The value of correctly diagnosing axial spondyloarthritis for patients and society

Webers, C.; Grimm, S.; Tubergen, A. van; Gaalen, F. van; Heijde, D. van der; Joore, M.; Boonen, A.

Seminars in Arthritis and Rheumatism 62 (2023) 152242

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

The value of correctly diagnosing axial spondyloarthritis for patients and society

Casper Webers [a,b,*], Sabine Grimm [c], Astrid van Tubergen [a,b], Floris van Gaalen [d], Désirée van der Heijde [d], Manuela Joore [c], Annelies Boonen [a,b]

[a] Department of Internal Medicine, Department of Rheumatology, Maastricht University Medical Centre, Maastricht, the Netherlands
[b] Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, the Netherlands
[c] Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre, Maastricht, the Netherlands
[d] Department of Rheumatology, Leiden University Medical Centre, Leiden, the Netherlands

ARTICLE INFO

Keywords:
Axial spondyloarthritis
Diagnosis
Value
Health economic evaluation
Cost-effectiveness
Quality of life

ABSTRACT

Objective: To demonstrate the value of diagnosing axSpA, by comparing health and costs associated with available diagnostic algorithms and perfect diagnosis.
Methods: Using data from SPACE and other cohorts, a model was developed to estimate health (quality-adjusted life-years, QALYs) and costs (healthcare consumption and work productivity losses) of different diagnostic algorithms for axSpA amongst patients with low back pain referred to a rheumatologist, over a 60-year horizon. The model combined a decision-tree (diagnosis) with a state-transition model (treatment). The three algorithms (Berlin [BER, highest specificity], Modification 1 [M1; less strict inflammatory back pain (IBP) criterion] and Modification 2 [M2; IBP not mandatory as entry criterion, highest sensitivity]) were compared. Changes in sensitivity/specificity were explored and the value of perfect diagnosis was investigated.
Results: For each correctly diagnosed axSpA patient, up to 4.7 QALYs and €60,000 could be gained/saved, considering a societal perspective. Algorithm M2 resulted in more health and lower costs per patient (24.23 QALYs; €157,469), compared to BER (23.96 QALYs; €159,423) and M1 (24.15 QALYs; €158,417). Hypothetical improvements in M2 sensitivity resulted in slightly more value compared to improvements in specificity. Perfect diagnosis can cost €7,500 per patient and still provide enough value.
Conclusion: Correct diagnosis of axSpA results in substantial health and cost benefits for patients and society. Not requiring IBP as mandatory for diagnosis of axSpA (algorithm M2) provides more value and would be preferable. A considerably more expensive diagnostic algorithm with better accuracy than M2 would still be considered good value for money.

Introduction

Axial spondyloarthritis (axSpA) is a chronic inflammatory disorder of the sacroiliac joints (SIJ) and spinal entheses [1]. Low back pain is the most characteristic symptom. AxSpA is considered a spectrum of disease, with a distinction between radiographic axSpA (r-axSpA; with structural damage to the SIJ on plain radiography [X-SIJ]) and non-radiographic axSpA (nr-axSpA; without damage on the X-SIJ). [2] Until 2009, the larger concept of axSpA was not formally recognized and only the r-axSpA subtype, traditionally referred to as ankylosing spondylitis (AS), was an accepted diagnosis. Following diagnosis, pharmacological and non-pharmacological treatment options are available. Nonsteroidal anti-inflammatory drugs (NSAIDs) are recommended as initial therapy. If those fail and the disease is active, biological disease-modifying antirheumatic drugs (bDMARDs) are indicated. [3]

Diagnosing axSpA is challenging and based on the judgement of the rheumatologist, who takes clinical and other parameters into account. The choice for a diagnostic work-up or diagnostic test is typically based on its test characteristics. However, these characteristics do not reflect the actual value of a diagnostic work-up for patients and for society: they provide insight into the extent of (mis)diagnosis that will occur, but not

* Corresponding author: Casper Webers, Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre, PO Box 5800, Maastricht 6202 AZ, The Netherlands.
E-mail address: cjp.webers@maastrichtuniversity.nl (C. Webers).

https://doi.org/10.1016/j.semarthrit.2023.152242

Available online 5 July 2023

the consequences. This is important, because correct and incorrect diagnosis of patients have important long-lasting consequences in terms of health (for patients) and costs (for society). Underdiagnosis or missed diagnosis results in missed opportunities for timely treatment and decreasing the burden of disease. Overdiagnosis, on the other hand, leads to potential overtreatment with (potentially expensive) drugs and put patients at risk of unnecessary adverse effects. Although these consequences are usually considered in an informal way when assessing diagnosis, they are not explicitly estimated in terms of health and costs.

In order to enhance recognition of patients with axSpA in rheumatology clinical practice, three algorithms for diagnosing axSpA in patients with chronic back pain have been developed and validated. [4,5] First, the original Berlin algorithm (Fig. 1) [4] used inflammatory back pain (IBP) according to the Calin criteria as obligatory entry criterion. [6] This has limited sensitivity, as not all patients with axSpA fulfil this criterion. [7–9] Later, the Berlin algorithm was changed, resulting in Modifications 1 and 2. [5] Modification 1 loosened the definition of IBP (at least 3/5 of the Assessment of SpondyloArthritis international Society [ASAS] IBP criteria instead of ≥4/5), [10] while Modification 2 removed IBP altogether as obligatory entry criterion, instead adding it as an SpA feature. [5] Both modifications have higher sensitivity, yet lower specificity. Modification 2 was accepted as the algorithm of choice by ASAS. [5] However, the health and economic value of diagnosing axSpA using any of these algorithms has never been investigated. It is unknown

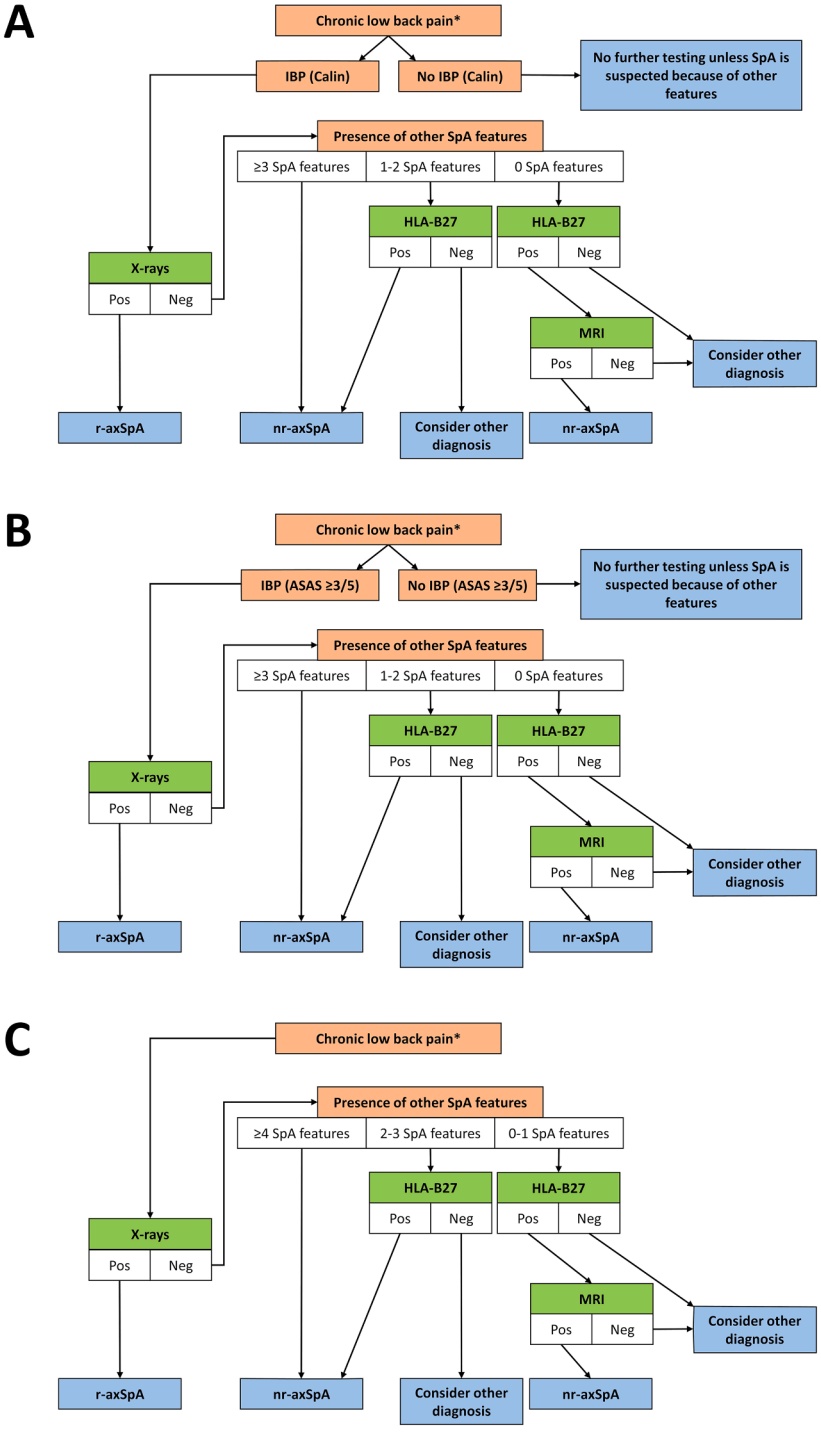

Fig. 1. Diagnostic algorithms for axSpA
The original Berlin algorithm (A), and ASAS Modifications 1 (B) and 2 (C) (adapted from: [4,5]). *Duration >3 months, onset before the age of 45.
axSpA, axial spondyloarthritis; ASAS, Assessment of SpondyloArthritis international Society; HLA-B27, human leucocyte antigen B27; IBP, inflammatory back pain; Neg, negative; nr-axSpA, non-radiographic axial spondyloarthritis; Pos, positive; r-axSpA, radiographic axial spondyloarthritis; SpA, spondyloarthritis.

to what extent it is acceptable (in terms of health and costs) to increase sensitivity, i.e. reducing underdiagnosis, at the expense of specificity, i.e. increasing overdiagnosis. In other words: would Modification 2 also be superior to the Berlin algorithm when long-term health and costs of underdiagnosis/overdiagnosis are considered? And even more important for future direction: none of these algorithms is perfect. How much can we gain in patient health by improving diagnosis of axSpA, and how much could this cost while still being worth it? And should we try to improve sensitivity or specificity?

For this study, our objective was to estimate the value (health and cost benefits) of correctly diagnosing axSpA for patients and society. Next, we compared the lifetime health and costs (cost-effectiveness) of the three available algorithms to diagnose axSpA, followed by treatment as currently recommended, from Dutch societal and healthcare perspectives. In addition, we investigated a perfect diagnosis (sensitivity and specificity 100%), to demonstrate the potential gain in costs and effects if no overdiagnosis/underdiagnosis occurred.

Material and methods

The development of the economic model is described in detail in Supplementary File S1.

Population, interventions, comparators

The modelled population consisted of patients referred to the rheumatologist with suspected axSpA, thus containing a mix of chronic low back pain (CLBP) patients with r-axSpA, nr-axSpA or non-axSpA CLBP (such as non-specific back pain). The characteristics of this modelled population were based on patient-level data on patient and disease characteristics in the SPondyloArthritis Caught Early (SPACE) cohort, at time of presentation to the rheumatologist. [5] In the remainder of this article, "CLBP" is specifically used for cases with non-axSpA CLBP. Subtype-specific parameters were used for r-axSpA and nr-axSpA, whenever relevant.

Three existing algorithms were investigated: the Berlin algorithm

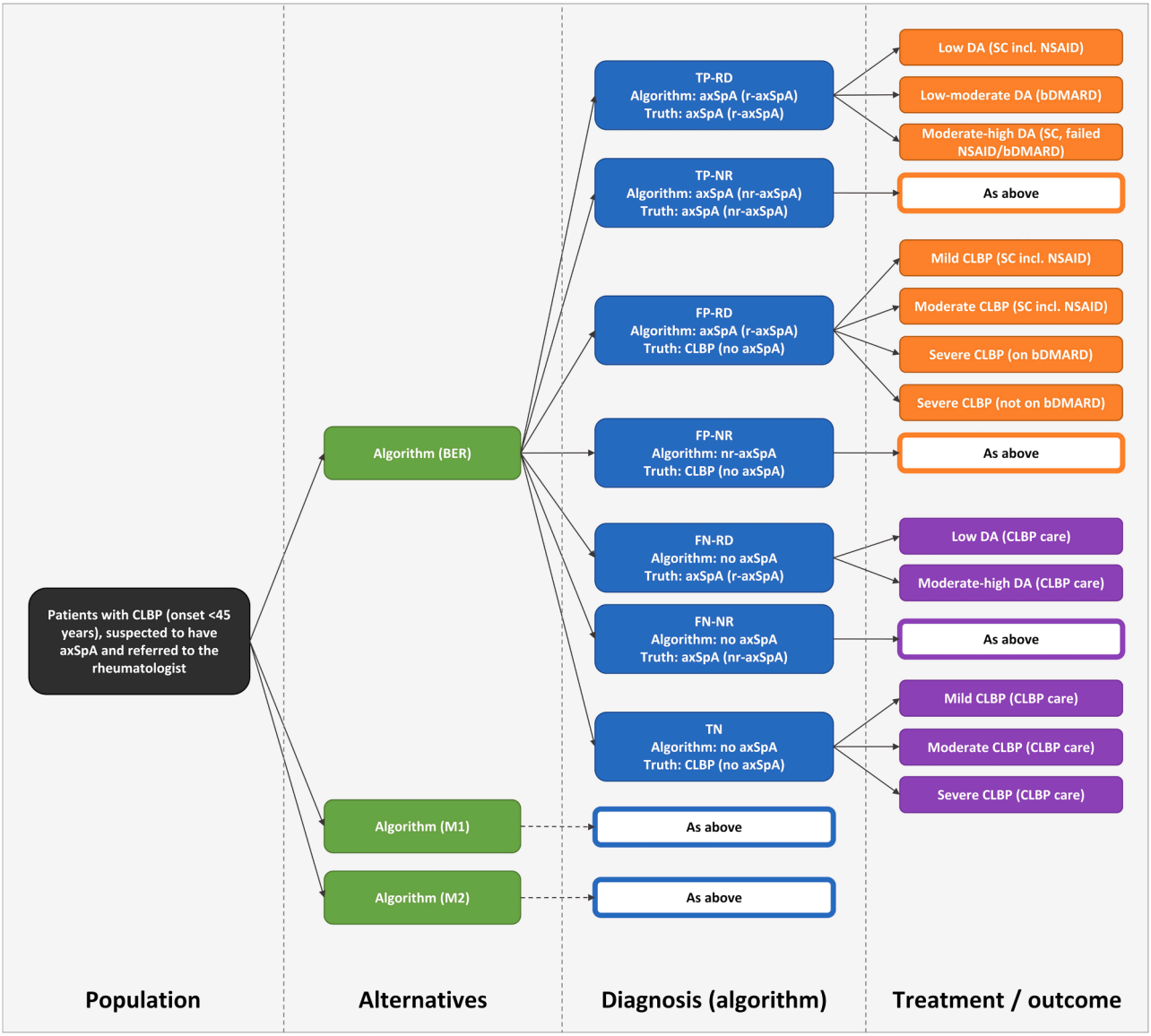

Fig. 2. Overview of the decision-tree component of the model
Decision-tree model of diagnosis of axSpA.
axSpA, axial spondyloarthritis; bDMARD, biological disease-modifying antirheumatic drug; BER, Berlin; CLBP, chronic low back pain; DA, disease activity; FN, false negative; FP, false positive; M1, Modification 1; M2, Modification 2; NR/nr-axSpA, non-radiographic axial spondyloarthritis; NSAID, nonsteroidal anti-inflammatory drug; RD/r-axSpA, radiographic axial spondyloarthritis; SC, supportive care; TN, true negative; TP, true positive.

(BER) and Modifications 1 (M1) and 2 (M2) (Fig. 1). In addition, two hypothetical algorithms were investigated. First, a perfect algorithm (PER; 100% sensitivity/specificity) was added. Second, an algorithm was added in which only r-axSpA would be perfectly diagnosed and treated, while nr-axSpA was not acknowledged and not treated as axSpA (RAD; 100% r-axSpA sensitivity/0% nr-axSpA sensitivity/100% specificity).

Model structure

The model consisted of two parts. Part 1 represented the diagnostic process, i.e. the performance of each algorithm, as a decision-tree (Fig. 2). Part 2 represented disease management, i.e. the consequences of (mis)diagnosis, as a state-transition model with a time horizon of 60 years (Fig. 3). This 'lifetime' time horizon was chosen to reflect the

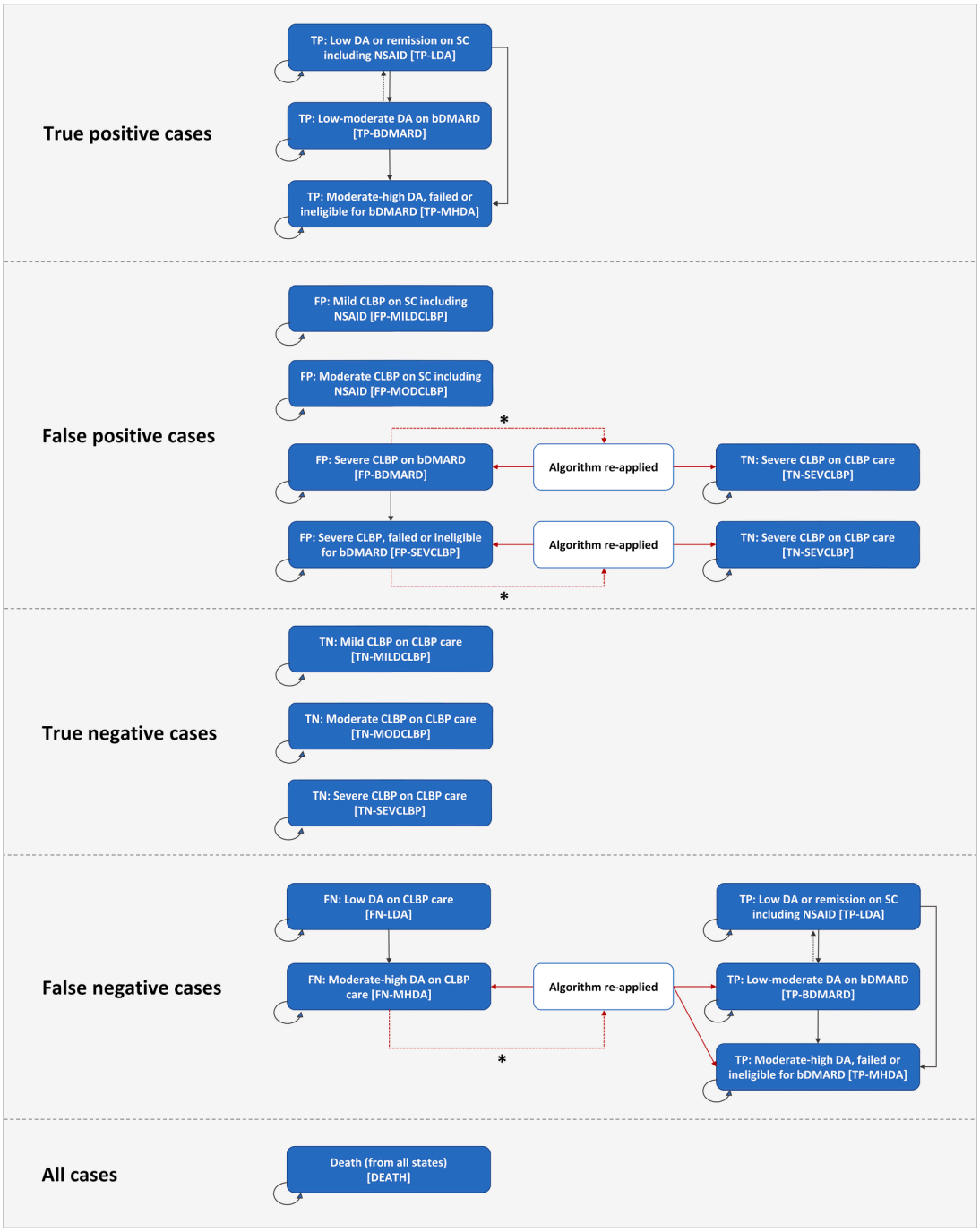

Fig. 3. Overview of the state-transition component of the model
State-transition model of axSpA and CLBP. After each cycle of 3 months, cases either remain in their current health state (circular arrows) or move to a different health state (straight black arrow). Note: the bDMARD-associated health states include several tunnel states (states in which cases can only spend one cycle, used to account for transition probabilities that are not fixed), not shown in the diagram above.
*Re-application of the algorithm is a one-time event, after which a proportion of misdiagnosed cases is correctly diagnosed (reconsideration of diagnosis) and enters a health state associated with the correct diagnosis.
bDMARD, biological disease-modifying antirheumatic drug; CLBP, chronic low back pain; DA, disease activity; FN, false negative; FP, false positive; NSAID, nonsteroidal anti-inflammatory drug; SC, supportive care; TN, true negative; TP, true positive.

potential long-term effects of adopting a certain diagnostic strategy. In a state-transition model, patients from a hypothetical cohort are assumed to reside in one of a finite number of health states at any point in time, and can make transitions between these health states over a series of discrete time periods (cycles), or remain in their current health state. [11,12] Costs and outcomes are assigned to each health state, allowing long-term costs and outcomes to be estimated.

Part 1: diagnosis (Fig. 2)

Each patient in the model entered the decision-tree at baseline and was "diagnosed" by one of the algorithms (which could be correct or incorrect, with varying misclassification depending on the algorithm's test characteristics). The external standard used to determine the patient's real diagnosis was an expert rheumatologist's opinion at the baseline assessment (time of presentation) in SPACE, who had all data available relevant for diagnosing axSpA. Of note, only SPACE participants whose rheumatologist had made the diagnosis (axSpA/no axSpA) with a level of confidence of at least 7 out of 10 (higher levels meaning more confidence in correct diagnosis) were used to inform the model. Patients were labelled true positive (TP; axSpA patients correctly labelled as "r-axSpA"/"nr-axSpA" by the algorithm), false positive (FP; CLBP patients mislabelled as "r-axSpA"/"nr-axSpA" by the algorithm), false negative (FN; axSpA patients mislabelled as "CLBP" by the algorithm) or true negative (TN; CLBP patients correctly labelled as "CLBP" by the algorithm). The flow of cases through this decision-tree was assumed to occur instantaneously.

Part 2: disease management (Fig. 3)

After patients received their algorithm-defined diagnostic label, they were assumed to be treated accordingly: TP and FP cases received axSpA treatment, TN and FN cases received CLBP treatment. Disease parameters and progression were assumed to be driven by the underlying true disease (axSpA or CLBP), and not the algorithm's diagnostic label ("axSpA" or "CLBP"). This part of the model was structured as a state-transition model with a 3-month cycle duration and 60-year time horizon, starting immediately after the decision-tree. Based on the outcome of the decision-tree, cases flowed into different health states, which were based on combinations of disease activity/severity states and receiving (or not) a bDMARD (see below). TP and FP cases were assumed to receive a 4-week NSAID trial before entering a health state.

For axSpA patients (TP/FN), Bath AS Disease Activity Index (BASDAI) and Bath AS Functional Index (BASFI) were used to model health outcome, as both domains captured by these measures are relevant when modelling health-related quality of life and costs in axSpA. [13,14] Three axSpA health states were defined: "low disease activity" without a bDMARD (LDA), "moderate/high disease activity" without a bDMARD (MHDA), and "on bDMARD" (BDMARD; mix of patients with low/moderate/high disease activity). These health states were chosen to reflect how disease activity (low versus moderate/high) influences clinical practice and resource use, with the addition of a bDMARD-specific health state as bDMARDs likely have substantial impact on both costs and effects. Patients could receive up to 3 sequential bDMARDs over time, and bDMARDs were assumed to be successfully tapered – but not discontinued – in a proportion of patients. Disease progression (accumulation of structural damage over time) was reflected by an increase in BASFI.

For CLBP patients (TN/FP), three CLBP health states were defined: "mild" (MILDCLBP), "moderate" (MODCLBP) and "severe" (SEVCLBP). These states were based on trajectory research in CLBP populations, [15–18] and the Global Burden of Disease (GBD) study. [19] As (C)LBP trajectories are relatively stable in individual patients, even over extended periods of time, CLBP cases were fixed within their health state, and no disease progression was assumed to occur. [16,18] For FP cases, a fourth health state was added ("severe, on bDMARD" [BDMARD]), as some of these patients were expected to receive (up to 3 sequential) bDMARDs.

A proportion of the cases initially misdiagnosed by the algorithm in the decision-tree (FP and FN) could switch to their correct counterpart states in the model (TN and TP, respectively) after 5 years and receive appropriate treatment (Fig. 3). This 'reconsideration of diagnosis' was carried out by re-applying the algorithm in the FP/FN cases with higher symptom burden (MHDA for axSpA cases, severe CLBP for CLBP cases) at 5 years after initial diagnosis, assuming the same test characteristics.

Data on diagnosis, disease (progression) and treatment

Data sources included patient-level data, published (group-level) data and guidelines. Data on AS were considered to also apply to r-axSpA. [20] Patient-level data from the SPACE cohort were a key source (Supplementary File S1, p. 36–37), and were analysed to inform demographics, test characteristics of the algorithms, disease severity of axSpA and CLBP, and productivity losses. [5] Patient characteristics and axSpA/CLBP disease parameters in SPACE were used to simulate the diagnostic algorithms and allocate cases to health states after diagnosis. Disease progression, transition probabilities and treatment of axSpA (including serious adverse events [SEAs]) were based on published data from observational cohorts (Supplementary File S1, p. 37–38). [21,22] Treatment of CLBP was based on published interventional and observational studies (see Supplementary File S1, p. 38). [23–26] Tables 1 and 2 present key model parameters and the diagnostic outcome of the algorithms, respectively. Supplementary File S2 contains all model parameters and their source.

Resource use and costs

The average diagnostic work-up costs of each algorithm were applied to all cases at the start of the state-transition model. Healthcare use and productivity losses for axSpA and CLBP were considered to be different and driven by health consumption behaviour matching the true underlying disease (axSpA or CLBP), and not the algorithm's diagnostic label ("axSpA" or "CLBP"). An exception to this were treatment choice and treatment-related healthcare consumption, as these would be closely linked to the algorithm's diagnostic label (e.g. it was assumed a patient labelled as "axSpA" by the algorithm would receive axSpA treatment, whether it was TP or FP).

For axSpA, resource use was modelled separately for each health state over time using data from the Outcome in Ankylosing Spondylitis International Study (OASIS). [14] Use of NSAIDs and bDMARDs depended on health state. For CLBP, resource use for severe cases was based on trial results, [27] and applied proportionally to cases with mild or moderate CLBP. [28] Medication costs and productivity losses for CLBP were estimated in a similar manner as for axSpA. Productivity losses were estimated by applying employment and work disability ratios (axSpA/CLBP vs. general population [29–31]) to general population rates, and by regression functions derived from SPACE for absenteeism (missed work days) and presenteeism (decreased at-work productivity). In the base-case analyses, both human capital and friction cost approaches were applied for valuation of productivity losses, [32] and presenteeism-associated costs were not included. The human capital approach counts any hour not worked as lost, while the friction cost approach only counts as lost those hours not worked until the patient is replaced.

Prices and costs were expressed in 2019 Euros. Unit costs for healthcare use, drugs and productivity losses were guided by the Dutch guideline for economic evaluations. [33] For bDMARDs, discounts were applied to list prices, as list prices were not representative of daily practice in the Netherlands according to local experts on drug pricing. Costs were discounted at 4% per year. [33]

Table 1
Key characteristics and parameters.

Parameter	Value					Source
General						
Horizon, years	60					Assumption (lifetime)
Discount rate per year, costs	4.0%					Dutch National Health Care Institute [33]
Discount rate per year, utility	1.5%					Dutch National Health Care Institute [33]
Demographics						
Age at time of first presentation at rheumatologist, years	31					SPACE
Proportion of males, axSpA	0.47					SPACE
Proportion of males, CLBP	0.27					SPACE
Prevalence of axSpA in referred population	0.36					SPACE
Prevalence of r-axSpA in referred population with axSpA	0.29					SPACE
Symptom duration in referred population, years	1.11					SPACE
Algorithm-dependant parameters	**Value**					**Source**
Diagnostics	**BER**	**M1**	**M2**	**PER**	**RAD**	
Algorithm sensitivity, r-axSpA,%	52%	84%	90%	100%	100%	SPACE
Algorithm sensitivity, nr-axSpA,%	50%	69%	76%	100%	0%	SPACE
Algorithm specificity,%	88%	82%	85%	100%	100%	SPACE
Average cost of algorithm per patient, euros*	€237	€258	€279	€0	€248	Dutch cost manual [33,49], NzA tariffs, SPACE
Algorithm-independent parameters	**Value**					**Source**
State transitions	**r-axSpA**		**nr-axSpA**			
Eligible for bDMARD if high disease activity,%	85%		75%			Assumption (expert opinion)
Probability of switching to second bDMARD if failure first bDMARD	0.47		0.47			Yahya, 2017 [48]
Probability of switching to third bDMARD if failure second bDMARD	0.47		0.47			Assumption (expert opinion)
Disease severity and treatment effects – axSpA	**r-axSpA**		**nr-axSpA**			
Baseline BASDAI/BASFI at time of diagnosis (starting NSAID)	3.5 / 2.5		4.3 / 3.3			SPACE
Baseline BASDAI/BASFI in those starting bDMARD	5.6 / 4.2		5.6 / 4.2			Ørnbjerg, 2019 (EuroSpA) [22]
Initial BASDAI/BASFI change when entering LDA state (effect NSAID)	−1.8 / −1.1		−1.8 / −1.1			Weighted effect in RCTs
Initial BASDAI/BASFI change when entering bDMARD state (effect bDMARD)	Range −3.1 to −1.7 / −2.2 to −1.2†					Ørnbjerg, 2019 (EuroSpA) [22]
Utility – axSpA						
Utility in health states (LDA, BDMARD, MHDA)	Based on BASDAI / BASFI / age					Mapping algorithm (Wailoo, 2015 [34])
Disease severity and treatment effects – CLBP	**TN**		**FP**			
Level of back pain (0–10) in 'mild CLBP' state	1.1		1.5			SPACE; assumed effect of CLBP care for TN
Level of back pain (0–10) in 'moderate CLBP' state	3.4		4.7			SPACE; assumed effect of CLBP care for TN
Level of back pain (0–10) in 'severe CLBP' state	5.4		7.4			SPACE; assumed effect of CLBP care for TN
Level of limitations in physical function§ (0–10) in 'mild CLBP' state	1.7		2.3			SPACE; assumed effect of CLBP care for TN
Level of limitations in physical function§ (0–10) in 'moderate CLBP' state	1.8		2.4			SPACE; assumed effect of CLBP care for TN

(*continued on next page*)

Table 1 (*continued*)

Parameter	Value		Source
Level of limitations in physical function[§] (0–10) in 'severe CLBP' state	3.0	4.1	SPACE; assumed effect of CLBP care for TN
Utility – CLBP	**TN**	**FP**	
Disutility compared to general population for mild CLBP	0.02	0.02	Global Burden of Disease Study, 2013 [19]; RCTs/CUAs
Disutility compared to general population for moderate CLBP	0.04	0.05	CLBP treatment (Smeets, 2009 [23]; Goossens, 2015 [24];
Disutility compared to general population for severe CLBP	0.27	0.37	Lamb, 2010 [25]; Artus, 2014 [26]; amongst others)
Medication-associated costs			
Cost of bDMARD, euros per cycle	Time-dependent (range €1444-€2911)		Medicijnkosten.nl (Dutch drug list prices) and expert opinion; assumed biosimilar preference

*Includes costs of visits to rheumatologist, and imaging (X-SIJ, MRI-SIJ) and HLA-B27 tests. The average costs reported here takes into account that, depending on the algorithm and presence of SpA features, not all patient undergo all tests. For PER, costs were set to €0, to investigate the value of an hypothetical perfect replacement test. For RAD, costs were assumed to only include visits to the rheumatologist and X-SIJ.

† Depending on 1st/2nd/3rd bDMARD and time since bDMARD initation.

§ In cases with CLBP, limitations in physical function were assessed with the Bath Ankylosing Spondylitis Functional Index (BASFI). Although not developed for use in CLBP, the BASFI was used in this case as it was the only available measure of physical function available. Of note, in our model, BASFI was only used for CLBP cases for modelling work productivity (where it actually outperformed back pain with regard to model fit in regression models).

axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; bDMARD, biological disease-modifying antirheumatic drug; BER, Berlin algorithm; CLBP, chronic low back pain; CUA, cost-utility analysis; FN, false negative; FP, false positive; LDA, low disease activity; M1, modification 1; M2, modification 2; MHDA, moderate/high disease activity; nr-axSpA, non-radiographic axSpA; NSAID, non-steroidal anti-inflammatory drug; NzA, Dutch Healthcare Authority (Nederlandse Zorgautoriteit); r-axSpA, radiographic axSpA; RCT, randomized controlled trial; SPACE, SPondyloArthritis Caught Early; TN, true negative; TP, true positive.

Table 2
Distribution of the modelled cohort population across diagnostic groups in the decision-tree, by algorithm.

Algorithm	SE_1 / SE_2 / SP (%)*	Proportion entering each diagnostic group (%)			
		TP (TP_1, TP_2)†	FP‡	FN (FN_1, FN_2)†	TN‡
Berlin (BER)	52 / 50 / 88	18 (6, 13)	8	18 (5, 13)	56
Modification 1 (M1)	84 / 69 / 82	27 (9, 18)	11	10 (2, 8)	52
Modification 2 (M2)	90 / 76 / 85	29 (10, 19)	10	7 (1, 6)	54
Perfect diagnosis (PER)	100 / 100 / 100	36 (11, 26)	0	0 (0, 0)	64
Only r-axSpA (RAD)	100 / 0 / 100	11 (11, 0)	0	26 (0, 26)	64

The proportions of the modelled cohort entering each diagnostic group represent the outcome of the algorithms at baseline (decision-tree). Of note, in the base-case analysis of the existing algorithms (BER, M1 and M2), some of the cases initially misdiagnosed in the decision-tree (FP and FN) could switch to their correct counterpart states (TN and TP, respectively) after 5 years. Consequently, after 5 years, the proportions of FP and FN in the modelled cohort will be lower than those reported in the table above (and TN and TP will be higher).

*SE_1 = sensitivity for r-axSpA, SE_2 = sensitivity for nr-axSpA, SP = specificity.

†Proportion of total cohort that has axSpA and is correctly diagnosed (TP) or underdiagnosed (FN) by the algorithm, for any type of axSpA (TP / FN), r-axSpA (TP_1 / FN_1) or nr-axSpA (TP_2 / FN_2).

‡Proportion of total cohort that has CLBP and is overdiagnosed (FP) or correctly excluded (TN).

axSpA, axial spondyloarthritis; CLBP, chronic low back pain; FN, false negative; FP, false positive; nr-axSpA, non-radiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; SE, sensitivity; SP, specificity; TN, true negative; TP, true positive.

Health utilities and quality-adjusted life-years (QALYs)

For axSpA, a mapping algorithm was used to estimate (EQ-5D) utility, based on age, BASDAI and BASFI. [34] Considering disease progression, utilities were calculated for each health state separately over time. A fixed SAE-associated disutility was proportionally applied to cases in medication-associated states, taking into account the probability of various SAEs occurring (see Supplementary File S2 for SAE-associated parameters and their source). [35–40]

For CLBP, utility was modelled by adding a CLBP-associated disutility to age/gender-matched general population EQ-5D utility. This disutility was based on disability weights and their distribution from the health states in the GBD study, [19] and was assumed to be smaller in TN cases (compared to FP cases) as these were assumed to receive CLBP care. SAE-associated disutility was applied in a similar manner as for axSpA. Utilities were discounted at 1.5% per year. [33]

Analysis

The model was developed in Excel 2016 (Microsoft Corporation). Analysis of patient-level data to derive parameter inputs was conducted in Stata SE Release 14.0 (StataCorp).

Absolute QALYs and costs were estimated and compared for correctly and incorrectly diagnosed patient groups. The incremental cost per QALY gained (incremental cost-utility ratio, ICUR) and incremental net monetary benefit (iNMB) were calculated for each comparison (e.g. M2 versus BER, M2 versus M1, PER versus M2). Also, the diagnostic accuracy of the optimal (most cost-effective) existing algorithm was changed, to assess whether future gains in sensitivity or specificity would result in most value. Finally, the hypothetical "only r-axSpA acknowledged" alternative (RAD) was compared against PER. As recommended in the Netherlands, the threshold for willingness-to-pay (WTP) was based on the burden of disease, and set at 20,000€/QALY. [41] Deterministic one-way sensitivity analysis (DOWSA) and

probabilistic sensitivity analysis (PSA) were conducted to explore the impact of uncertainty. For PSA, model parameters were randomly sampled 1000 times from their appropriate distributions. In scenario analyses, key assumptions were changed. For all analyses, societal (including healthcare, informal care and work-related costs) and healthcare (only healthcare costs) perspectives were adopted.

Results

Value of a correct diagnosis

In the modelled cohort, 36% had axSpA (Table 1). To demonstrate the value of a correct diagnosis, QALYs and costs of TP and TN cases were compared with FN and FP cases, respectively. Over 60 years, the average TP case accumulated 2.9 more QALYs than an FN case (range 1.7–4.7, depending on axSpA subtype/severity), and the average TN case accumulated 2.3 more QALYs than an FP case (range 0.2–3.5, depending on severity) (Supplementary Tables S1.1-S1.3). Similarly, a correct versus incorrect diagnosis of axSpA or non-axSpA resulted in cost savings for society ranging from €10,000 to over €60,000.

Comparing existing algorithms (BER/M1/M2)

The distribution of the modelled cohort population across diagnostic groups (TP/FP/FN/TN) for each algorithm is shown in Table 2. In the base-case analysis, using a societal perspective, application of the M2 algorithm resulted in more health and lower costs per patient (24.23 QALYs; €157,469) compared to BER (23.96 QALYs; €159,423) and M1 (24.15 QALYs; €158,417) over a 60-year time horizon, thus dominating BER and M1 (Table 3). From a healthcare perspective, BER was associated with lowest costs (€91,833) compared to M1 (€92,890) and M2 (€92,837). Notwithstanding, M2 was cost-effective, with an ICUR of €3495/QALY compared to BER, and dominating M1 (Table 3).

The uncertainty in the following parameters had most impact on the cost-effectiveness of the existing algorithms: time axSpA cases spend in the LDA state (before transitioning to the BDMARD/MHDA states), discount on bDMARD price, severity of axSpA in the TP population, sensitivity of the algorithms, and prevalence of axSpA (Fig. 4, Supplementary Figures S2.1-S2.8). In all DOWSA analyses, M2 remained cost-effective compared to BER and M1. At a WTP threshold of 20,000€/QALY, the probability of M2 being most cost-effective was 92.9%, compared to 6.3% for M1 and 0.8% for BER (societal perspective, Supplementary Figures S1.1-S1.6). In all scenario analyses, M2 also remained most cost-effective (Supplementary Tables S2.1-S2.17).

Improving diagnostic accuracy

Hypothetical improvements in sensitivity of M2 (the most cost-effective algorithm) resulted in most QALYs, while improvements in specificity led to lowest costs, the former resulting in slightly more value (iNMB €2380 vs €2050 for 10% increase in sensitivity or specificity, respectively, when compared to base-case M2) (Supplementary Tables S3.1-S3.3). These results were largely driven by the assumed prevalence of axSpA in the modelled cohort: higher prevalence favoured sensitivity, while lower prevalence favoured specificity (Supplementary Figure S3.1). Also, uncertainty was substantial, as the probability of improving sensitivity to be superior in terms of health and costs to improving specificity was 48% at a WTP threshold of 20,000€/QALY (Supplementary Figures S4.1-S4.2).

Compared to M2, application of the perfect diagnosis (PER) led to more QALY (24.44 [PER] vs 24.23 [M2] QALYs) and less costs (€154,247 vs €157,469) per patient (Table 3). This comparison resulted in an iNMB of €7467 over 60 years. This implies that a perfect diagnostic test could cost almost €7500 while still providing enough value in terms of health and costs to be acceptable. Scenario analyses provided varying estimates (Supplementary Tables S4.1-S4.17). If it was assumed that any misdiagnosis would be corrected after 5 years, a perfect diagnostic test could cost €2606 (Supplementary Table S4.11). In uncertainty analyses of the perfect scenario, similar parameters as for the existing algorithms were the main drivers of costs and effects (Supplementary Figures S5.1-S5.3).

Discussion

In this analysis, we demonstrated for the first time the health and cost gains associated with correctly diagnosing or ruling out axSpA. Of

Table 3
Pairwise comparisons of cost-effectiveness outcomes between currently available and hypothetical diagnostic algorithms over a 60-year time horizon.

Perspective	Algorithm	SE_1 / SE_2 / SP (%)*	Total costs (€)	Total QALYs	iCosts (€)†	iQALY†	ICUR (€/QALY)	iNMB†,‡
Existing algorithms								
Societal - friction cost	Modification 2	90 / 76 / 85	157,469	24.23	Ref.	Ref.	Ref.	Ref.
	Modification 1	84 / 69 / 82	158,417	24.15	948	−0.08	Dominated by M2	−2569
	Berlin	52 / 50 / 88	159,423	23.96	1954	−0.27	Dominated by M2	−7412
Societal - human capital	Modification 2	90 / 76 / 85	334,250	24.23	Ref.	Ref.	Ref.	Ref.
	Modification 1	84 / 69 / 82	335,436	24.15	1186	−0.08	Dominated by M2	−2807
	Berlin	52 / 50 / 88	336,927	23.96	2678	−0.27	Dominated by M2	−8136
Healthcare	Berlin	52 / 50 / 88	91,883	23.96	Ref.	Ref.	Ref.	Ref.
	Modification 2	90 / 76 / 85	92,837	24.23	954	0.27	3495	4504
	Modification 1	84 / 69 / 82	92,890	24.15	53	−0.08	Dominated by M2	−1674
Perfect diagnosis								
Societal - friction cost	Modification 2	90 / 76 / 85	157,469	24.23	Ref.	Ref.	Ref.	Ref.
	Perfect	100 / 100 / 100	154,247	24.44	−3222	0.21	Dominates M2	7467
Societal - human capital	Modficiation 2	90 / 76 / 85	334,250	24.23	Ref.	Ref.	Ref.	Ref.
	Perfect	100 / 100 / 100	330,348	24.44	−3901	0.21	Dominates M2	8146
Healthcare	Modification 2	90 / 76 / 85	92,837	24.23	Ref.	Ref.	Ref.	Ref.
	Perfect	100 / 100 / 100	92,129	24.44	−709	0.21	Dominates M2	4953

*SE_1 = sensitivity for r-axSpA, SE_2 = sensitivity for nr-axSpA, SP = specificity.
† Calculated between adjacent non-dominated algorithms.
‡ Calculated using a willingness-to-pay threshold of 20,000€/QALY.
iCost, incremental cost; ICUR, incremental cost-utility ratio; iNMB, incremental net monetary benefit; iQALY, incremental QALY; M2, Modification 2 algorithm; QALY, quality-adjusted life-year; Ref, reference (comparator); SE, sensitivity; SP, specificity.

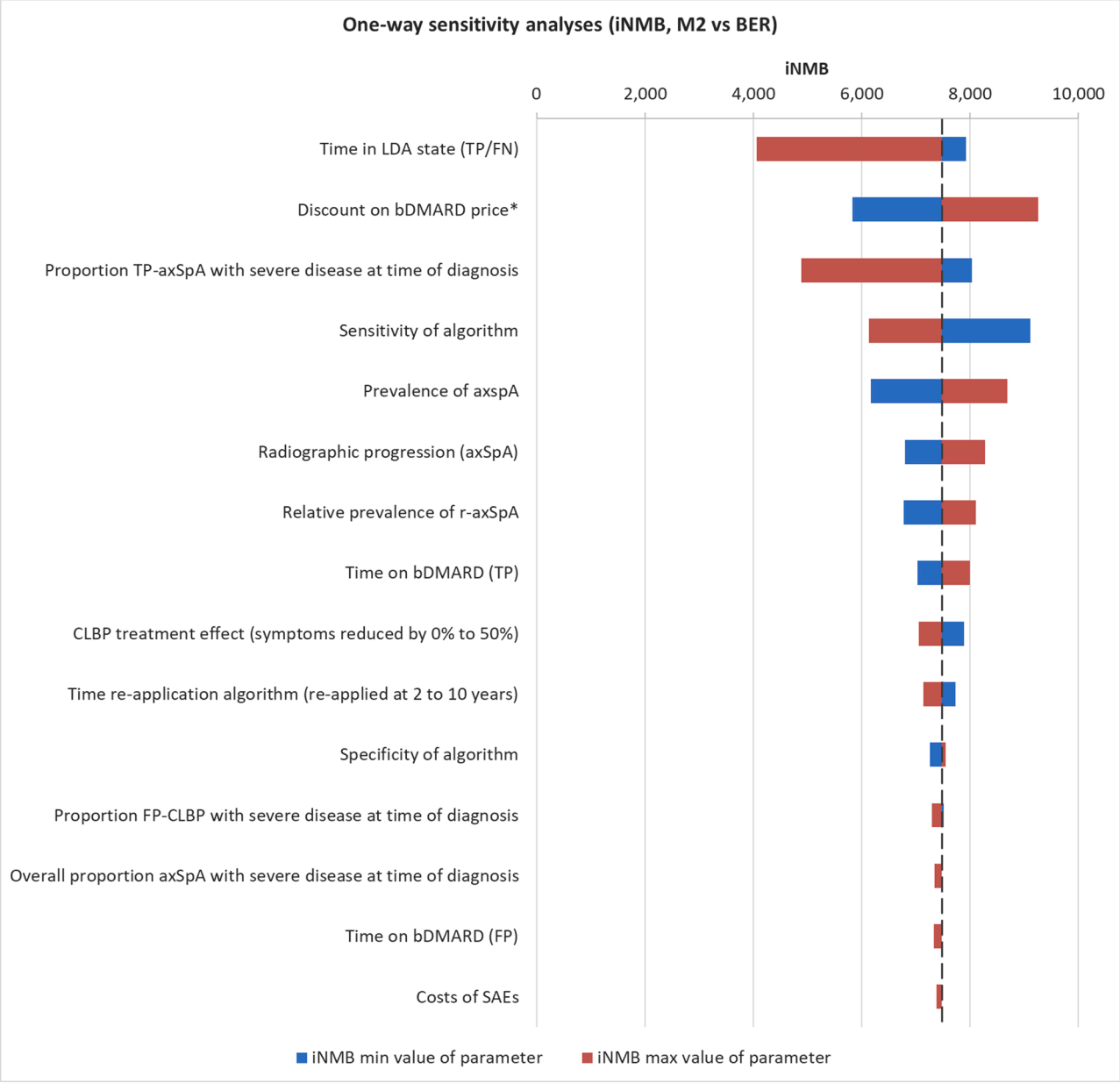

Fig. 4. Results of deterministic one-way sensitivity analysis of M2 versus BER algorithm, societal perspective
Results of deterministic one-way sensitivity analysis of the M2 algorithm (intervention) compared to the original BER algorithm (comparator), using a societal perspective with a friction costs approach over a 60-year time horizon. iNMB was calculated using a willingness-to-pay threshold of 20,000€/QALY. Minimum and maximum values of the parameters reflect the bounds of the 95% confidence interval of that parameter, unless otherwise stated. The dashed line (separating the coloured bars) represents the iNMB of the base-case analysis of the M2 algorithm compared to the BER algorithm.
*Percentage range of discounts on bDMARD prices is confidential data and has been removed from figure.
axSpA, axial spondyloarthritis; BER, Berlin algorithm; bDMARD, biological disease-modifying antirheumatic drug; CLBP, chronic low back pain; FN, false negative; FP, false positive; iNMB, incremental net monetary benefit; LDA, low disease activity; M2, Modification 2; r-axSpA, radiographic axial spondyloarthritis; SAE, serious adverse event; TP, true positive.

several available diagnostic algorithms for axSpA in the Dutch setting, the M2 algorithm (in which presence of IBP is not obligatory for diagnosis) provided the most health and lowest costs. Previously already recommended by ASAS, these findings further support M2 as the diagnostic algorithm of choice. Importantly, a considerably more expensive diagnostic algorithm with better accuracy than M2 could still be cost-effective. Approaches to such an algorithm could be aimed at increasing either sensitivity or specificity, although the net gains by increasing sensitivity could be slightly larger.

To our knowledge, this is the first study to estimate the health and economic consequences of (in)correct diagnosis of axSpA. One previous study did investigate the cost-effectiveness of utilizing conventional radiography of the sacroiliac joints compared to MRI, or both, for the diagnosis of axSpA from a US healthcare perspective. [42] However, this study differed substantially in scope and methodology.

The current study provides an additional and relevant perspective on diagnosis of axSpA. Traditional research of diagnostics does not fully take the consequences of diagnosis into account. In health-economic studies such as ours, the consequences of both diagnosis and misdiagnosis can be estimated. We demonstrated that incorrect diagnosis (overdiagnosis or underdiagnosis) is associated with substantial loss of health for the individual patient, and economic burden for society.

However, we also observed that if an initially incorrect diagnosis is corrected within a few years, the long-term consequences are limited.

The differences in costs and effects between the algorithms might be considered small to modest. This can be explained by the trade-off between sensitivity and specificity: with M2 being more sensitive than BER (more TP, less FN cases), there is less underdiagnosis. However, as M2 is also less specific (more FP, less TN cases), there is more overdiagnosis and overtreatment of non-axSpA cases. Apparently, when comparing these algorithms, these opposing 'forces' are relatively well-balanced in terms of health and costs. Another consideration is that, despite the small differences in costs, if these diagnostic algorithms would be used in all referred patients with suspected axSpA, the between-algorithm differences in budget impact would be substantial.

Our data indicate that further improvement of diagnosis would still be worthwhile, even if this would be more costly. The debate on diagnosing axSpA seems to mainly focus on overdiagnosis, which is understandable. With the increasing focus on early diagnosis, the risk of overdiagnosis (and overtreatment) is real. However, our results indicated that underdiagnosis might be an equally or more important aspect, as improvement of sensitivity resulted in at least equal gains as improving specificity. The finding that a perfect diagnostic test for axSpA could cost €7500 per referred patient is promising for further research on new diagnostic biomarkers and tests. To put this number into context: the current diagnostic work-up for axSpA has a cost of approximately €250–300. amongst the parameters most influential on cost-effectiveness in the current analysis were the prevalence of axSpA and disease severity of referred patients. The prevalence of axSpA affects the relative importance of sensitivity and specificity. The disease severity affects the potential gains in health and reductions in costs when improving diagnostic accuracy. If the axSpA prevalence or disease severity differs in other settings, results will also differ. Costs of bDMARDs also had notable impact on results. Even at reduced cost, bDMARD costs remained an important driver of costs in management of axSpA. Future cost reduction would greatly improve the value of diagnosis and allow for more costly diagnostic strategies.

In the diagnostic algorithms for axSpA, the conventional radiography of the sacroiliac joints (X-SI) is positioned quite high. For M2, it is actually the first step after entry. The role of the X-SI in the diagnostic work-up for axSpA has been debated, and it is possible that this might change in the future. However, as far as we are aware, the algorithms used in this study are the only ones currently available, and they are being used in current practice. Also, M2 is endorsed by ASAS. It was not our goal to change these algorithms, but to investigate the consequences of using them in terms of health and costs. Future studies should assess the impact of changes to these algorithms (such as a different positioning of X-SI).

Our study was focused on accuracy of diagnosis of axSpA by the rheumatologist. In practice, diagnosis of axSpA by the rheumatologist also depends on referral of suspected cases by non-rheumatologists (e.g. general practitioners). One relevant study in this area is a health-economic analysis of referral strategies for axSpA in primary care in the United Kingdom (UK), which showed there was considerable uncertainty as to which strategy would be optimal compared to current UK practice. [43] However, in this study it was assumed that the rheumatologist's diagnosis of axSpA/non-axSpA in the referred population was perfect. In reality, improving referral of suspected axSpA will likely change the composition and case-mix of the referred population that is seen by the rheumatologist, and this might affect the performance of the rheumatologist and of the diagnostic algorithms investigated in this paper. The interplay between referral and diagnosis of axSpA is very relevant, but has yet to be investigated. This requires studies that take into account both referral and diagnosis.

The current study has several limitations. First, we identified several areas of uncertainty, and had to make assumptions for these. Relevant assumptions were those regarding healthcare utilisation by axSpA and CLBP cases, and the occurrence of long-term absenteeism in axSpA and CLBP (we only had short-term patient-level data for this). Second, as long-term data on EQ-5D is lacking in axSpA, we used a mapping algorithm to estimate EQ-5D utility. This algorithm was based on BASDAI and BASFI, which were well studied in longitudinal studies of axSpA. Third, we only re-applied the algorithm at a single point in time, due to technical limitations of the model. In practice, diagnosis is not a static phenomenon, but can change over time. This is why we conducted sensitivity analyses assuming re-consideration after a rather short time (2 years), which confirmed the results. Finally, the external (real-life) validity of our results depends on complete adherence to guidelines for diagnosis and management in daily practice, which is not always the case. [44] Although this is not a limitation of the model or this study, it can affect the results in practice.

Strengths of this work include the conceptualization, development and validation of the model, that were based on International Society for Pharmacoeconomics and Outcomes Research guidelines. [45,46] Demographic and diagnostic parameters were derived from the cohort in which the algorithms were developed and modified (SPACE). We used local and external experts to 'validate' the model structure and input parameters. Finally, we opted for a modelling technique (state-transition model) that is relatively easy to use and interpret by others, and overcame the main limitations of this type of model by utilizing several technical work-arounds (see Supplementary File S1).

The generalizability (transferability) of an economic evaluation is often limited due to regional differences in health systems, unit costs and social security regulations. The current analysis was conducted from the Dutch perspective. Our model structure, including management of axSpA and CLBP, and the modelled diagnostics and effects are likely transferable to other regions. The modelled costs, especially those due to work productivity losses, are likely to differ by country.

Going forward, this model can be used to evaluate potential diagnostic biomarkers for axSpA. In addition, the model could be extended to incorporate referrals from general practitioners and non-rheumatology specialists (as discussed above). This is especially important because prevalence of axSpA, which is directly linked to patient referral, was an important driver of results.

In conclusion, correct diagnosis of axSpA is associated with substantial health benefits for the patient and cost savings for society. The M2 algorithm (that does not require IBP for axSpA diagnosis) is acceptable in terms of costs and effects when compared to the original Berlin algorithm, although differences between algorithms are modest. Furthermore, a perfect diagnostic test could cost substantially more (€7500 per referred patient) but still provide good value for money. It is worthwhile to invest in more accurate diagnosis in axSpA.

Funding

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

CRediT authorship contribution statement

Casper Webers: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft. **Sabine Grimm:** Methodology, Software, Validation, Investigation, Writing – review & editing. **Astrid van Tubergen:** Investigation, Writing – review & editing. **Floris van Gaalen:** Investigation, Writing – review & editing. **Désirée van der Heijde:** Investigation, Writing – review & editing. **Manuela Joore:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision. **Annelies Boonen:** Conceptualization, Methodology, Investigation, Writing – original draft, Supervision.

Declaration of Competing Interest

CW has nothing to disclose. SG has nothing to disclose. AvT reports grants from Pfizer, UCB; grants and consulting fees from Novartis; consulting fees from Galapagos; outside the submitted work and paid to the institution. FvG reports grants from Pfizer, Reuma Nederland, Stichting Vrienden van Sole Mio, Assessment of SpondyloArthritis international Society (ASAS); consulting fees from MSD, Novartis, UCB, Eli Lilly, AbbVie, BMS; and is member of ASAS Executive Committee and ASAS treasurer (unpaid); outside the submitted work. DvdH reports personal fees from AbbVie, Bayer, BMS, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Lilly, Novartis, Pfizer, Takeda, UCB Pharma, outside the submitted work; and is Associate editor of Annals of the Rheumatic Diseases, Editorial board member of Journal of Rheumatology and RMD Open, advisor of ASAS and Director of Imaging Rheumatology bv. MJ has nothing to disclose. AB reports grants from Abbvie; consulting fees or honoraria from UCB, Galapagos, Abbvie, Pfizer, Novartis; outside the submitted work and paid to the institution.

Ethics approval and consent to participate

Not applicable.

Data availability

All data relevant to this study are published in the article or in the supplementary files. The model developed for this study is available for collaborative purposes, upon reasonable request. Proposals should be directed to the corresponding author.

Acknowledgements

The authors would like to thank Miranda van Lunteren (Rheumatology, Leiden University Medical Centre, Leiden, the Netherlands) for her assistance with data management of SPACE, Stephen Palmer (Centre for Health Economics, University of York, York, United Kingdom) for consultation on the York model and providing the research team access to the York model [47] and Fariz Yahya (Royal National Hospital for Rheumatic Diseases, Bath, United Kingdom and University of Malaya, Kuala Lumpur, Malaysia) for providing additional data on bDMARD drug survival in the BRITSpA study. [48] Some of the supplementary figures were inspired by the work of McAllister et al. (NICE Guideline 65, [43] in particular Appendix H: Full health economics report). Finally, the authors would like to thank prof. Robert Landewé for his helpful comments on a previous version of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2023.152242.

References

[1] Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017;390:73–84.
[2] Rudwaleit M, van der Heijde D, Landewe R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777–83.
[3] Ramiro S, Nikiphorou E, Sepriano A, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82:19–34.
[4] Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535–43.
[5] van den Berg R, de Hooge M, Rudwaleit M, et al. ASAS modification of the Berlin algorithm for diagnosing axial spondyloarthritis: results from the SPondyloArthritis caught early (SPACE)-cohort and from the assessment of spondyloarthritis international society (ASAS)-cohort. Ann Rheum Dis 2013;72:1646–53.
[6] Calin A, Porta J, Fries JF, et al. Clinical history as a screening test for ankylosing spondylitis. JAMA 1977;237:2613–4.
[7] Rojas-Vargas M, Munoz-Gomariz E, Escudero A, et al. First signs and symptoms of spondyloarthritis–data from an inception cohort with a disease course of two years or less (REGISPONSER-Early). Rheumatology (Oxford) 2009;48:404–9.
[8] Poddubnyy D, Callhoff J, Spiller I, et al. Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care. RMD Open 2018;4:e000825.
[9] Rudwaleit M, Landewe R, van der Heijde D, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis 2009;68:770–6.
[10] Sieper J, van der Heijde D, Landewe R, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the assessment of spondyloarthritis international society (ASAS). Ann Rheum Dis 2009;68:784–8.
[11] Drummond M, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015. Fourth edition. editor.
[12] Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998;13:397–409.
[13] Machado P, Landewe R, Braun J, et al. A stratified model for health outcomes in ankylosing spondylitis. Ann Rheum Dis 2011;70:1758–64.
[14] Tran-Duy A, Boonen A, van de Laar MA, et al. A discrete event modelling framework for simulation of long-term outcomes of sequential treatment strategies for ankylosing spondylitis. Ann Rheum Dis 2011;70:2111–8.
[15] Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. Am J Epidemiol 2006;163:754–61.
[16] Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. BMJ Open 2013;3:e003838.
[17] Chen Y, Campbell P, Strauss VY, et al. Trajectories and predictors of the long-term course of low back pain: cohort study with 5-year follow-up. Pain 2018;159:252–60.
[18] Kongsted A, Kent P, Axen I, et al. What have we learned from ten years of trajectory research in low back pain? BMC Musculoskelet Disord 2016;17:220.
[19] Salomon JA, Haagsma JA, Davis A, et al. Disability weights for the global burden of disease 2013 study. Lancet Glob Health 2015;3:e712–23.
[20] Boel A, Molto A, van der Heijde D, et al. Do patients with axial spondyloarthritis with radiographic sacroiliitis fulfil both the modified New York criteria and the ASAS axial spondyloarthritis criteria? Results from eight cohorts. Ann Rheum Dis 2019;78:1545–9.
[21] Ramiro S, van der Heijde D, Sepriano A, et al. Spinal radiographic progression in early axial spondyloarthritis: five-year results From the DESIR cohort. Arthritis Care Res (Hoboken) 2019;71:1678–84.
[22] Ornbjerg LM, Brahe CH, Askling J, et al. Treatment response and drug retention rates in 24 195 biologic-naive patients with axial spondyloarthritis initiating TNFi treatment: routine care data from 12 registries in the EuroSpA collaboration. Ann Rheum Dis 2019;78:1536–44.
[23] Smeets RJ, Severens JL, Beelen S, et al. More is not always better: cost-effectiveness analysis of combined, single behavioral and single physical rehabilitation programs for chronic low back pain. Eur J Pain 2009;13:71–81.
[24] Goossens ME, de Kinderen RJ, Leeuw M, et al. Is exposure in vivo cost-effective for chronic low back pain? A trial-based economic evaluation. BMC Health Serv Res 2015;15:549.
[25] Lamb SE, Hansen Z, Lall R, et al. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. Lancet 2010;375:916–23.
[26] Artus M, van der Windt D, Jordan KP, et al. The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. BMC Musculoskelet Disord 2014;15:68.
[27] Boonen A, van den Heuvel R, van Tubergen A, et al. Large differences in cost of illness and wellbeing between patients with fibromyalgia, chronic low back pain, or ankylosing spondylitis. Ann Rheum Dis 2005;64:396–402.
[28] Ekman M, Jonhagen S, Hunsche E, et al. Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting. Spine (Phila Pa 1976) 2005;30:1777–85.
[29] Webers C, Vanhoof L, van Genderen S, et al. Employment and the role of personal factors among patients with ankylosing spondylitis: a Dutch cross-sectional case-control study. RMD Open 2018;4:e000680.
[30] Castillo-Ortiz JD, Ramiro S, Landewe R, et al. Work outcome in patients with Ankylosing Spondylitis: results from a 12-year followup of an international study. Arthritis Care Res (Hoboken) 2016;68:544–52.
[31] Shmagel A, Foley R, Ibrahim H. Epidemiology of chronic low back pain in US adults: data from the 2009-2010 National Health and Nutrition Examination survey. Arthritis Care Res (Hoboken) 2016;68:1688–94.
[32] van den Hout WB. The value of productivity: human-capital versus friction-cost method. Ann Rheum Dis 2010;69:i89–91.
[33] Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg [Guideline for economic evaluations in healthcare]. Diemen: Zorginstituut Nederland [National Health Care Institute of the Netherlands]; 2015.
[34] Wailoo A, Hernandez M, Philips C, et al. Modeling health state utility values in Ankylosing Spondylitis: comparisons of direct and indirect methods. Value Health 2015;18:425–31.
[35] Farkouh ME, Kirshner H, Harrington RA, et al. Comparison of lumiracoxib with naproxen and ibuprofen in the therapeutic arthritis research and gastrointestinal event trial (TARGET), cardiovascular outcomes: randomised controlled trial. Lancet 2004;364:675–84.

[36] Silverstein FE, Faich G, Goldstein JL, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib Long-term Arthritis Safety study. JAMA 2000;284:1247–55.
[37] Cannon CP, Curtis SP, FitzGerald GA, et al. Cardiovascular outcomes with etoricoxib and diclofenac in patients with osteoarthritis and rheumatoid arthritis in the multinational etoricoxib and diclofenac arthritis long-term (MEDAL) programme: a randomised comparison. Lancet 2006;368:1771–81.
[38] Sepriano A, Regel A, van der Heijde D, et al. Efficacy and safety of biological and targeted-synthetic DMARDs: a systematic literature review informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. RMD Open 2017;3:e000396.
[39] Wallis D, Thavaneswaran A, Haroon N, et al. Tumour necrosis factor inhibitor therapy and infection risk in axial spondyloarthritis: results from a longitudinal observational cohort. Rheumatology (Oxford) 2015;54:152–6.
[40] Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database Syst Rev 2011;2:CD008794.
[41] Vijgen S, Van Heesch F, Obradovic M. Ziektelast in de praktijk: de theorie en praktijk van het berekenen van ziektelast bij pakketbeoordelingen [Disease burden in practice]. Diemen: Zorginstituut Nederland; 2018.
[42] Gorelik N, Tamizuddin F, Rodrigues TC, et al. Comparison between radiography and magnetic resonance imaging for the detection of sacroiliitis in the initial diagnosis of axial spondyloarthritis: a cost-effectiveness study. Skeletal Radiol 2020;49:1581–8.
[43] Spondyloarthritis in over 16s: diagnosis and management. London: national institute for health and care excellence; 2017.
[44] Rees F, Peffers G, Bell C, et al. Compliance with NICE guidance on the use of anti-TNF alpha agents in ankylosing spondylitis: an east and west Midlands regional audit. Clin Med (Lond) 2012;12:324–7.
[45] Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. Value Heal 2012;15:804–11.
[46] Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force–7. Value Health 2012;15:843–50.
[47] Corbett M, Soares M, Jhuti G, et al. Tumour necrosis factor-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. Health Technol Assess 2016;20:1–334. v-vi.
[48] Yahya F, Gaffney K, Hamilton L, et al. Tumour necrosis factor inhibitor survival and predictors of response in axial spondyloarthritis-findings from a United Kingdom cohort. Rheumatology (Oxford) 2018;57:619–24.
[49] Kanters TA, Bouwmans CAM, van der Linden N, et al. Update of the Dutch manual for costing studies in health care. PLoS ONE 2017;12:e0187477.