4, pain score improvement of 3.12, and 113 degree total ROM improvement compared with baseline. Seven case series (n=275 patients, 474 hips) followed for a median of 7.4 years demonstrated a median Harris Hip Score (HHS) improvement of 55 points (5 studies). ROM improvements were substantial across studies, but reported differently (preventing aggregation of results). Only 2 studies reported verifying the diagnosis of AS according to current criteria. Results were described as 65%-85% "good/excellent" in 2 studies.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	THA	Control	Relative (95% CI)	Absolute		
Mortality												
Health Status: Pain (follow-up 20-46 months; 0-10, Better indicated by lower values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	41	-	-	Mean 3.12 lower	⊕OOO VERY LOW	CRITICAL
Functional Status: ROM (total degrees) (follow-up 20-46 months; Better indicated by higher values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	41	-	-	mean 113 higher (26 to 0 higher)	⊕OOO VERY LOW	NOT IMPORTANT
Heterotopic ossification (follow-up 20-46 months)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	6/41 (14.6%)	-	-	-	⊕OOO VERY LOW	NOT IMPORTANT
Harris Hip Score (follow-up 20-46 months; range of scores: 0-100; Better indicated by higher values)												
1	observational studies	very serious[1]	no serious inconsistency	very serious[2]	no serious imprecision	none	41	-	-	MD 39.4 higher (34.9 to 43.9 higher)	⊕OOO VERY LOW	NOT IMPORTANT

[1] Sample size limited; no controls
[2] Reported on resurfacing, rather than THA

PICO 25 includes RCT :	None
PICO 25 includes Observational studies/case series :	Bangjian 2012[2]; Bhan 2008[3]; Joshi 2002[4]; Tang 2000[5]; Sochart 1997[6]; Brinker 1996[7]; Bhan 1996[8]; Li 2009[1]

1. Li J, Xu W, Xu L, Liang Z. Hip resurfacing arthroplasty for ankylosing spondylitis. J Arthroplasty 2009;24:1285-91.

2. Bangjian H, Peijian T, Ju L. Bilateral synchronous total hip arthroplasty for ankylosed hips. Int Orthop 2012;36:697-701.

3. Bhan S, Eachempati KK, Malhotra R. Primary cementless total hip arthroplasty for bony ankylosis in patients with ankylosing spondylitis. J Arthroplasty 2008;23:859-66.

4. Joshi AB, Markovic L, Hardinge K, Murphy JC. Total hip arthroplasty in ankylosing spondylitis: an analysis of 181 hips. J Arthroplasty 2002;17:427-33.

5. Tang WM, Chiu KY. Primary total hip arthroplasty in patients with ankylosing spondylitis. J Arthroplasty 2000;15:52-8.

6. Sochart DH, Porter ML. Long-term results of total hip replacement in young patients who had ankylosing spondylitis. Eighteen to thirty-year results with survivorship analysis. J Bone Joint Surg Am 1997;79:1181-9.

7. Brinker MR, Rosenberg AG, Kull L, Cox DD. Primary noncemented total hip arthroplasty in patients with ankylosing spondylitis. Clinical and radiographic results at an average follow-up period of 6 years. J Arthroplasty 1996;11:802-12.

8. Bhan S, Malhotra R. Bipolar hip arthroplasty in ankylosing spondylitis. Arch Orthop Trauma Surg 1996;115:94-9.

PICO 26. In adults with AS and severe kyphosis, is elective spinal osteotomy more effective than no surgery in improving outcomes?

Summary: This PICO centered on two clinical scenarios: 1) correction cervical or cervicothoracic kyphosis [CK]; and 2) correction of thoracolumbar deformity [TLD]. Correction of spinal pseudoarthrosis was not considered elective and therefore excluded from this PICO. This PICO was not directly addressed by any RCTs.

A prior systematic review (Etame 2008)[1] reported on 227 patients in 6 case series of **CK**, which represent the majority of the relevant studies to date. It found inconsistent/non-standardized preoperative and postoperative evaluations, although results reported "high likelihood of success" and restoration of horizontal gaze in all patients, complications were common (26.9% to 87.5%), and the peri-operative mortality rate was 2.6%. A single more recent case series published in 2013 demonstrated similar findings (Koller 2013)[2]. Studies prior to 1999 addressing **TLD** were evaluated and summarized in a prior systematic review (van Royen 1999)[3] that reported on 856 patients in 41 case-series managed by three different surgical techniques. It found inconsistent/non-standardized preoperative and postoperative evaluations and no appreciable differences in mean postoperative correction and complication rates between the three surgical techniques. Peri-operative mortality was 4% and the risk of permanent neurologic sequelae of surgery was 4.9%.

TLD was indirectly addressed by 6 case-series since 1998. The studies followed 271 patients for a mean of 4.0 years, demonstrating a mean total lumbar correction of 39 degrees. Virtually all studies only reported degrees of correction, which were extremely consistent across studies. One study reported consistent SF-36 improvements across all scales compared to baseline.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 26 includes Meta-analysis:	Etame 2008[1]; van Royen 1999[3]
PICO 26 includes RCT:	None
PICO 26 includes Observational studies :	Koller 2013[2]; Wang 2010[4]; Kiaer 2010[5]; van Royen 1998[6]; Min 2007[7]; Chen 2001[8]; Chang 2005[9]

1. Etame AB, Than KD, Wang AC, La MF, Park P. Surgical management of symptomatic cervical or cervicothoracic kyphosis due to ankylosing spondylitis. Spine (Phila Pa 1976) 2008;33:E559-E564.

2. Koller H, Meier O, Zenner J, Mayer M, Hitzl W. Non-instrumented correction of cervicothoracic kyphosis in ankylosing spondylitis: a critical analysis on the results of open-wedge osteotomy C7-T1 with gradual Halo-Thoracic-Cast based correction. Eur Spine J 2013;22:819-32.

3. Van Royen BJ, De Gast A. Lumbar osteotomy for correction of thoracolumbar kyphotic deformity in ankylosing spondylitis. A structured review of three methods of treatment. Ann Rheum Dis 1999;58:399-406.

4. Wang Y, Zhang Y, Mao K, Zhang X, Wang Z, Zheng G, et al. Transpedicular bivertebrae wedge osteotomy and discectomy in lumbar spine for severe ankylosing spondylitis. J Spinal Disord Tech 2010;23:186-91.

5. Kiaer T, Gehrchen M. Transpedicular closed wedge osteotomy in ankylosing spondylitis: results of surgical treatment and prospective outcome analysis. Eur Spine J 2010;19:57-64.

6. Van Royen BJ, de K leuver M, Slot GH. Polysegmental lumbar posterior wedge osteotomies for correction of kyphosis in ankylosing spondylitis. Eur Spine J 1998;7:104-10.

7. Min K, Hahn F, Leonardi M. Lumbar spinal osteotomy for kyphosis in ankylosing spondylitis: the significance of the whole body kyphosis angle. J Spinal Disord Tech 2007;20:149-53.

8. Chen IH, Chien JT, Yu TC. Transpedicular wedge osteotomy for correction of thoracolumbar kyphosis in ankylosing spondylitis: experience with 78 patients. Spine (Phila Pa 1976) 2001;26:E354-E360.

9. Chang KW, Chen YY, Lin CC, Hsu HL, Pai KC. Closing wedge osteotomy versus opening wedge osteotomy in ankylosing spondylitis with thoracolumbar kyphotic deformity. Spine (Phila Pa 1976) 2005;30:1584-93.

IRITIS

PICO 27. In adults with AS, is treatment of acute episodes of iritis by an ophthalmologist more effective in decreasing the severity, duration, or complications of episodes compared to no ophthalmologist care?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 28. In adults with AS, is prescription of topical glucocorticoids for prompt at-home use in the event of eye symptoms effective in decreasing the severity or duration of iritis episodes compared to no at-home use?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 29. In adults with AS and iritis, are TNFi monoclonal antibodies more effective than etanercept in decreasing recurrences of iritis?

Summary: This PICO was not directly addressed by any head-to-head RCTs. Four observational studies or pooled analyses of RCTs compared rates of iritis between patients treated with etanercept and either infliximab (4 studies) or adalimumab (2 studies). All studies reported higher rates among patients treated with etanercept than with infliximab/adalimumab, with relative risks of 8.6, 2.3, 22.7, and infinity.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	TNFi monoclonals	Control: Etanercept	Relative (95% CI)	Absolute		
Iritis flare Rate/100 Pt-Yrs (follow-up 2-16 years; Better indicated by lower values)												
4	observational studies[1]	serious	no serious inconsistency (NC)	no serious indirectness	serious[2]	strong association[3]	339[4]	113[5]	-	mean 28.7 lower (unable to calculate CI)[6]	⊕OOO VERY LOW	IMPORTANT

[1] 3 cohort studies and study of 1 pooled data from RCTs
[2] Unclear how flare was defined and rates varies substantially between cohort studies
[3] Substantial and consistently greater flares for etanercept across all 4 studies
[4] Either infliximab or adalimumab (only 15 total on adalimumab)
[5] Etanercept
[6] Mean rate in etanercept 31.9 flares/100PY; mean rate for monoclonals: 3.2 flares/100PY

PICO 29 includes RCT :	None
PICO 29 includes Observational studies :	Guignard 2006[1]; Braun 2005[2]; Cobo-Ibanez 2008[3]; Fouache 2009[4]

1. Guignard S, Gossec L, Salliot C, Ruyssen-Witrand A, Luc M, Duclos M, et al. Efficacy of tumour necrosis factor blockers in reducing uveitis flares in patients with spondylarthropathy: a retrospective study. Ann Rheum Dis 2006;65:1631-4.

2. Braun J, Baraliakos X, Listing J, Sieper J. Decreased incidence of anterior uveitis in patients with ankylosing spondylitis treated with the anti-tumor necrosis factor agents infliximab and etanercept. Arthritis Rheum 2005;52:2447-51.

3. Cobo-Ibanez T, del Carmen OM, Munoz-Fernandez S, Madero-Prado R, Martin-Mola E. Do TNF-blockers reduce or induce uveitis? Rheumatology (Oxford) 2008;47:731-2.

4. Fouache D, Goeb V, Massy-Guillemant N, Avenel G, Bacquet-Deschryver H, Kozyreff-Meurice M, et al. Paradoxical adverse events of anti-tumour necrosis factor therapy for spondyloarthropathies: a retrospective study. Rheumatology (Oxford) 2009;48:761-4.

PICO 30. In adults with AS who develop iritis while treated with a TNFi, is switching the TNFi more effective in decreasing recurrences of iritis than continuing the same TNFi?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

INFLAMMATORY BOWEL DISEASE

PICO 31. In adults with AS and inflammatory bowel disease, are certain NSAIDs more likely to worsen IBD symptoms than other NSAIDs?

Summary: This PICO was not directly addressed by any studies. Indirect evidence is available from the IBD literature (i.e. patients generally without AS). From this literature, we identified a single RCT that compared celecoxib for a two week exposure to placebo. No statistically significant differences in IBD relapse rates were observed.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 31 includes RCT :	Sandborn 2006[1]
PICO 31 includes Observational studies :	None

1. Sandborn WJ, Stenson WF, Brynskov J, Lorenz RG, Steidle GM, Robbins JL, et al. Safety of celecoxib in patients with ulcerative colitis in remission: a randomized, placebo-controlled, pilot study. Clin Gastroenterol Hepatol 2006;4:203-11.

PICO 32. In adults with AS and inflammatory bowel disease, are certain TNFi more effective in improving outcomes than other TNFi?

Summary: This PICO was directly addressed by 1 study that pooled data from multiple RCTs and 2 open label studies of TNFi for patients with AS. The pooled data study was subsequently revised (with data from an additional study included) in a second report. Infliximab was superior to etanercept, and adalimumab was not statistically different from either. The studies demonstrated wide confidence intervals and high risk of bias. This PICO was indirectly addressed by a single RCT demonstrating the ineffectiveness of etanercept in patients with inflammatory bowel disease without a diagnosis of AS (Sandborn, 2001) [1].

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	Infliximab	Other TNFs (etanercept)	Relative (95% CI)	Absolute		
IBD flares (follow-up 14-156 weeks; measured with: IBD flare or onset; Better indicated by lower values)												
1	randomized trials[1]	very serious[2]	very serious[3]	very serious[4]	very serious[5]	reporting bias[4]	366	419	-	mean 2 lower (0 to 9 higher)	⊕OOO VERY LOW	IMPORTANT

[1] Pooled data from 8 RCTs (1 added later for adalimumab) + 2 open studies.
[2] Double blind and open label studies included
[3] Reviewed literature - multiple studies of unknown quality
[4] Some of the rationale is based on scant (small studied) of observed efficacy of these agents in IBD without AS. It's unclear whether this effect translates into outcomes for IBD in the setting of AS.

[5] Post hoc analysis (Gao) published with support from the pharmaceutical company that markets adalimumab substantially changed the result for adalimumab. These revised results suggest adalimumab produced results in between infliximab and etanercept, but was not statistically different from either.

PICO 32 includes RCT :	Sandborn 2001[1]
PICO 32 includes Observational studies :	Braun 2007[2]; Gao 2012[3]

1. Sandborn WJ, Hanauer SB, Katz S, Safdi M, Wolf DG, Baerg RD, et al. Etanercept for active Crohn's disease: a randomized, double-blind, placebo-controlled trial. Gastroenterology 2001;121:1088-94.

2. Braun J, Baraliakos X, Listing J, Davis J, van der Heijde D, Haibel H, et al. Differences in the incidence of flares or new onset of inflammatory bowel diseases in patients with ankylosing spondylitis exposed to therapy with anti-tumor necrosis factor alpha agents. Arthritis Rheum 2007;57:639-47.

3. Gao X, Wendling D, Botteman MF, Carter JA, Rao S, Cifaldi M. Clinical and economic burden of extra-articular manifestations in ankylosing spondylitis patients treated with anti-tumor necrosis factor agents. J Med Econ 2012;15:1054-63.

PREVENTIVE CARE

PICO 48. In adults with AS, is group or individual self-management education more effective than no formal self-management education in improving outcomes?

Summary: This PICO was directly addressed by 5 RCTs and a non-randomized prospective controlled study (Gross 1981)[1]. All but one of the interventions (Gross 1981)[1] focused upon educational programs related to the performance (self-management) of physical therapy and most relied upon in-person instruction (some used mailed video instructions). The remaining study examined the effect of educational support groups to enhance AS knowledge, adherence to exercise, and knowledge of AS treatments. Descriptions of the specific intervention components were often unclear. The intervention was favored in 8 outcomes while the control group was favored in 1 outcome, and one outcome was not statistically significant.

Quality of Evidence Across All Critical Outcomes: Low ⊕⊕OO

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Other considerations	group or self management education (exercise)	Control: no education	Relative (95% CI)	Absolute		
Health Status: BASDAI (Better indicated by lower values)												
4	randomized trials	serious[1]	no serious inconsistency (64%)	no serious indirectness	no serious imprecision	none	136	136	-	MD 0.66 lower (1.44 to 0.11 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Pain (Better indicated by lower values)												
3	randomized trials	no serious risk of bias	serious[2] (91%)	no serious indirectness	no serious imprecision	none	120	120	-	MD 0.21 higher (0.64 to 1.06 higher)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Fatigue (Better indicated by lower values)												
2	randomized trials	serious[3]	no serious inconsistency (56%)	no serious indirectness	no serious imprecision	none	45	40	-	MD 0.87 lower (2.12 to 0.38 lower)	⊕⊕⊕O MODERATE	CRITICAL
Health Status: Stiffness (Better indicated by lower values)												
1	randomized trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	20	22	-	MD 1.4 lower (2.73 to 0.07 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: SF-36 mental (Better indicated by higher values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 0.12 higher (0.03 to 0.21 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: Depression (Better indicated by lower values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 3.89 lower (5.31 to 2.47 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Health Status: SF-36 social (Better indicated by higher values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 0.09 higher (0.03 lower to 0.21 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: BASFI (Better indicated by lower values)												
4	randomized trials	no serious risk of bias[1]	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	136	136	-	MD 0.87 lower (1.34 to 0.4 lower)	⊕⊕⊕⊕ HIGH	CRITICAL

Functional Status: SF-36 physical (Better indicated by higher values)												
1	randomized trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25	18	-	MD 0.18 higher (0.07 to 0.29 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Functional Status: BASMI (Better indicated by lower values)												
2	randomized trials	no serious risk of bias	no serious inconsistency (0%)	no serious indirectness	no serious imprecision	none	36	38	-	MD 1.86 lower (2.79 to 0.93 lower)	⊕⊕⊕⊕ HIGH	NOT IMPORTANT

[1] High attrition in some studies
[2] Variable findings across studies
[3] Small numbers
[4] Intervention poorly described

PICO 48 includes RCT :	Masiero 2011[2]; Durmus 2009[3]; Sweeney 2002[4]; Kraag 1990[5]; Widberg 2009[6]
PICO 48 includes Observational studies :	Gross 1981[1]

1. Gross M, Brandt KD. Educational support groups of patients with ankylosing spondylitis: a preliminary report. Patient Couns Health Educ 1981;3:6-12.

2. Masiero S, Bonaldo L, Pigatto M, Lo NA, Ramonda R, Punzi L. Rehabilitation treatment in patients with ankylosing spondylitis stabilized with tumor necrosis factor inhibitor therapy: a randomized controlled trial. J Rheumatol 2011;38:1335-42.

3. Durmus D, Alayli G, Cil E, Canturk F. Effects of a home-based exercise program on quality of life, fatigue, and depression in patients with ankylosing spondylitis. Rheumatol Int 2009;29:673-7.

4. Sweeney S, Taylor G, Calin A. The effect of a home based exercise intervention package on outcome in ankylosing spondylitis: a randomized controlled trial. J Rheumatol 2002;29:763-6.

5. Kraag G, Stokes B, Groh J, Helewa A, Goldsmith C. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis--a randomized controlled trial. J Rheumatol 1990;17:228-33.

6. Widberg K, Karimi H, Hafstrom I. Self- and manual mobilization improves spine mobility in men with ankylosing spondylitis--a randomized study. Clin Rehabil 2009;23:599-608.

PICO 49. In adults with AS, is screening for osteopenia/osteoporosis with DEXA scanning yearly, every other year, every five years, more effective than screening after insufficiency fractures or no screening in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 50. In adults with AS and syndesmophytes or spinal fusion, is screening for osteopenia/osteoporosis with DEXA scanning of the hip or other non-spine sites more effective than DEXA scanning of the spine in improving outcomes?

Summary: This PICO was not directly addressed by any studies. The validity of various imaging approaches has been assessed in patients with AS. Some studies have found that low bone density may be significantly more common at the femoral neck (measured by DEXA) than at the lumbar spine but others have found the converse.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 50 includes RCT :	None
PICO 50 includes Observational studies :	Karberg 2005[1]; Toussirot 2001[2]

1. Karberg K, Zochling J, Sieper J, Felsenberg D, Braun J. Bone loss is detected more frequently in patients with ankylosing spondylitis with syndesmophytes. J Rheumatol 2005;32:1290-8.

2. Toussirot E, Michel F, Wendling D. Bone density, ultrasound measurements and body composition in early ankylosing spondylitis. Rheumatology (Oxford) 2001;40:882-8.

PICO 51. In adults with AS, is fall evaluation and counseling more effective than no evaluation and counseling in improving outcomes?

Summary: This PICO was not directly addressed by any studies.
Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 52. In adults with AS, is screening for cardiac conduction defects with electrocardiogram at diagnosis, yearly, every other year, or every five years more effective than no screening in improving outcomes?

Summary: This PICO was not directly addressed by any studies.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 53. In adults with AS, is screening for valvular heart disease with echocardiogram at diagnosis, yearly, every other year, or every five years more effective than no screening in improving outcomes?

Summary: This PICO was not directly addressed by any RCTs and was indirectly addressed by 1 observational cohort study. A cross-sectional component of this study examined aortic and valvular abnormalities in patients with AS (n=44) and controls (n=30). Abnormalities were discovered in 82% (AS) vs. 27% (controls). Follow-up of 25 patients revealed new aortic root or valve abnormalities in 24%, worsening of existing valve regurgitation in 12%, and resolution of abnormalities in 20%. Twenty percent of patients developed heart failure, underwent valve replacement, had a stroke, or died, as compared with 3% of control subjects (unclear duration of follow-up).

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 53 includes RCT :	None
PICO 53 includes Observational studies :	Roldan 1998[1]

1. Roldan CA, Chavez J, Wiest PW, Qualls CR, Crawford MH. Aortic root disease and valve disease associated with ankylosing spondylitis. J Am Coll Cardiol 1998;32:1397-404.

DISEASE ACTIVITY MEASURES

PICO 54. In adults with active or stable AS, is regular interval use and monitoring of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) or AS Disease Activity Score (ASDAS) more effective than usual care without monitoring of the BASDAI or ASDAS in improving outcomes?

Summary: This PICO was not directly addressed by any studies, as confirmed by a recent systematic review and our own systematic search. Four RCTs included in this systematic review used these measures as a decision point to determine subsequent therapy, but this was not regular interval use. Many studies have assessed the validity of these measures.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 54 includes Meta-analysis:	Schoels 2014[1]
PICO 54 includes RCT :	None
PICO 54 includes Observational studies :	None

1. Schoels MM, Braun J, Dougados M, Emery P, Fitzgerald O, Kavanaugh A, et al. Treating axial and peripheral spondyloarthritis, including psoriatic arthritis, to target: results of a systematic literature search to support an international treat-to-target recommendation in spondyloarthritis. Ann Rheum Dis 2014;73:238-42.

PICO 56. In adults with active or stable non-radiographic axial SpA, is regular interval use and monitoring of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) or AS Disease Activity Score (ASDAS) more effective than usual care without monitoring of the BASDAI or ASDAS in improving outcomes?

Summary: This PICO was not directly addressed by any studies. See PICO 54

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 55. In adults with AS, is regular interval use and monitoring of C-reactive protein (CRP) levels or erythrocyte sedimentation rate (ESR) more effective than usual care without regular CRP or ESR monitoring in improving outcomes?

Summary: This PICO was not directly addressed by any studies, as confirmed by a recent systematic review and our own systematic search. Two RCTs used these measures as a decision point to determine subsequent therapy, but this was not regular interval use. Many studies have assessed the validity of these measures.

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO

PICO 55 includes Meta-analysis:	Schoels 2014[1]
PICO 55 includes RCT :	None
PICO 55 includes Observational studies :	None

1. Schoels MM, Braun J, Dougados M, Emery P, Fitzgerald O, Kavanaugh A *et al.* Treating axial and peripheral spondyloarthritis, including psoriatic arthritis, to target: results of a systematic literature search to support an international treat-to-target recommendation in spondyloarthritis. Ann Rheum Dis 2014;73:238-42.

PICO 57. In adults with non-radiographic axial SpA, is regular interval use and monitoring of C-reactive protein (CRP) levels or erythrocyte sedimentation rate (ESR) more effective than usual care without regular CRP or ESR monitoring in improving outcomes?

Summary: This PICO was not directly addressed by any studies. See PICO 55

Quality of Evidence Across All Critical Outcomes: Very low ⊕OOO