Appendix H: Full health economics report

H.1 Original health economic analysis – decision problem

The GDG identified the recognition of axial spondyloarthritis as its key priority for original health economic analysis. The group advised that delayed diagnosis is a significant issue in all spondyloarthritis, but that people with axial symptoms are subject to particularly damaging delays, invariably because their symptoms are misidentified as mechanical back pain. The GDG emphasised that, if people with axial disease could be identified more reliably when they first present, they would gain access to effective treatments, improving their quality of life and their chances of long-term treatment response.

Table 1: Research questions

RQ	What are the indications (signs, risk factors, test or scan findings) for referral for specialist advice at initial diagnosis in suspected axial disease?

There is a historical delay to diagnosis of people with spondyloarthritis, particularly axial disease and therefore the Committee were interested in economic modelling that could provide analysis to help combat this problem.

Table 2: PICO

Population	People presenting with low back pain with a duration of more than 3 months and with an onset before the age of 45.
Intervention	Criteria on which to refer to specialist care
Comparator	Current practice/ alternative referral criteria
Outcomes	A cost–utility analysis was constructed based on the quality of life (in quality adjusted life years[QALYs]) and costs of different strategies in terms of their ability to identify spondyloarthritis.

H.1.1 Systematic review of published cost–utility analyses

A literature search was conducted jointly for all review questions in this guideline by applying standard health economic filters to a clinical search for spondyloarthritis (see Appendix C). A total of 9,970 references was retrieved, of which none were retained for these review questions.

H.2 Original cost–utility model – methods

H.2.1 Overview of the model

H.2.1.1 Modelled population(s) and intervention(s)

The population model reflects that outlined above (Table 2).

The model uses a patient perspective for outcomes and an NHS perspective for costs, in line with the Guidelines Manual (2012).

H.2.1.2 Model structure

We built a Markov model with a 3-month cycle length, and a lifetime time horizon to estimate the costs and utility of being correctly identified as having spondyloarthritis, and having the disease missed and therefore being treated for chronic low back pain of unknown cause. The

model is run for a lifetime horizon which therefore provided estimates of the quality of life and costs (over a lifetime) of disease depending on whether people get the correct diagnosis.

The model has a decision-tree element which proportionally allocates the cohort to the outcomes following a diagnostic test (true positive, false negative, false positive and true negative). The costs and utility of each of the referral criteria then become a summation of the proportion of the cohort multiplied by estimated costs and utilities for each of the diagnostic outcomes.

We assume that people who are correctly identified as true negatives incur no further costs beyond those incurred in primary care as part of the initial referral strategy, nor have a quality of life impact.

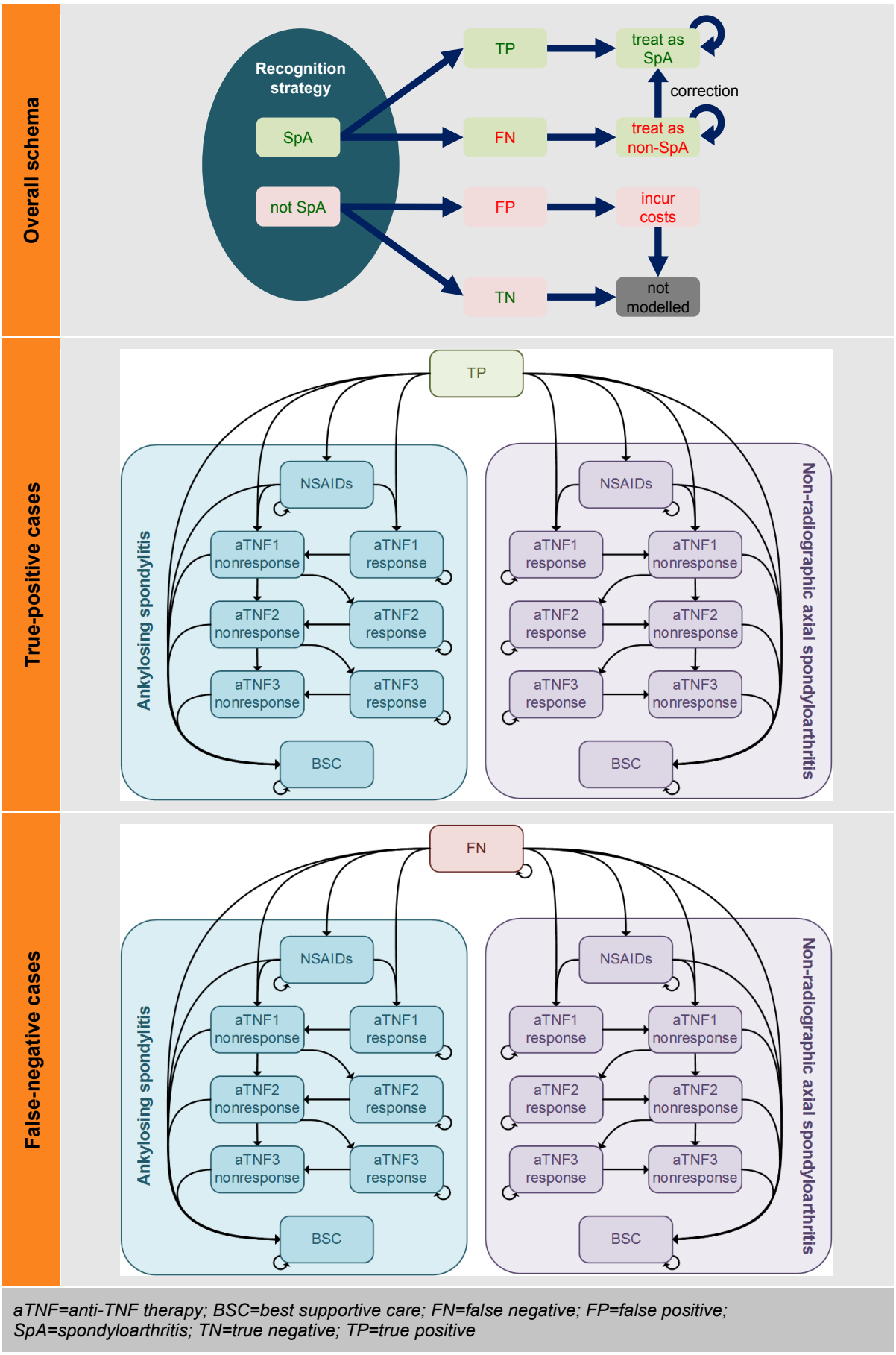
The false positives incur the costs of referral and the cost of a diagnostic work-up in specialist care. We assume in this model that the specialists correctly identify that these people do not have spondyloarthritis and therefore at this point they stop accruing costs altogether.

Accordingly, we designed the original model to estimate quality of life and costs (over a lifetime) of people who are and are not correctly referred, having presented with symptoms that might indicate axial spondyloarthritis. It has a 3-month cycle length and a lifetime time horizon, and adopts a patient perspective for outcomes and an NHS perspective for costs, in line with the Guidelines Manual (2012).

In reflection of the diagnostic accuracy evidence, the simulated population comprises people with chronic back pain of at least 3 months' duration that began at age 45 or younger. Using data from a large inception study (Rudwaleit et al., 2009), the ankylosing spondylitis (AS) cohort was assumed to be 64% male with an average age of 30.4 (95%CI: 29.0 to 31.8), and the non-radiographic axial spondyloarthritis (nrAxSpA) cohort was 43% male and had a mean age of 33.2 (95%CI: 31.8 to 34.6).

Figure 1 provides a schematic depiction of the model structure. It shows that we model each recognition strategy in terms of its ability to categorise people into true-positive and true-negative diagnoses (with complementary probabilities of false-negative and false-positive diagnoses, respectively). The long-term costs and QALYs associated with people who do not have spondyloarthritis are not modelled: it is assumed that the specialists to whom false-positive cases are incorrectly referred will identify their true-negative status, so only the costs of specialist diagnostic work-up are modelled. Where true-negative cases are concerned, the choice of referral strategy makes no difference to the future costs and quality of life of people who are correctly identified as not having spondyloarthritis, so there is no need to estimate these.

A simplified treatment pathway is assumed for true-positive cases: for most people, first-line treatment is with NSAIDs (although a proportion of people will be contraindicated and proceed directly to biological DMARDs, unless they are also contraindicated for these, in which case they can only receive best supportive care [BSC]). Up to 3 lines of anti-TNF therapy are modelled, in reflection of technology appraisal guidance TA383. The BSC state is designed to represent the care of people who cannot take – or whose disease no longer responds to – any disease-modifying therapy. The model assumes that a proportion of people in this state are referred to a chronic pain management service.



aTNF=anti-TNF therapy; BSC=best supportive care; FN=false negative; FP=false positive; SpA=spondyloarthritis; TN=true negative; TP=true positive

Figure 1: Structure of original cost–utility model

The false-negative pathway is identical to the true-positive version, with the critical exception that people remain in the false-negative state (where they are treated as if they have mechanical low-back pain) until their true diagnosis is uncovered. The likelihood of late diagnosis is parameterised used evidence from a survey of 1,630 people with ankylosing spondylitis in the UK (NASS 2013).

The model allocates cases of spondyloarthritis proportionally between diagnoses of ankylosing spondylitis (AS) and non-radiographic axial spondyloarthritis (nrAxSpA); wherever evidence exists for differential effects between these categories, this is reflected in the model. There are no transitions between the AS and nrAxSpA subgroups. Although it is acknowledged that some people are first diagnosed with non-radiographic disease that subsequently becomes radiographically overt, this is also true of participants in the studies used to populate the nrAxSpA pathway (most notably, RCTs of the effectiveness of biological DMARDs). Therefore, the 'non-radiographic' states in the model can be interpreted as 'axial spondyloarthritis that did not meet radiographic criteria at the time of initial diagnosis'. The GDG advised that, in practice, no difference would be expected between people presenting with symptoms that turn out to be ankylosing spondylitis and people who ultimately receive a diagnosis of non-radiographic axial disease. Therefore, there is no reason to suspect that different referral criteria and different pathways should be considered for people with different presenting symptoms.

Discounting of costs and effects was at 3.5%.

The 3-monthly cycle represents an appropriate time period as the disease, although progressive, is often gradual in the worsening of symptoms (with the exception of flare). Therefore we would not expect large and significant changes in health to occur between cycles. The evidence of treatment effects is measured at 3 months for biological therapies and therefore a 3-monthly cycle is also convenient for this purpose.

H.2.2 Prevalence of axial spondyloarthritis

A critical parameter in this model is the proportion of people presenting with ≥3 months' back pain that began at age ≤45 who truly have axial spondyloarthritis. No ideal source was found for this parameter. Hamilton et al. (2015) found that 1.8% of people of all ages with back pain meet ASAS criteria for axial spondyloarthritis; however, their study is not confined to the presenting population of interest and includes some people with a known diagnosis of ankylosing spondylitis. Van Hoeven et al. (2015) found a much higher prevalence of 16.4% in their study identifying people aged 18–45 with back pain in a GP database. The wide range of these estimates indicates the degree of uncertainty that exists around the true value of the parameter. In discussion with the GDG, a base case value of 5% was agreed as reasonable; this was thought to be a relatively conservative estimate.

H.2.3 Projecting BASDAI and BASFI

The average BASDAI and BASFI of simulated patients in each state is projected using evidence on natural history and treatment effect. These are used to project quality of life (using a published mapping function [Wailoo et al., 2015]) and background healthcare costs (data from Boonen et al., 2003, as implemented by Corbett et al., 2016). Because BASDAI and BASFI are projected to rise at a steeper trajectory in occult disease than when people are receiving appropriate treatment, people who are diagnosed later have higher values which, in turn, translate into worse quality of life and higher background healthcare costs.

Models that seek to quantify the benefits of prompt detection of a condition must estimate the harms of delayed detection, and unavoidably face the problem that the natural history of an occult condition is, by definition, unknowable. In this model, we had a particular need to estimate the progression of BASDAI and BASFI over time in people whose spondyloarthritis is undiagnosed. We did this by using evidence from Aggarwal and Malaviya (2009). The

authors dichotomise their population at the median duration of pre-diagnosis symptoms into those with an 'early' diagnosis of ankylosing spondylitis (mean diagnostic delay of 3 years) and those with a 'late' diagnosis (mean diagnostic delay of 10.5 years). This evidence shows significantly higher BASDAI and BASFI in the late diagnoses than in the earlier, from which a year-on-year increase in values in occult disease can be inferred.

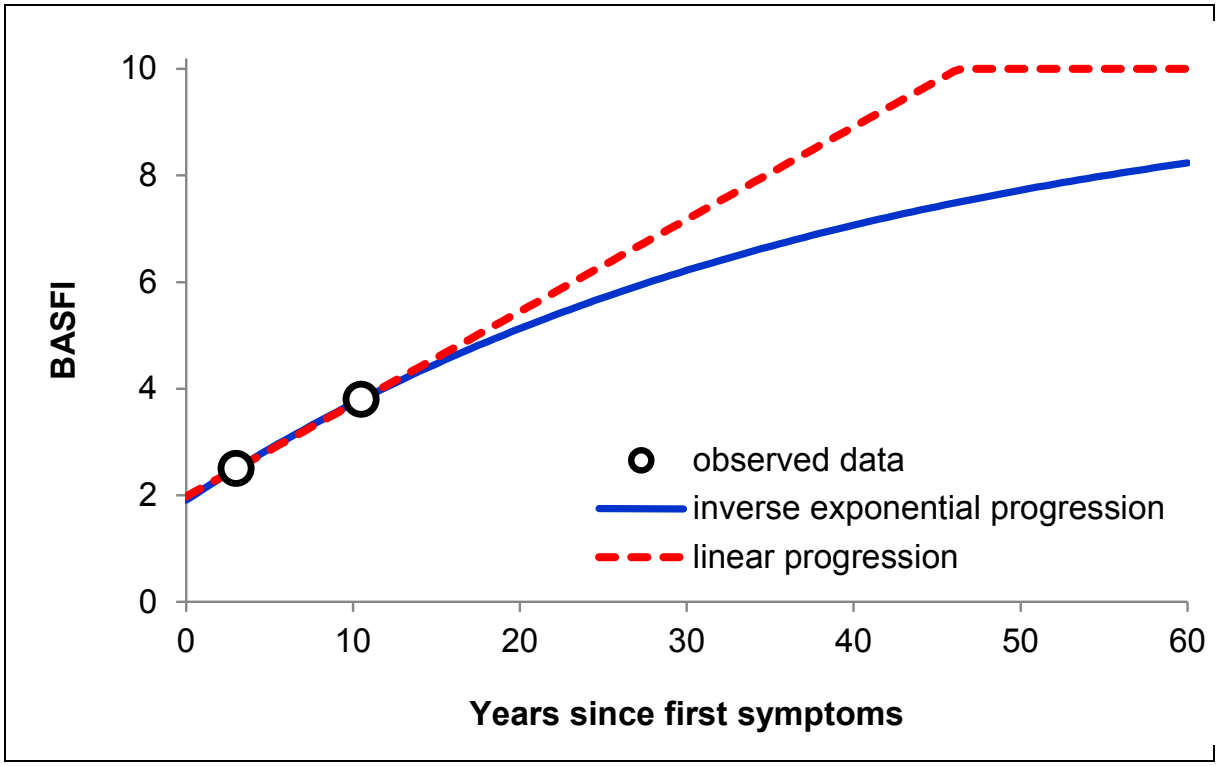


Figure 2: Illustrative natural history of BASFI in occult disease, using inverse exponential extrapolation

However, it is necessary to extrapolate this trend over an extended timeframe, and a simple linear increasing trend would, over a lifetime model, ultimately exceed the bounds of the scale (that is, result in BASFIs and BASDAIs greater than 10). In discussion with the GDG, it was agreed that the best solution to this problem would be to use a constrained function – an inverse exponential function – that (a) reflects a slowly diminishing level of year-on-year progression over time, and (b) can be guaranteed to keep values in a sensible realm. Inverse exponential growth has these properties (in other words, an exponential decline that is inverted and then scaled such that its characteristic of never declining to 0 is translated into a property of never increasing to 10). The approach is illustrated in Figure 2.

Using this approach, the natural history of BASFI at time *t* is estimated as

$$BASFI_t = 10 - (10 - BASFI_{T0}) \left(\frac{10 - BASFI_{TNF}}{10 - BASFI_{T0}} \right)^{\frac{t}{t_{symp}}} \quad (1)$$

, where $BASFI_{T0}$ represents BASFI at the point of first consultation, $BASFI_{TNF}$ represents average BASFI at the start of anti-TNF therapy, and t_{symp} is total time with symptoms before commencing anti-TNFs.

$BASFI_{T0}$ is not a trivial quantity to estimate, because it relates to the true BASFI at the earliest moment an individual could have been diagnosed, not their BASFI at the moment of diagnosis. However, it can be approximated by 'winding back' the natural history from a

known BASFI associated with a known duration of symptoms; in this case, we use BASFI and duration of symptoms reported at commencement of anti-TNF therapy.

In this way, $BASFI_{T0}$ can be estimated as a function of known quantities – BASFI at diagnosis ($BASFI_{diag}$), t_{symp}, time with symptoms before initial presentation (t_{preNHS} – from survey data [NASS 2013]), time between diagnosis and beginning of anti-TNFs (approximated by time on NSAIDs before starting anti-TNFs – t_{NSAIDs}) – and our knowledge about progression of occult disease from Aggarwal and Malaviya (2009):

$$BASFI_{T0} = \left(10 - BASFI_{diag}\right)\left(\frac{10 - BASFI_{late}}{10 - BASFI_{early}}\right)^{\frac{-t_{symp}-t_{preNHS}-t_{NSAIDs}}{t_{late}-t_{early}}} \quad \textbf{(2)}$$

$$= (10 - 4.13)\left(\frac{10 - 3.8}{10 - 2.5}\right)^{\frac{-13.66-3.46-6.64}{10.5-3}}$$

$$= 3.575$$

An identical formulation is used to calculate BASDAI profiles.

H.2.4 Effectiveness of treatments for axial spondyloarthritis

For true-positive diagnoses of axial spondyloarthritis, the model simulates first-line NSAIDs followed by up to 3 lines of anti-TNF therapy. These have effects on the cohort's BASDAI and BASFI and, in the case of anti-TNFs, response and non-response is explicitly modelled.

H.2.4.1 NSAIDs

The effectiveness of NSAIDs, in terms of BASFI and BASDAI reduction, is drawn from an NMA of data from the Cochrane review of NSAIDs for axial SpA (Kroon et al. 2015). Although Kroon et al. (2015) include 35 RCTs in their overall review, only 3 report BASDAI and/or BASFI data for the relevant comparisons; all are exclusively in people with ankylosing spondylitis (Dougados et al., 2001; Sieper et al., 2008; van der Heijde et al., 2005).

In our NMAs, 2 classes of NSAIDs – COX-2 inhibitors and traditional NSAIDs – are compared with each other and placebo in a 3-node analysis. The effect of NSAIDs as an overall class is then calculated as a weighted average of results for COX-2 and traditional NSAIDs, according to the assumed proportion of people taking each – for the base case, the GDG advised that around 30% of people take COX-2 inhibitors.

For both BASDAI and BASFI, random-effects models provided a superior fit to the data (as measured by DIC), so these were preferred.

Table 3: BASFI – relative effectiveness of all pairwise combinations (row -v- column)

Placebo		
-1.327 (-2.005, -0.657)	**COX-2 NSAID**	
-1.012 (-1.662, -0.313)	0.315 (-0.221, 0.909)	**Traditional NSAID**

Values given are mean differences (row versus column). The point estimate reflects the median of the posterior distribution, and numbers in parentheses are 95% credible intervals.

Table 4: BASDAI – relative effectiveness of all pairwise combinations (row -v- column)

Placebo		
-2.031 (-3.274, -0.814)	**COX-2 NSAID**	
-1.899 (-3.120, -0.645)	0.131 (-0.750, 1.096)	**Traditional NSAID**

Values given are mean differences (row versus column). The point estimate reflects the median of the posterior distribution, and numbers in parentheses are 95% credible intervals.

For consistency with complexities of modelling anti-TNFs (see below), the effect of NSAIDs is applied in a proportional way in the base case. To enable this, the average BASFI and BASDAI at baseline of participants in the meta-analysed RCTs was calculated, and the difference expressed proportionally. For example, average baseline BASFI was 4.560 and weighted average mean difference of NSAIDs -v- placebo is −1.110, so NSAIDs are estimated to reduce BASFI to 75.7% of baseline level: (4.560−1.110)/4.560=75.7%. The impact of applying changes in an absolute manner was explored in sensitivity analysis.

Duration of NSAID therapy

The average duration of NSAID monotherapy is estimated using the duration of diagnosed disease at baseline in RCTs of first-line anti-TNF therapy. The GDG advised that this should provide a good approximation of time on NSAIDs, because people will have invariably commenced routine NSAID therapy at diagnosis, and would only become eligible for anti-TNF trials when that proved inadequate to control their symptoms. All RCTs reporting this value were pooled in random-effects meta-analyses, which arrived at average durations of 6.6 years for people with ankylosing spondylitis and 2.9 years for people with non-radiographic axial spondyloarthritis. In the absence of more detailed evidence, the rate of transition from NSAIDs alone to anti-TNFs was assumed to be constant, so a quarterly transition probability could be estimated using a standard rate-to-probability calculation as

$$\begin{aligned} TP &= 1-e^{-(1/m)\times cycleLength} \\ &= 1-e^{-(1/6.6)\times 0.25} \\ &= 0.0369 \end{aligned} \quad \textbf{(3)}$$

, where *m* is the mean duration of NSAID monotherapy.

However, the GDG advised that some people with axial spondyloarthritis achieve permanent symptom control with NSAIDs alone, and never progress to need anti-TNFs. For this reason, a proportion of the cohort – 25%, in the base case, as estimated by the GDG – remain in the NSAIDs state until they die.

H.2.4.2 Anti-TNF therapy

The effectiveness of anti-TNF therapy was based on the systematic review and evidence synthesis undertaken by the Centre for Reviews and Dissemination and the Centre for Health Economics at the University of York to support the recent NICE multiple technology assessment of TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis [TA383] (Corbett et al. 2016).

Anti-TNFs are treated as a single class, because the York synthesis shows there are no material differences between different agents. Therefore, the effectiveness of separate agents is not modelled (although their costs are accounted for in a weighted sum; see below).

Effects of anti-TNFs on BASDAI and BASFI

The synthesis model estimates change in BASDAI and BASFI for people receiving anti-TNFs or placebo conditional upon their BASDAI response status (that is, whether their BASDAI reduced by at least 50%, thereby achieving 'responder' status and mandating continued therapy). For a description of their methods and posterior mean results, see table 77 in Corbett et al. (2016) and accompanying text.

Whereas, in the York model, the effects of anti-TNFs could be estimated in a single population with single mean baseline values for BASDAI and BASFI, our model needed to estimate the effectiveness of treatment in cohorts with varying baseline values (that is, with a proportion of people commencing therapy in every cycle of the model and BASDAI and BASFI values changing along the way). The theoretically optimal way to calculate these effects would be to re-estimate the synthesis model many hundreds of times over, with new cycle-specific baseline values each time (and then re-estimate each of these again in every probabilistic iteration of the model in PSA). This is self-evidently unfeasible. Therefore, a simple approximation was adopted, where response to anti-TNFs was assumed to be proportional to starting value.

Table 5: Mean treatment effects of anti-TNF therapies, derived from York synthesis model

	Anti-TNFs		Placebo	
	Responders	Non-responders	Responders	Non-responders
Ankylosing spondylitis				
BASDAI				
Baseline	6.12			
12 weeks	0.94	5.43	1.12	5.97
Proportional effect	15.3%	88.8%	18.3%	97.5%
BASFI				
Baseline	5.28			
12 weeks	1.12	5.63	1.80	5.42
Proportional effect	21.2%	106.5%	34.1%	102.6%
Non-radiographic axial spondyloarthritis				
BASDAI				
Baseline	6.42			
12 weeks	1.14	5.23	1.20	5.80
Proportional effect	17.8%	81.5%	18.7%	90.4%
BASFI				
Baseline	4.92			
12 weeks	0.68	6.12	1.07	5.33
Proportional effect	13.8%	124.4%	21.8%	108.4%

For example, a small proportion of simulated patients with ankylosing spondylitis commence anti-TNFs in the first cycle of the model, if they are correctly diagnosed at their first presentation and contraindicated to NSAIDs. Such people start the model with a BASFI of 3.575, so the proportion who respond to anti-TNFs begin the next cycle with a BASFI of $3.575 \times 0.212 = 0.758$. However, people who live with a false-negative diagnosis for 10 years and then spend another 10 years on NSAIDs will have a BASFI of 5.912 by the time they start anti-TNFs; for them, response to anti-TNFs would lead to a BASFI of $5.912 \times 0.212 = 1.253$.

Disease duration and probability of response to anti-TNFs

The probability of responding to anti-TNFs is also dependent on duration of disease. Rudwaleit et al. (2004) show that the odds of response to anti-TNFs reduce by 6% for every year of symptoms a person has experienced – that is, disease duration (per year) is associated with an odds ratio of 0.94 (95%CI 0.89, 0.99). In ankylosing spondylitis, the York evidence synthesis suggests that the probability of response to anti-TNFs is 0.42, and we also know that, in the RCTs that contribute to the synthesis, the mean disease duration at baseline was 13.66 years. Combining these pieces of evidence, it is straightforward to work backwards and forwards in time to estimate probability of response at earlier and later timepoints:

$$\text{logit}(p_t) = \text{logit}(0.42) + \ln(0.94)(t - 13.66) \quad \textbf{(4)}$$

$$\therefore\ p_t = \frac{1}{1 + e^{-\left(\ln\left[\frac{0.42}{1-0.42}\right] + \ln[0.94][t-13.66]\right)}}$$

Figure 3 provides an illustration of the relationship between duration of symptoms and probability of response.

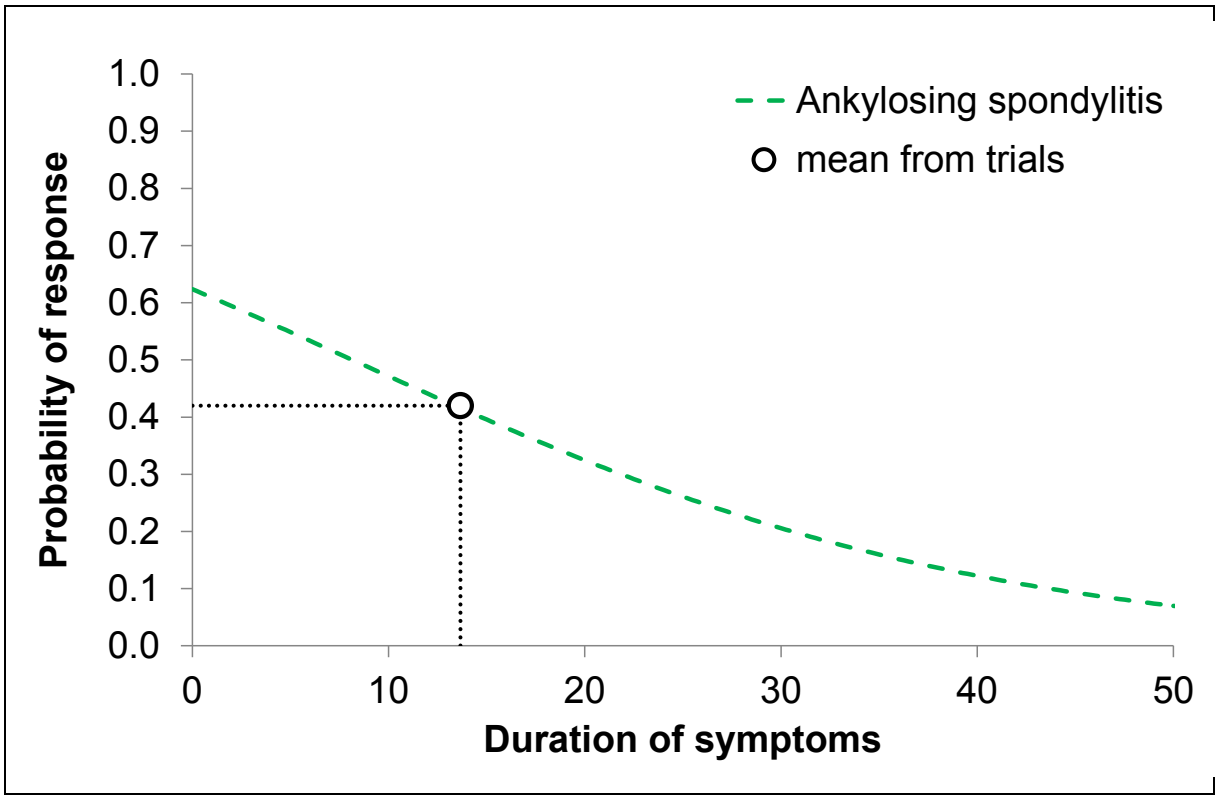


Figure 3: Duration-dependent probability of response to anti-TNF therapy

A separate function is estimated for non-radiographic axial spondyloarthritis; however, it is noteworthy that – by chance or otherwise – it is indistinguishable from the ankylosing spondylitis one. The nrAxSpA RCT evidence synthesised by York comprises a population with mean disease duration of 6.504 years and a probability of response of 0.53; according to the function estimated for ankylosing spondylitis, people with a disease duration of 6.504 years have a probability of response of 0.529.

Probability of response to second- and third-line anti-TNFs

The recommendations of TA383 suggest that alternative anti-TNF drugs may be tried for people whose disease has not responded – or has lost response – to the initial choice. To reflect this, up to 3 lines of therapy are simulated. Analysis of data from Glintborg et al. (2013) suggests that the odds ratio for response to second-line therapy, compared with first-line, is 0.502 (95%CI 0.402, 0.626) and the same value for third-line therapy is 0.364 (95%CI 0.249, 0.533). These values are used to modify equation (4) to estimate duration-dependent probability of response to second- and third-line anti-TNFs.

Duration of anti-TNF therapy

Simulated patients who enter the response state for each line of anti-TNFs have a probability of discontinuing treatment due to intolerance or lost response. Such people have a probability of discontinuing anti-TNF therapy altogether or commencing a subsequent line of treatment. These parameters were derived from Glintborg et al. (2013), and are tabulated in Table 6.

Table 6: Discontinuation of anti-TNFs (from Glintborg et al., 2013)

	Probability of remaining on therapy after 2 years	Implied quarterly probability of discontinuation	Probability of discontinuing anti-TNFs altogether at this point
First line	0.58 (95%CI 0.55, 0.61)	0.066	0.16 (95%CI 0.14, 0.18)
Second line	0.47 (95%CI 0.42, 0.52)	0.090	0.18 (95%CI 0.15, 0.22)
Third line	0.49 (95%CI 0.41, 0.57)	0.086	-

H.2.4.3 Best supportive care

On entry to the BSC state, BASDAI and BASFI are altered using evidence from the control arms of the anti-TNF synthesis (see Table 5); thereafter, the natural history trajectories are followed (see H.2.2).

Treatments applied in the BSC state are assumed to have no disease-modifying effect so, while they incur costs (see below), they have no impact on BASDAI or BASFI. This is not quite the same thing as saying that they have no effect; rather it is assumed that it is necessary to commit resources to the specified treatments in order to maintain BASDAI and BASFI (and, by extension, quality of life) at the levels that observed in the natural history of the disease.

A proportion of the cohort are assumed to receive an intensive course of chronic pain management. In the absence of any evidence on the number of people receiving such care, the GDG suggested that the proportion of people with AxSpA experiencing fibromyalgia would provide a reasonable surrogate estimate. Azevedo et al. (2010) report an incidence of 15.5% over a mean period of 16.59 years. This corresponds to a quarterly probability of 0.0025.

People are also assumed to receive physiotherapy and/or hydrotherapy, with additional analgesia for some (see costs, below).

H.2.4.4 False-negative states

For the states of the model in which people with spondyloarthritis have not yet been correctly diagnosed, BASDAI and BASFI progress according to natural history (see H.2.2). Costs of erroneous treatment for mechanical back-pain are included (see below).

H.2.4.5 Mortality

All-cause mortality is estimated using national mortality statistics. People with spondyloarthritis, whether diagnosed or undiagnosed, have an increased mortality rate which is represented in the model with the gender-specific standardised mortality rates reported by Bakland et al. (2011) applied to the average mortality rate for the population as a whole. This evidence suggests that men are subject to an increased hazard of death by a factor of 1.63 (95%CI 1.29, 1.97); the analogous figure for women is 1.38 (95%CI 0.48, 2.28). This results in people with spondyloarthritis in the model having an average life expectancy of 78.0, whereas a population of the same starting age without spondyloarthritis could expect to live, on average, to 82.0.

H.2.5 Health-related quality of life

Throughout the model, quality of life is estimated as a function of age, BASDAI and BASFI using a published mapping function (Wailoo et al., 2015). This is based on a sample of 1,615 observations of 516 Welsh people with ankylosing spondylitis (mean age 54.42; 76.5% male; see Atkinson et al., 2010, for a description of the study). We followed the authors' recommendation to use their 4-component mixture model. Estimates of utility are calculated for every state at every cycle in the model, and these are aggregated according to state occupancy.

No events in the model are associated with additional quality of life decrements or benefits.

H.2.6 Costs

The cost of each of the resource use elements within the model are obtained from a number of standard sources. Where these sources do not provide the unit cost needed to parameterise the cost of a resource use variable within the model then a search is conducted for unit costs generated from costing studies or within trials. Where the parameter is a key component of the model, a tailored systematic review can be conducted to locate the most appropriate unit cost.

The Prescription Pricing Authority drug tariff database is used for prices of drugs. The database is updated monthly therefore a single month’s tariff is used for all analysis to maintain consistency.

NHS Reference costs are used as the source of unit costs for inpatient and outpatient procedures as well as hospital stay information.

The Personal Social Services Research Unit (PSSRU) generates the Unit Costs for Health and Social Care report which includes costs for both community and hospital-based healthcare staff.

Where an appropriate reference cost cannot be sourced from national tariffs and the cost variable used is from a relevant published study, the value is inflated to current prices using the HCIS inflation indices.

H.2.6.1 Background NHS costs

Throughout the model, nonspecific background NHS costs are estimated as a function of BASFI using an approach that has been adopted in previous models of axial spondyloarthritis (data from Boonen et al., 2003, as implemented by Corbett et al., 2016):

$$annual\ NHS\ costs(£) = 1284 \times e^{0.213 BASFI} \quad \textbf{(5)}$$

This suggests that people with axial spondyloarthritis incur NHS costs ranging from £1,284 to £10,825 per year, depending on their level of functional impairment.

H.2.6.2 Specific modelled costs

H.2.6.2.1 NSAIDs

NSAID costs apply in NSAID monotherapy states, but the model also assumes that a majority of people continue to take them alongside anti-TNFs and in the BSC states. The proportion of continued use is estimated from baseline data on coprescription of NSAIDs in the anti-TNF RCTs, pooled using random-effects meta-analysis (with logistic transformation – that is, on a log-odds scale). Parameters are given in Table 7.

Table 7: NSAID costs

Item	Value (95%CI)	Source
Monthly etoricoxib (90mg daily)	£22.96	NHS drugs tariff 2016
Monthly naproxen (1g daily)	£2.76	
% COX-2 use	30.0% (19.0%, 42.4%)	GDG opinion
% PPI use alongside NSAIDs	100.0%	
Monthly PPI (omeprazole 10mg daily)	£1.03	NHS drugs tariff 2016
Quarterly NSAID + PPI costs	£29.55	
% continuing NSAIDs in anti-TNF & BSC states (AS)	86.1% (67.1%, 99.1%)	Meta-analysis of anti-TNF RCTs
% continuing NSAIDs in anti-TNF & BSC states (nrAxSpA)	70.5% (51.6%, 100.0%)	

H.2.6.2.2 Anti-TNFs

Drug costs

The approach to costing anti-TNFs was closely based on the York model (Corbett et al., 2016).

Acquisition costs of the drugs were drawn from BNF (2016), as prices are not available in the NHS drug tariff or eMIT. Standard dosages as detailed in each drug's SPC were assumed.

Administration costs depend on mode of administration. For subcutaneous drugs, a single nurse-led appointment (followed by self-administration) is assumed. For infliximab, which is delivered intravenously, an appointment is needed for each administration, with the cost taken from the NHS reference cost for a standard chemotherapy appointment.

Tests are accounted for at initiation of therapy and for ongoing monitoring. Resource use is as in Corbett et al. (2016), with the exception of testing for tuberculosis, where the GDG advised that the York model may have substantially underestimated costs by accounting for a single Mantoux test. In reality (and in accordance with updated NICE guidance that postdates TA383), candidates for anti-TNF therapy will receive an IGRA test, and those who have a positive result will be offered prophylactic therapy for latent TB. Accordingly, our model accounts for these costs.

Costs are summarised in Table 8.

Infliximab costs are substantially higher than those for subcutaneous therapies, both because of drug costs and increased administration costs. Costs are also somewhat higher for non-radiographic disease than for ankylosing spondylitis. This is because additional imaging is assumed (1 MRI at initiation; 1 X-ray per year during monitoring). The slightly higher costs for infliximab drug and administration in this population reflect the slightly higher probability of response (the infliximab SPC states that the initiation phase should be discontinued if there is no sign of response).

Table 8: Anti-TNF-related costs

Drug	First 3 months				Subsequent cycles			
	Drug	Admin	Tests	Total	Drug	Admin	Tests	Total
Ankylosing spondylitis								
Adalimumab	£2,112.84	£49.00	£497.73	£2,659.57	£2,296.71	–	£76.59	£2,373.30
Certolizumab[a]	–	£49.00	£497.73	£546.73	£2,331.67	–	£76.59	£2,408.26
Etanercept	£2,145.00	£49.00	£497.73	£2,691.73	£2,331.67	–	£76.59	£2,408.26
Golimumab[b]	£2,288.91	£49.00	£497.73	£2,835.64	£2,288.91	–	£76.59	£2,365.50
Infliximab	£5,638.97	£978.64	£497.73	£7,115.35	£3,127.80	£499.37	£76.59	£3,703.76
Weighted ave.[c]				£2,991.65				£2,477.05
Non-radiographic axial spondyloarthritis								
Adalimumab	£2,112.84	£49.00	£666.73	£2,828.57	£2,296.71	–	£83.14	£2,379.84
Certolizumab[a]	–	£49.00	£666.73	£715.73	£2,331.67	–	£83.14	£2,414.80
Etanercept	£2,145.00	£49.00	£666.73	£2,860.73	£2,331.67	–	£83.14	£2,414.80
Golimumab[b]	£2,288.91	£49.00	£666.73	£3,004.64	£2,288.91	–	£83.14	£2,372.05
Infliximab	£5,797.95	£1,006.23	£666.73	£7,470.91	£3,127.80	£499.37	£83.14	£3,710.31
Weighted ave.[c]				£3,173.57				£2,483.60

[a] *Subject to a public-domain PAS in which the manufacturer provides the drug free for the first 3 months*
[b] *Subject to a public-domain PAS in which the manufacturer provides the 100mg dose of golimumab at the same cost as the 50mg dose*
[c] *Weighted according to expected prescription frequency: 51% adalimumab; 0% certolizumab; 35% etanercept; 7% golimumab; 7% infliximab (NASS 2013)*

A weighted average of anti-TNF costs is applied in the model, with weights drawn from a large survey of UK patients, showing which agents are most commonly prescribed (NASS 2013). In the base case, this is

H.2.6.2.3 Chronic pain management

The costs of a course of chronic pain management were estimated using unpublished data from East Sussex Healthcare NHS Trust, who have asked us to keep details confidential. The total expected cost for people referred to such services is £898.94 for true positives (as part of best supportive care for known spondyloarthritis) and £887.44 for false negatives (as part of management of assumed mechanical back pain). Psychologist, physiotherapist and occupational therapist time is accounted for. The slight difference in totals reflects differences in the assumed involvement of musculoskeletal specialists in the referral pathway.

H.2.6.2.4 Best supportive care

In best supportive care states, it is assumed that, over and above the healthcare provision accounted for in background costs estimated via BASFI (see 0), 50% of people will receive an additional course of physiotherapy and 50% will receive an additional course of hydrotherapy. Additional analgesia is also accounted for.

Physiotherapy

Resource use was estimated using unpublished data obtained by a GDG member from East Sussex Healthcare NHS Trust. A 9-session course is assumed, with an average of 11 people seen across a 3-hour session. 3 hours of band-7 time, 3 hours of band-5 time and 0.5 hours of band-2 time are required. This results in a total cost of £225.82 per person for a complete course.

Hydrotherapy

Resource use and pool costs were estimated using unpublished data obtained by a GDG member from East Sussex Healthcare NHS Trust. A 9-session course is assumed, with an average of 20 people seen across a 3-hour session. 3 hours of band-7 time, 3 hours of band-5 time and 3 hours of band-2 time are required. The useable hourly cost of the pool is estimated as £10.20. This results in a total cost of £164.97 per person for a complete course.

Analgesia

It is assumed that the same proportion of people experiencing severe enough pain to warrant a chronic pain management course will also require additional analgesia. The GDG advised on common agents and dosages based on their experience – see Table 9.

Table 9: Additional analgesia for chronic pain management

Drug	% using	Usual dosage	Unit cost	Cost per 3mo cycle
Amitriptyline	50%	20mg od	£0.91 for 28 x 10mg	£2.97
Buprenorphine patch	30%	40% 5mcg/hr 39% 10mcg/hr 21% 20mcg/hr	5mcg/hr £17.60 10mcg/hr £31.55 20mcg/hr £57.46	£122.59
Gabapentin	20%	600mg tds	£6.74 for 100	£3.69
Tramadol	40%	50mg tds	£2.90 for 100	£3.18
Total				**£132.42**

H.2.6.2.5 Treatment for mechanical back pain in false negatives

The costs of treatment incurred in false-negative cases that are erroneously managed as mechanical back pain are drawn from the STarT Back risk stratification and management economic analysis (Whitehurst et al., 2012). We assumed that nobody with occult spondyloarthritis would present with back pain that would be stratified as low risk, and apportioned cases proportionally between medium and high risk – see Table 10.

Table 10: Costs of treatment for mechanical back pain in false-negative cases

Risk group	Proportion	Annual cost
Medium risk	61.9%	£258.06
High risk	38.1%	£421.41
Weighted average cost per 3mo cycle		**£80.06**

H.2.6.2.6 Diagnostic work-up

When a simulated referral strategy refers people for specialist diagnosis, it is necessary to account for the costs that are incurred in establishing whether that person does or does not have spondyloarthritis (true-positive or false-positive referral).

The unit costs used are tabulated in Table 11.

Table 11: Unit costs of elements of diagnostic work-up

Item	Cost	Source
GP appointment	£45.00	PSSRU (2015)
X-ray	£30.23	NHS Reference Costs (2014/15)
MRI	£162.38	NHS Reference Costs (2014/15)
HLA-B27	£37.55	Avg Trust Prices
ESR	£2.98	TA 199
CRP	£6.62	Henriksson et al. (2010), Corbett et al. (2016)
DEXA	£59.44	NHS Reference Costs (2014/15)
Ultrasound (outpatient)	£55.00	NHS Reference Costs (2014/15)
Specialist consultation	£137.00	NHS Reference Costs (2014/15)

No publicly available 'list' price is available for HLA B27 assay; the GDG advised that each laboratory has a separate negotiated tariff for such tests. Therefore, information on current cost was obtained from GDG members – 3 were able to provide details of cost in their locality. These were averaged, producing an estimate of £37.55 per test.

The GDG provided an estimate of resource use for true-positive and false-positive pathways (Table 12 and Table 13, respectively). These are similar – of course, the true status of the person is, by definition, unknown at the time of referral – however, false-positive referrals are predicted to cost somewhat more, because test results will invariably be negative, in these cases, whereas true-positive cases may be identified conclusively after relatively little investigation.

The diagnostic accuracy – sensitivity and specificity – of the components is required to calculate flow through the diagnostic pathway. These values are derived from systematic review and evidence synthesis performed for this guideline – see full guideline sections 4.2 and 4.3.

Table 12: Diagnostic work-up in true-positive cases

Assumption	Value (95%CI)
Proportion specialist consultation	1.0000
Proportion having X-ray as first imaging test	0.6500
Proportion having 1 X-ray	0.8571
Proportion having 3 X-rays	0.1429
Sensitivity of X-ray	0.292 (0.210, 0.392)
Proportion having no further tests if positive on X-ray	0.7000
Proportion having MRI if positive on X-ray	0.3000
Proportion having MRI if negative on X-ray	1.0000
Proportion having MRI as first imaging test	0.3500
Sensitivity of MRI	0.473 (0.441, 0.504)
Proportion having no further tests if positive on MRI	1.0000
Proportion having further tests if negative on MRI	0.2000
Proportion having initial HLA B27 test independent of imaging	0.4000
Proportion having HLA B27 if negative on X-ray	1.0000
Sensitivity of HLA B27	0.683 (0.571, 0.776)
Proportion ESR	1.0000
Proportion CRP	1.0000
Proportion diagnosed SpA having DEXA	1.0000
Total costs	**£434.34**

Table 13: Diagnostic work-up in false-positive cases

Assumption	Value
Proportion Specialist consultation	1.0000
Proportion having X-ray as first imaging test	0.6500
Proportion having 1 X-ray	0.8571
Proportion having 3 X-rays	0.1429
Specificity of X-ray	0.985 (0.939, 0.997)
Proportion having MRI when negative on X-ray	0.9000
Proportion having no further tests when negative on X-ray	0.1000
Proportion having MRI first	0.3500
Specificity of MRI	0.989 (0.940, 0.998)
Proportion having no further tests when negative on MRI	0.6500
Proportion having further tests when negative on MRI	0.3500
No. of additional MRIs for people with negative findings	2
No. of additional specialist consultations for people with negative findings	2
Proportion having initial HLA B27 test independent of imaging	0.4000
Proportion having HLA B27 when negative on X-ray	1.0000
Specificity of HLA B27	0.848 (0.770, 0.903)
Proportion ESR	1.0000
Proportion CRP	1.0000
Total cost	**£558.83**

H.3 Calculating the cost effectiveness of recognition strategies

Rather than simulating each possible strategy individually, the model calculates the discounted lifetime costs and QALYs expected from true-positive and false-negative cases (as well as the costs associated with false-positive referrals). The costs and effects of any strategy can then be calculated as an average of the relevant outputs, weighted according to the proportion of TP, FN and FP cases the strategy is predicted to produce (which is, in turn, a simple function of the sensitivity and specificity of the strategy and the true prevalence of axial spondyloarthritis in the presenting population).

Strategies that were evaluated in the model are described in Table 14. Evidence of the appropriate type – that is, following people with possible AxSpA until final diagnosis, regardless of whether they met particular criteria – is limited, and dominated by reports from 2 cohorts (Braun et al. 2011, 2013; van Hoeven et al. 2015).

The evidence from the systematic literature reviews conducted for the guideline (see full guideline, section 4.1) forms the basis of the strategies to be modelled. In addition, a paper by Sieper et al. (2013) provided a way of estimating the correlations between variables. Although the referral strategies explored in the trial are excluded from analysis as only the referred cohort were evaluated for their disease status, thus presenting an incomplete picture of the accuracy of the strategies for referral, the tables of results include many different combinations of variables that may form a referral rule, which can be deconstructed to identify how the presence of individual factors may be related to one another. The evidence of the diagnostic accuracy of individual factors could then be used along with the estimates of correlation to generate hypothetical strategies.

A supplement to the evidence identified in the review of evidence for RQ12 (Braun, 2014 & van Hoeven et al, 2015) comes from a piece of analysis published in an abstract in which the cohort used to develop the CaFaSpA referral rules (van Hoeven et al, 2015) is used to validate alternative referral criteria.

Table 14: Evaluated strategies

Study	Strategy	Description
van Hoeven (2015)	>=x	A score of x or more on the CaFaSpA scoring system: • positive ASAS IBP questionnaire (1pt) • family history (1pt) • good response to NSAIDs (1pt) • duration >5yr (0.5pt)
van Hoeven (ASAS) – validation of ASAS referral criteria in CaFaSpA cohort (van Hoeven et al. 2015)	>=x	x or more criteria from the ASAS referral criteria met (as validated in the CaFaSpA cohort): • IBP • arthritis, enthesitis or dactylitis • psoriasis, IBD or uveitis • family history • good response to NSAIDs • elevated CRP or ESR • HLA-B27 positivity • Sacroiliitis on imaging (if available)
van Hoeven (SSB27) – combinations of features assessed in CaFaSpA cohort (van Hoeven et al. 2014)	>=x	x or more criteria (signs, symptoms and/or HLA-B27 positivity): • IBP • arthritis • enthesitis • dactylitis • psoriasis • IBD • uveitis • family history • good response to NSAIDs • elevated CRP • HLA-B27 positivity
Braun (2011)	>=x	x or more criteria for the recognition of axial SpA met: • age at onset ≤35 • wakening in the second half of the night • alternating buttock pain • improvement by NSAIDs within 48h • improvement by movement, not rest.
Braun (2013)	Buttock OR HLA B27	either buttock pain or HLA-B27 positivity
Braun (2013)	2-step	2 or more of the following: • improvement by movement • buttock pain • history of psoriasis or • HLA-B27 positivity

Study	Strategy	Description
Braun (2013)	>=x	x or more of the following criteria: • age at onset of chronic BP ≤35 • waking during the second half of the night • buttock pain • improvement by movement • improvement by NSAIDs within 48 h • first-grade relatives with AS • history of arthritis • history of enthesitis • history of psoriasis • HLA-B27 positivity
Sieper (2013)	as specified	specified combinations of features
HLA B27	alone	from evidence synthesis for this guideline (see @@)

H.3.1 Accuracy and cost of referral strategies

The sensitivity and specificity of referral strategies is derived directly from reported data.

The 'up-front' cost of each possible strategy is obviously a component of relative cost effectiveness. Because all strategies require a primary consultation, the costs of this initial step are ignored; however, the costs of tests and any additional consultations required to make a referral decision are estimated.

In particular, several possible strategies included HLA B27 testing as a potential cost. Where HLA B27 testing was a necessary component of the strategy, the cost was simply added on. More commonly, however, HLA B27 appears on a list of possible factors to be considered (see Table 14). In these cases, it will, in practice, frequently not be necessary to undertake the test – if a person fulfils the necessary minimum number of criteria using signs, symptoms and risk factors alone, there would be no benefit in establishing HLA B27 status before making a referral. Unfortunately, it is not possible to estimate the proportion of cases of which this is true without access to more detailed information than is published in the relevant papers. Therefore, we took the conservative decision to assume all such strategies incurred the costs of an HLA B27 test, although this will somewhat overestimate the true costs of the strategy.

A second complication to the estimation of HLA-B27 costs is that, if the test is undertaken as part of a referral strategy, it will not need to be repeated in specialist setting (as it has a binary, permanent result). Therefore, the costs of establishing true diagnosis in people who are referred with HLA B27 results are slightly lower than in other cases (see H.2.6.2.6). Rather than calculating separate diagnosis costs for every combination of true-positive and false-positive referrals, we simply calculated the amount each is reduced by when the need for the test is obviated, and deducted this from the 'up-front' test cost. The reduction turned out to be -£30.97 for every true-positive referral with HLA-B27 results and -£37.23 for every false-positive one.

Table 15: Accuracy and cost of evaluated strategies

Source	Strategy	Cost	Sens	Spec
	'Current practice'	£0.00	10.66%	99.44%
	'Refer everyone'	£0.00	100.00%	0.00%
van Hoeven (2015)	>=1.0	£0.00	92.63%	39.05%
van Hoeven (2015)	>=1.5	£0.00	74.74%	57.64%
van Hoeven (2015)	>=2	£0.00	41.05%	82.44%
van Hoeven (2015)	>=2.5	£0.00	28.42%	88.22%
van Hoeven (ASAS)	>=1	£7.20	99.73%	18.55%
van Hoeven (ASAS)	>=2	£21.91	99.73%	60.13%
van Hoeven (ASAS)	>=3	£31.21	66.85%	86.45%

Source	Strategy	Cost	Sens	Spec
van Hoeven (ASAS)	>=4	£34.75	30.39%	96.45%
van Hoeven (ASAS)	>=5	£35.59	9.39%	98.82%
van Hoeven (ASAS)	>=6	£35.87	2.76%	99.61%
Braun (2011)	>=4	£0.00	47.79%	86.12%
Braun (2011)	>=3	£0.00	78.76%	46.41%
Braun (2011)	>=2	£0.00	96.46%	17.22%
Braun (2013)	Buttock pain OR HLA B27	£6.80	89.72%	40.31%
Braun (2013)	Buttock pain AND HLA B27	£33.78	45.79%	93.72%
Braun (2013)	2-step	£21.64	80.37%	75.39%
Braun (2013)	>=1	£1.57	99.07%	2.62%
Braun (2013)	>=2	£3.23	97.20%	7.33%
Braun (2013)	>=3	£10.08	93.46%	26.70%
Braun (2013)	>=4	£23.04	85.98%	63.35%
Braun (2013)	>=5	£34.34	53.27%	95.29%
Braun (2013)	>=6	£35.82	23.36%	99.48%
Braun (2013)	>=7	£35.82	4.67%	99.48%
Braun (2013)	>=8	£35.91	0.46%	99.74%
Sieper (2013)	B27+EAM (both)	£34.88	23.11%	96.83%
Sieper (2013)	B27+EAM (either)	£19.28	79.33%	67.18%
Sieper (2013)	B27+IBP (both)	£34.55	47.68%	95.90%
Sieper (2013)	B27+IBP (either)	£10.41	88.06%	48.22%
Sieper (2013)	B27+ response to NSAIDs (both)	£34.35	43.31%	95.31%
Sieper (2013)	B27+ response to NSAIDs (either)	£9.60	90.73%	46.34%
Sieper (2013)	HLA B27, Family history & EAM (2 of 3)	£32.18	42.91%	89.97%
Sieper (2013)	HLA B27, EAM, response to NSAIDs (2 of 3)	£25.29	68.49%	77.15%
Sieper (2013)	EAM & family history (both)	£0.00	10.44%	99.23%
Sieper (2013)	EAM & family history (either)	£0.00	48.00%	65.89%
Sieper (2013)	EAM & response to NSAIDs (both)	£0.00	14.83%	90.64%
Sieper (2013)	EAM & response to NSAIDs (either)	£0.00	85.14%	45.40%
Sieper (2013)	Family history, EAM & response to NSAIDs	£0.00	43.43%	84.54%
Sieper (2013)	Family history, EAM & IBP	£0.00	55.92%	72.30%
Sieper (2013)	EAM, IBP & response to NSAIDs	£0.00	85.29%	47.78%
van Hoeven (SSB27)	>=1	£10.88	99.73%	28.95%
van Hoeven (SSB27)	>=2	£24.00	64.09%	66.05%
van Hoeven (SSB27)	>=3	£32.10	27.07%	88.95%
HLA B27	alone	£30.63	68.25%	84.81%

Two scenarios are presented in addition to the empirically parameterised strategies:

- An approximation of 'current practice'. The sensitivity of the scenario is based on data showing that 10.7% of people with spondyloarthritis are currently diagnosed within 3 months of their first presentation (NASS 2013). The specificity is calculated to match the GDG's belief that approximately 50% of people currently referred with suspected spondyloarthritis are ultimately diagnosed with the condition. This is a function of prevalence as well as sensitivity; at the base-case prevalence of 5%, the specificity of current practice is calculated as 99.4%.
- A demonstration of what would happen if everyone meeting the population criteria was referred to specialist care. This is easily simulated by specifying a sensitivity of 100% – that is, everyone with SpA becomes a true-positive referral – and a specificity of 0% – that is, no one who does not have SpA becomes a true-negative non-referral.

Both scenarios are assumed to bear no cost.

H.4 Results

H.4.1.1 Clinical outcomes from the model

The aggregated effect of changes in BASFI and BASDAI consequent on effective treatment is shown in Figure 4 and Figure 5. Note that, in these graphs, false-negative status refers to cases that were not referred at initial presentation. As time goes on, an increasing proportion of these have their true diagnosis discovered and their status corrected; hence, the profiles converge over time.

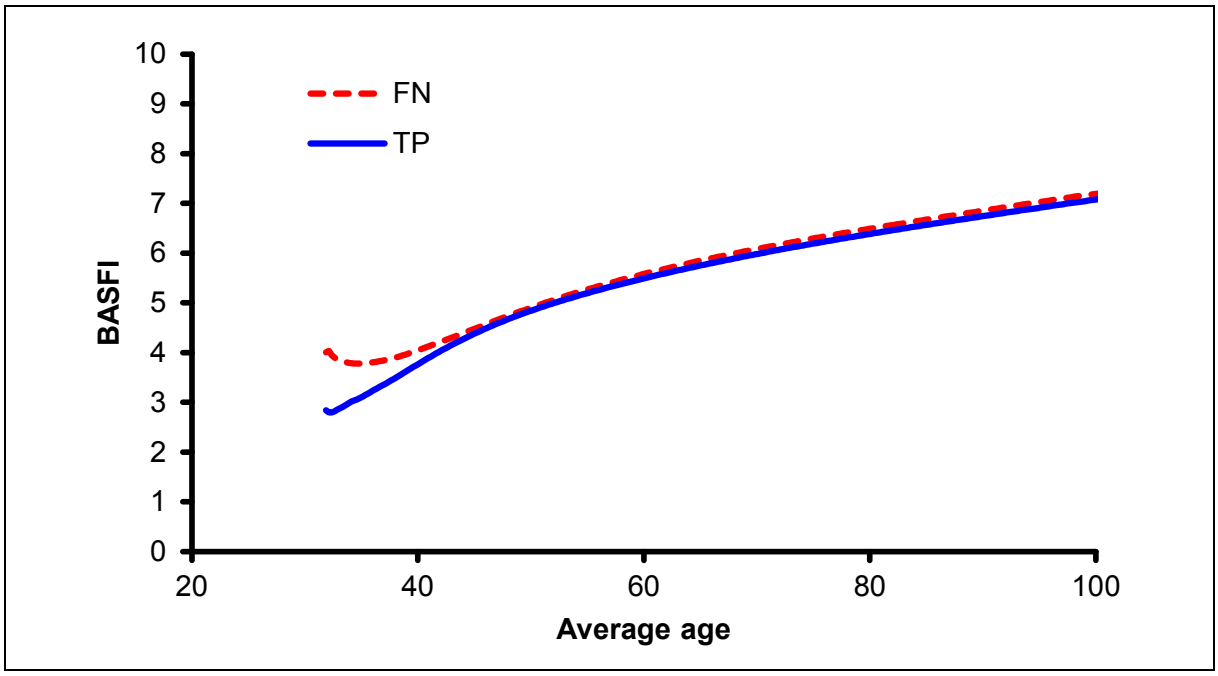


Figure 4: Modelled BASFI for true-positive referrals and false-negative non-referrals

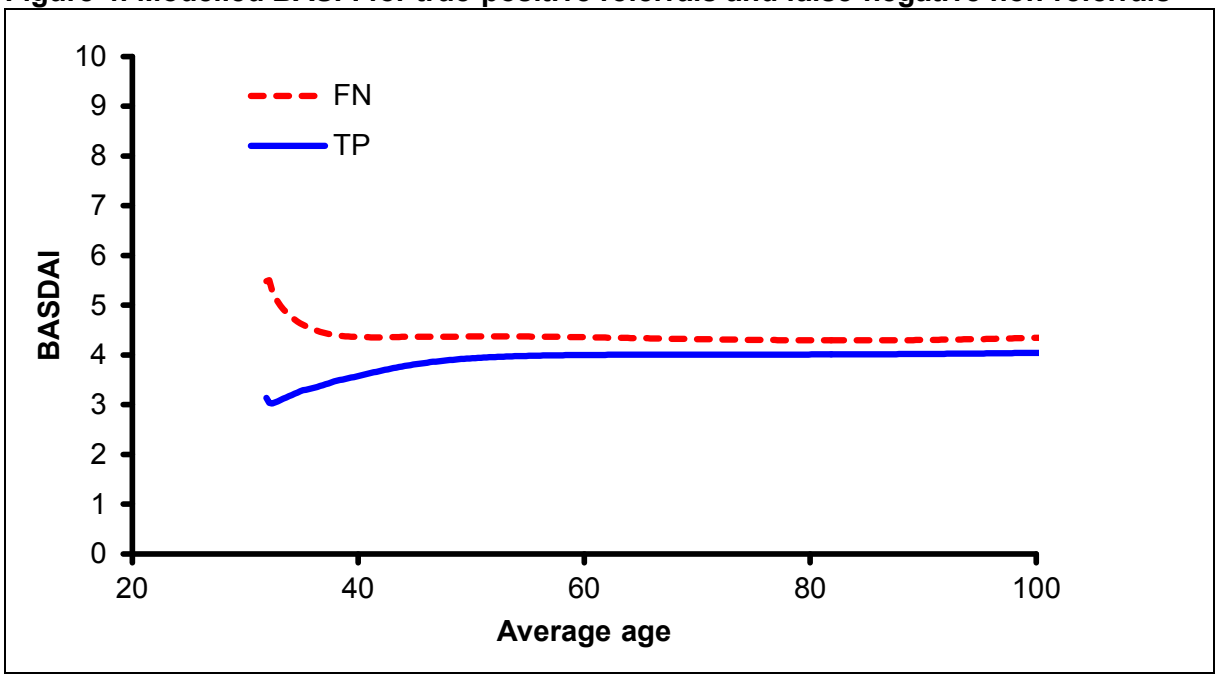


Figure 5: Modelled BASDAI for true-positive referrals and false-negative non-referrals

State occupancy over time for true-positive referrals and false-negative non-referrals is shown in Figure 6. The considerable dwell-time with false-negative status is conspicuous in the latter. It is also clear that, owing to lower chance of response, the average person spends less time receiving anti-TNFs if their diagnosis is delayed.

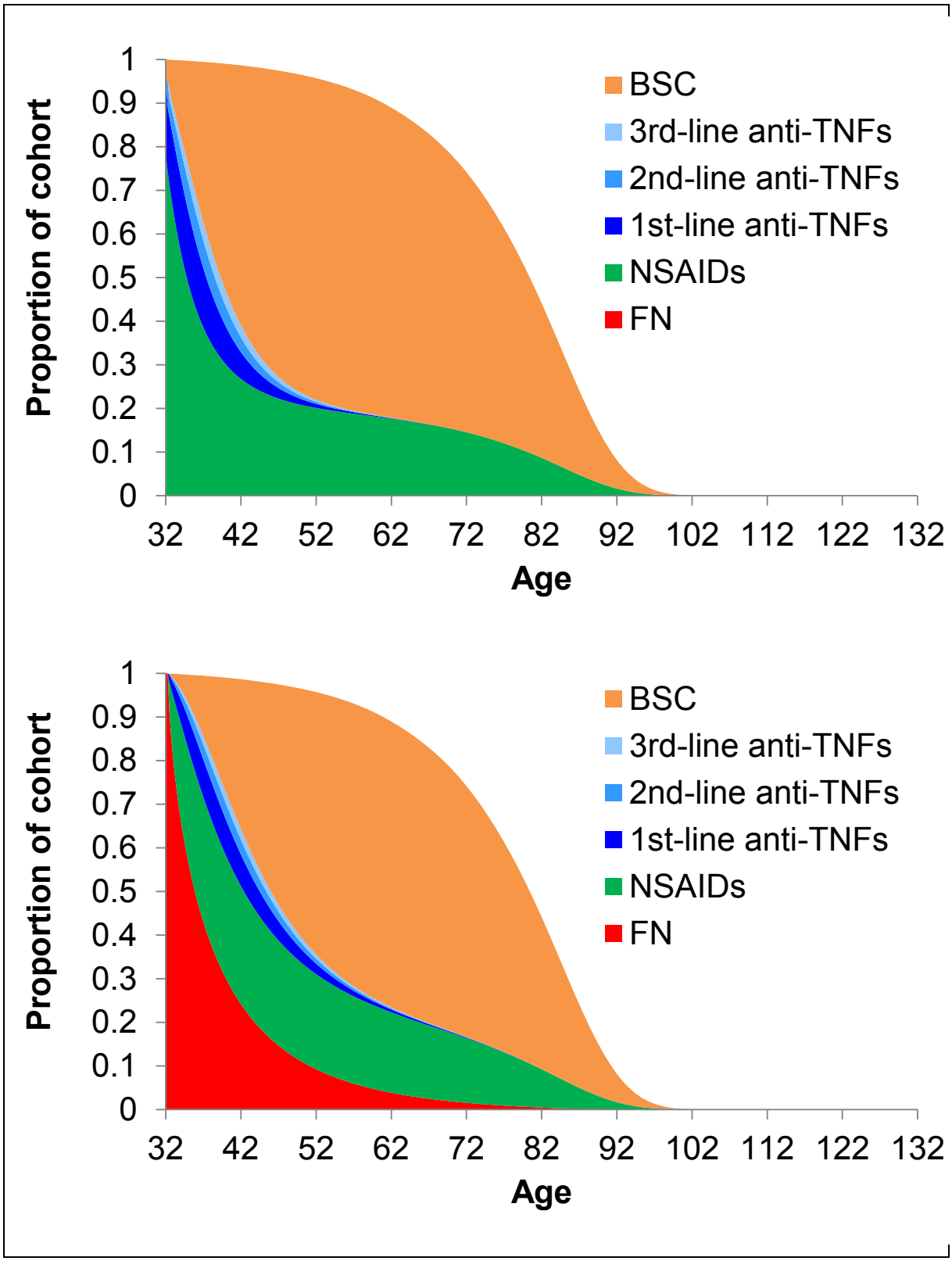


Figure 6: Modelled state occupancy for true-positive referrals (top) and false-negative non-referrals (bottom)

H.4.1.2 Base-case cost–utility results

H.4.1.2.1 Theoretical trade-off between sensitivity and specificity

Any model with a broadly diagnostic focus is concerned with establishing the optimal balance between sensitivity and specificity of possible approaches – this is, how to minimise false-negative decisions while not causing an excess of false-positive cases.

For this reason, it is instructive to explore the theoretical relationship between sensitivity and specificity of referral strategies before proceeding to analyse any real-world approaches. Figure 7 depicts relative cost effectiveness (measured as net monetary benefit with an arbitrary threshold) for any combination of these parameters.

It shows that, in this model, the trade-off between sensitivity and specificity is critically influenced by assumed prevalence of disease. At low prevalences (e.g. the 1.8% reported by Hamilton et al., 2015; see H.2.2), specificity is somewhat more important than sensitivity In contrast, as prevalence becomes higher (e.g. the 16.4% reported by van Hoeven et al., 2015; see H.2.2), sensitivity becomes the dominant determinant of cost effectiveness – it can be seen that strategies with very high sensitivity have similar value for money, almost regardless of how specific they are. At the intermediate prevalence of 5% favoured by the GDG for our base-case analysis, there is a finer balance: highly sensitive strategies tend to provide best value for money but, if corresponding specificity is low, similar results can be achieved by strategies with high specificity and only moderate sensitivity.

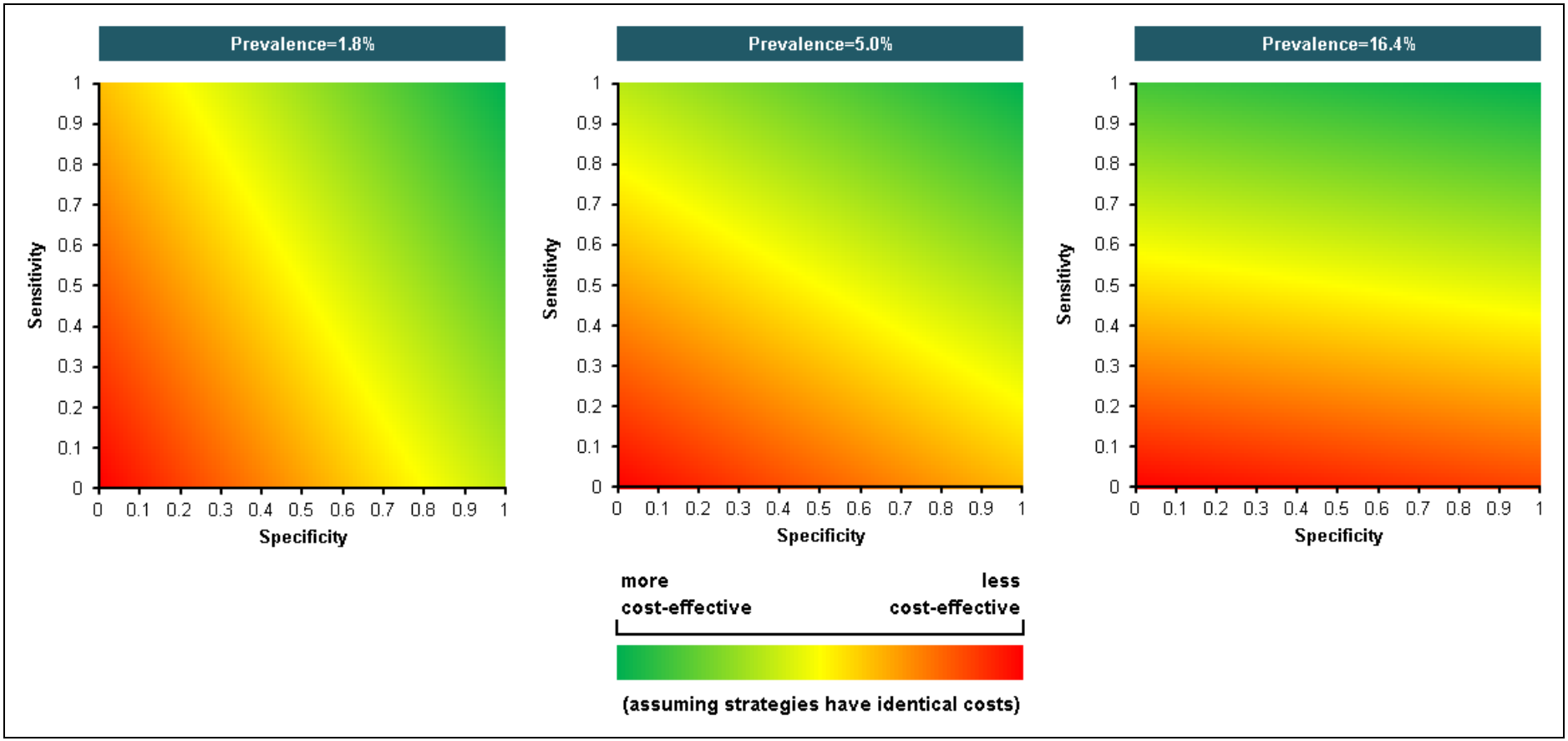


Figure 7: Cost–utility model: illustrative relationship between prevalence and cost effectiveness

H.4.1.2.2 Deterministic base case

Table 16: Base-case deterministic cost–utility results – costs and QALYs associated with diagnostic outcomes

	Discounted lifetime costs			Discounted lifetime QALYs
	Background	Specific	Total	
True positives	£79,951	£27,356	£107,307	14.571
False negatives	£83,684	£21,282	£104,966	13.534
False positive	–	£559	£559	–
True negatives	–	–	–	–

The model predicts that, on average, a person with axial spondyloarthritis who is correctly referred for specialist assessment at their first contact with healthcare services accrues just over 1 QALY more, over their lifetime, than a similar person who is not referred. However, timely referral is also estimated slightly to increase lifetime healthcare costs. This is because more people end up receiving costly interventions – notably biological DMARDs – earlier in their disease course (and remain on them for longer). This additional expense is partially offset by a reduction in background healthcare costs, with the net result that the average true-positive referral costs the NHS around 2% more, over their lifetime, than the average false negative. The costs accrued by specialist care in identifying the negative disease status of false-positive referrals is estimated at £559 each.

It follows that the expected lifetime discounted costs and QALYs of any referral strategy may be estimated as follows:

$$Costs = C + vn107{,}307 + v(1-n)104{,}966 + (1-v)(1-p)559 \quad \textbf{(6)}$$

$$QALYs = vn14.5\ 71 + v(1-n)13.534 \quad \textbf{(7)}$$

, where C is the initial cost of the strategy (for example, the cost of any tests administered or extra appointments required), v is the true prevalence of spondyloarthritis among people with chronic back pain of ≥3 months' duration with age of onset ≤45, n is the sensitivity of the strategy and p is the specificity of the strategy.

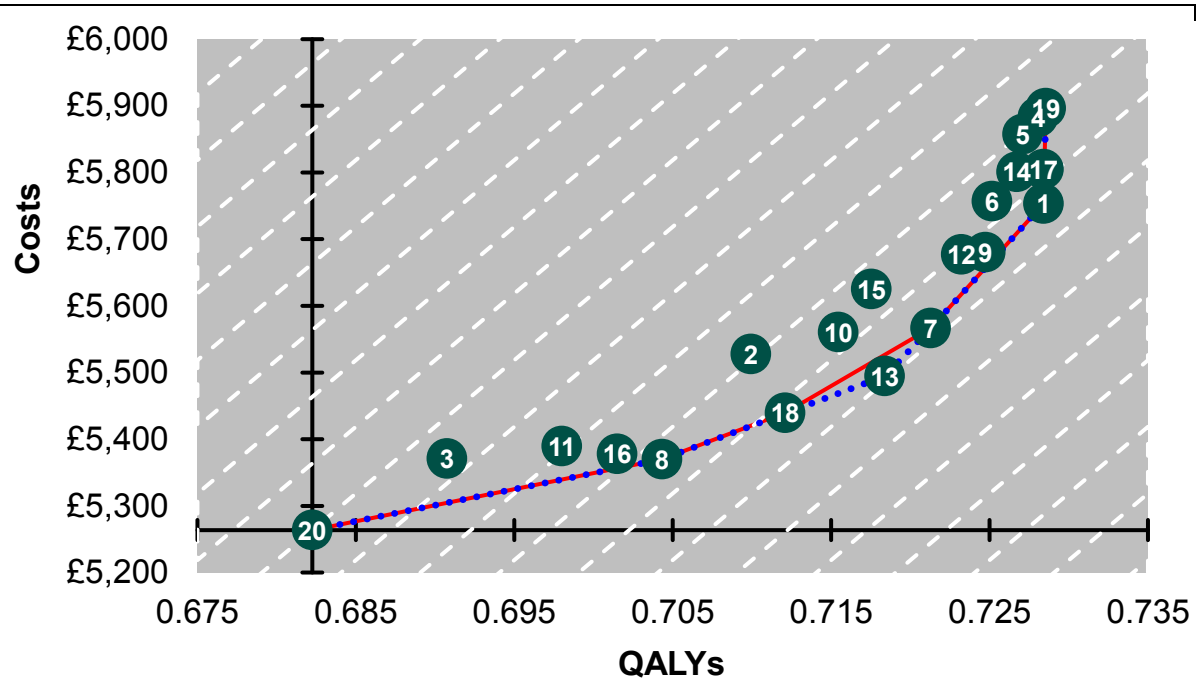
Base-case cost–utility results are tabulated in Table 17 and illustrated in Figure 8. Results are presented for 18 of the strategies for which published data were available, as well as 2 additional scenarios – one that provides an approximation of 'current practice' and one that shows what would happen if everyone was referred to specialist care. The former is based on data on the proportion of people who are diagnosed on first presentation (NASS 2013); the latter is easily simulated with a sensitivity of 100% – that is, everyone with SpA becomes a true-positive referral – and a specificity of 0% – that is, no one who does not have SpA becomes a true-negative non-referral.

One strategy with apparently good sensitivity (>80%) and specificity (>75%) is the Braun (2013) '2-step' algorithm, in which people with possible SpA are referred on the basis of clinical questions and/or HLA-B27 positivity. However, the GDG expressed doubts about how methodologically sound, clinically meaningful and practically replicable the proposed algorithm is (especially in its reliance on reported both-sided buttock pain; see full guideline section 4.1.4). For this reason, incremental results are shown for a decision-space that includes this strategy and one that excludes it.

Table 17: Base-case deterministic cost–utility results – possible strategies

Strategy	Accuracy		Costs (£)	Effects (QALYs)	Incremental					
					Including Braun (2013) 2-step			Excluding Braun (2013) 2-step		
	Sensitivity (%)	Specificity (%)			Costs (£)	Effects (QALYs)	ICER (£/QALY)	Costs (£)	Effects (QALYs)	ICER (£/QALY)
'Current practice'	10.7%	99.4%	£5,264	0.6823						
Braun (2013): >=5	53.3%	95.3%	£5,370	0.7043	£106	0.0221	£4,812	£106	0.0221	£4,812
Van Hoeven (SSB27): >=3	27.1%	88.9%	£5,371	0.6908	£1	-0.0136	dominated	£1	-0.0136	dominated
Braun (2011): >=4	47.8%	86.1%	£5,378	0.7015	£8	-0.0028	dominated	£8	-0.0028	dominated
Van Hoeven (2015): >=2	41.1%	82.4%	£5,390	0.6980	£20	-0.0063	dominated	£20	-0.0063	dominated
HLA B27: alone	68.3%	84.8%	£5,439	0.7121	£69	0.0078	ext. dom.	£69	0.0078	£8,943
Braun (2013): 2-step	80.4%	75.4%	£5,495	0.7184	£125	0.0140	£8,876	–	–	–
Van Hoeven (SSB27): >=2	64.1%	66.1%	£5,528	0.7099	£33	-0.0084	dominated	£88	-0.0022	dominated
Van Hoeven (2015): >=1.5	74.7%	57.6%	£5,561	0.7155	£66	-0.0029	dominated	£121	0.0034	ext. dom.
Braun (2013): >=4	86.0%	63.4%	£5,567	0.7213	£72	0.0029	£24,747	£127	0.0092	£13,839
Braun (2011): >=3	78.8%	46.4%	£5,625	0.7175	£58	-0.0037	dominated	£58	-0.0037	dominated
Braun (2013): Buttock OR HLA B27	89.7%	40.3%	£5,677	0.7232	£110	0.0019	ext. dom.	£110	0.0019	ext. dom.
Van Hoeven (2015): >=1.0	92.6%	39.0%	£5,680	0.7247	£114	0.0034	ext. dom.	£114	0.0034	ext. dom.
Van Hoeven (SSB27): >=1	99.7%	28.9%	£5,753	0.7284	£187	0.0071	£26,199	£187	0.0071	£26,199
Braun (2013): >=3	93.5%	26.7%	£5,757	0.7252	£4	-0.0032	dominated	£4	-0.0032	dominated
Braun (2011): >=2	96.5%	17.2%	£5,801	0.7267	£48	-0.0017	dominated	£48	-0.0017	dominated
van Hoeven (ASAS): >=1	99.7%	18.6%	£5,805	0.7284	£52	0.0000	ext. dom.	£52	0.0000	ext. dom.
Braun (2013): >=2	97.2%	7.3%	£5,857	0.7271	£104	-0.0013	dominated	£104	-0.0013	dominated
Braun (2013): >=1	99.1%	2.6%	£5,883	0.7281	£130	-0.0003	dominated	£130	-0.0003	dominated
'Refer everybody'	100.0%	0.0%	£5,896	0.7286	£143	0.0001	£992,832	£143	0.0001	£992,832

ext. dom. = extendedly dominated



(1) Van Hoeven (SSB27): >=1
(2) Van Hoeven (SSB27): >=2
(3) Van Hoeven (SSB27): >=3
(4) Braun (2013): >=1
(5) Braun (2013): >=2
(6) Braun (2013): >=3
(7) Braun (2013): >=4
(8) Braun (2013): >=5
(9) Van Hoeven (2015): >=1.0
(10) Van Hoeven (2015): >=1.5
(11) Van Hoeven (2015): >=2
(12) Braun (2013): Buttock OR HLA B27
(13) Braun (2013): 2-step
(14) Braun (2011): >=2
(15) Braun (2011): >=3
(16) Braun (2011): >=4
(17) van Hoeven (ASAS): >=1
(18) HLA B27: alone
(19) 'Refer everybody'
(20) 'Current practice'

Red continuous line shows cost-effectiveness frontier excluding Braun (2013) 2-step; blue dotted line shows frontier including Braun (2013) 2-step
Dashed grey lines in background indicate ICER gradients of £20,000/QALY

Figure 8: Base-case deterministic results – cost–utility plane

QALY gains appear small, in absolute terms; however, it should be remembered that a substantial majority (95%, in the base case) of simulated patients in the model do not have SpA and, thus, experience no benefit or harm from better or worse recognition of SpA. This means that the substantial gains in quality of life for the minority of people who do have SpA appear, on face value, to be diluted by the experience of people who do not have disease. For example, under the Braun (2013) >=4 strategy, the average person with SpA gains 0.781 QALYs compared with estimated current practice. However, it is necessary to account for people without disease in the denominator of cost-per-QALY calculations, as the costs they incur are important constituents of the numerator.

If it is considered credible that the reported results of the Braun (2013) '2-step' algorithm can be replicated in NHS practice, then it is likely to be judged the optimal strategy. Compared with approximated current practice, it produces 0.036 QALYs per person at an incremental cost of £231. Several strategies have somewhat superior sensitivity and, as a consequence, somewhat superior effectiveness but, because all these strategies are also less specific than the '2-step', the incremental benefit they provide comes at an additional cost that exceeds £20,000 per QALY gained. For example, Braun (2013) >=4, being 5% more sensitive than the '2-step', is associated with 0.003 extra QALYs (approximately 1 quality-adjusted life-day) but, because it is 12% less specific, it also costs £72 per presenting person more. This produces an ICER of £24,750 per QALY gained.

If the Braun (2013) '2-step' algorithm is excluded from the decision space, Braun (2013) >=4 is likely to be considered to represent the best balance of costs and benefits. Compared with approximated current practice, it produces 0.039 QALYs per presenting person at an incremental cost of around £300 and, in incremental analysis, it is associated with an ICER of £13,800 compared with the next-cheapest non-dominated alternative (HLA-B27 alone). Again, slightly more QALYs may be gained by other strategies; in this case, van Hoeven (2014) >=1 (which is 100% sensitive but only 29% specific) generates 0.007 extra QALYs at an incremental cost approaching £200 per presenting case, leading to an ICER of £26,200 per QALY gained compared with Braun (2013) >=4.

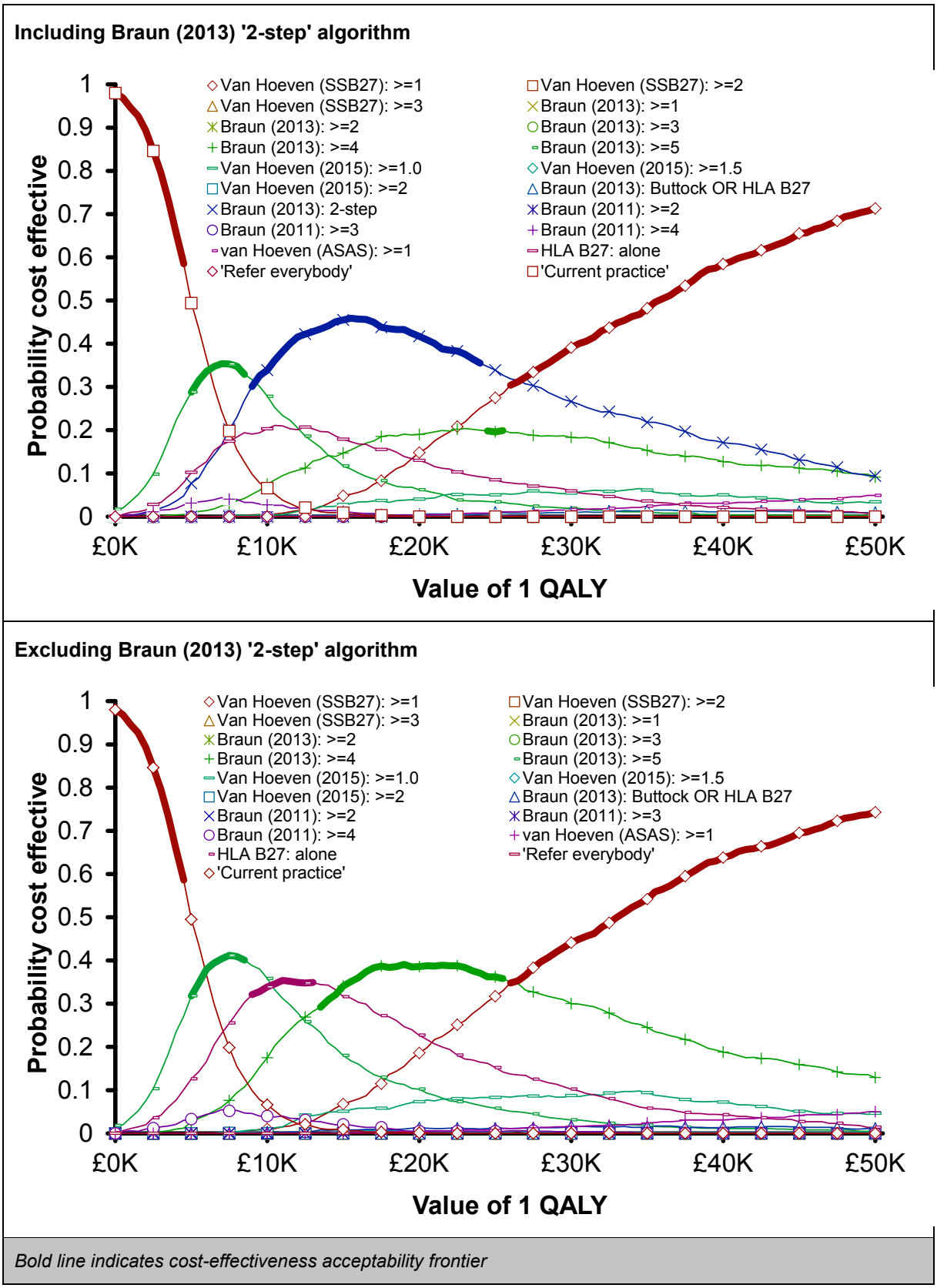
It is a collateral benefit of the approach adopted in this model that it is not necessary to access the model to estimate the cost effectiveness, compared with 'current practice', of any referral strategy for which an estimate of sensitivity and specificity is available, by extending equations (6) and (7) to an incremental comparison with costs and QALYs estimated for 'current practice' in Table 17:

$$ICER\text{ -}v\text{- }current = \frac{C + vn107{,}307 + v(1-n)104{,}966 + (1-v)(1-p)559 - 5{,}264}{vn14.571 + v(1-n)13.534 - 0.682} \quad \textbf{(8)}$$

H.4.1.3 PSA

Outputs of probabilistic sensitivity analysis are consistent with the deterministic base case (see Figure 9). If QALYs are valued at £20,000 each, there is a 99.9% probability that one of the referral strategies simulated represents better value for money than current practice.

If Braun (2013) '2-step' is included in the decision-space, there is a 42% probability that it is optimal at a threshold of £20,000/QALY. If it is omitted, the probability that Braun (2013) >=4 is optimal at that threshold is 39%.



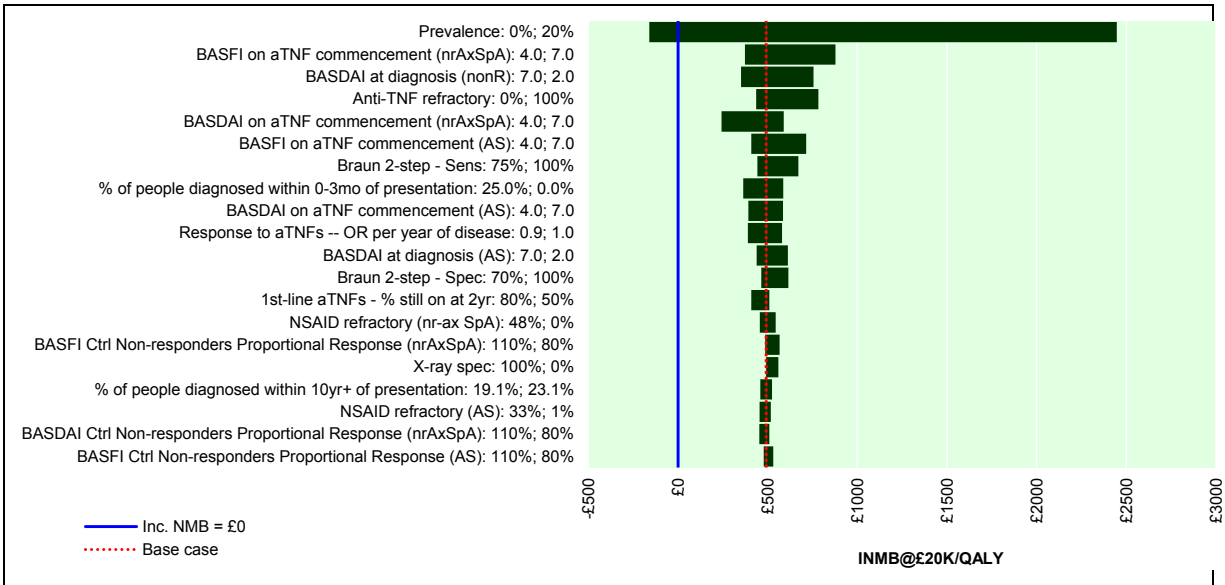
Bold line indicates cost-effectiveness acceptability frontier

Figure 9: Probabilistic sensitivity analysis – cost-effectiveness acceptability curve and frontier

H.4.1.4 Deterministic sensitivity analysis

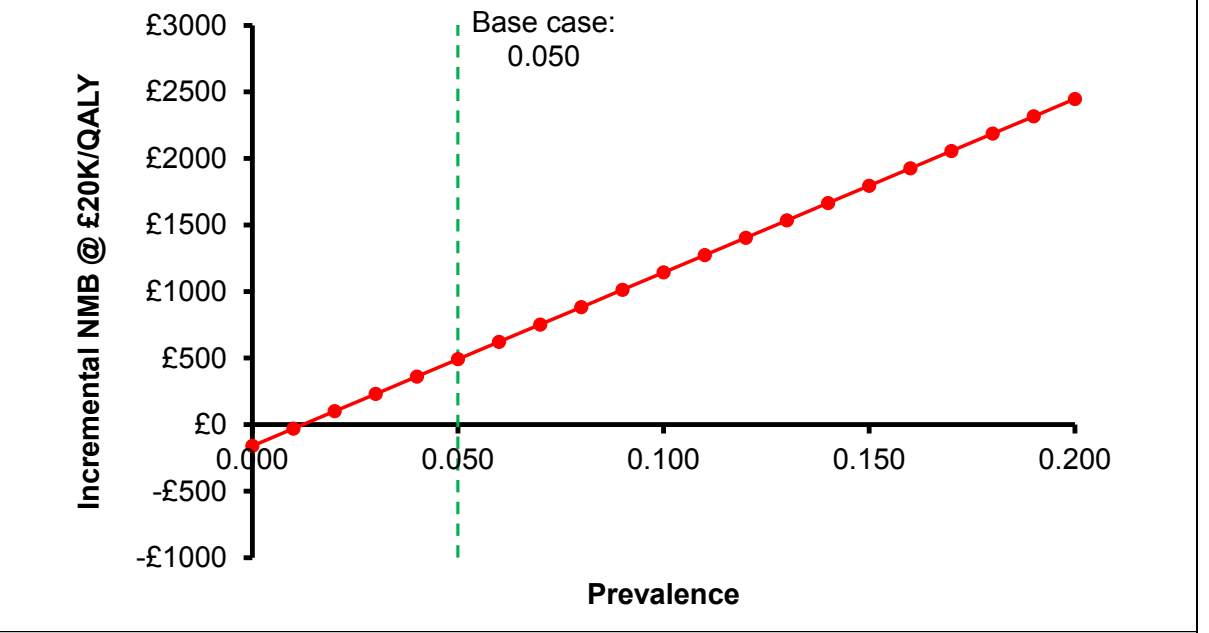
One-way sensitivity analysis shows that:

- Braun (2013) '2-step' would only represent poor value for money compared with current practice if true prevalence was 1.5% or lower (base case 5%). No other parameter alterations make current practice look optimal (see Figure 10 and Figure 11).



Values less than 0 indicate current practice provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step
20 most influential parameters only shown

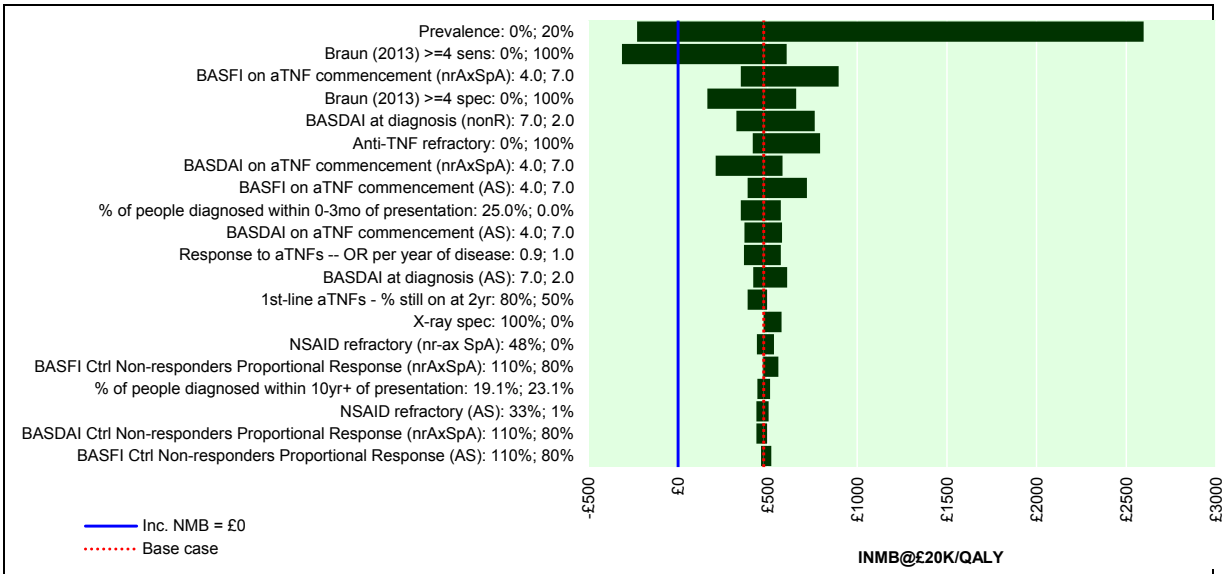
Figure 10: One-way sensitivity analysis – Braun (2013) 2-step -v- current practice



Values less than 0 indicate current practice provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step

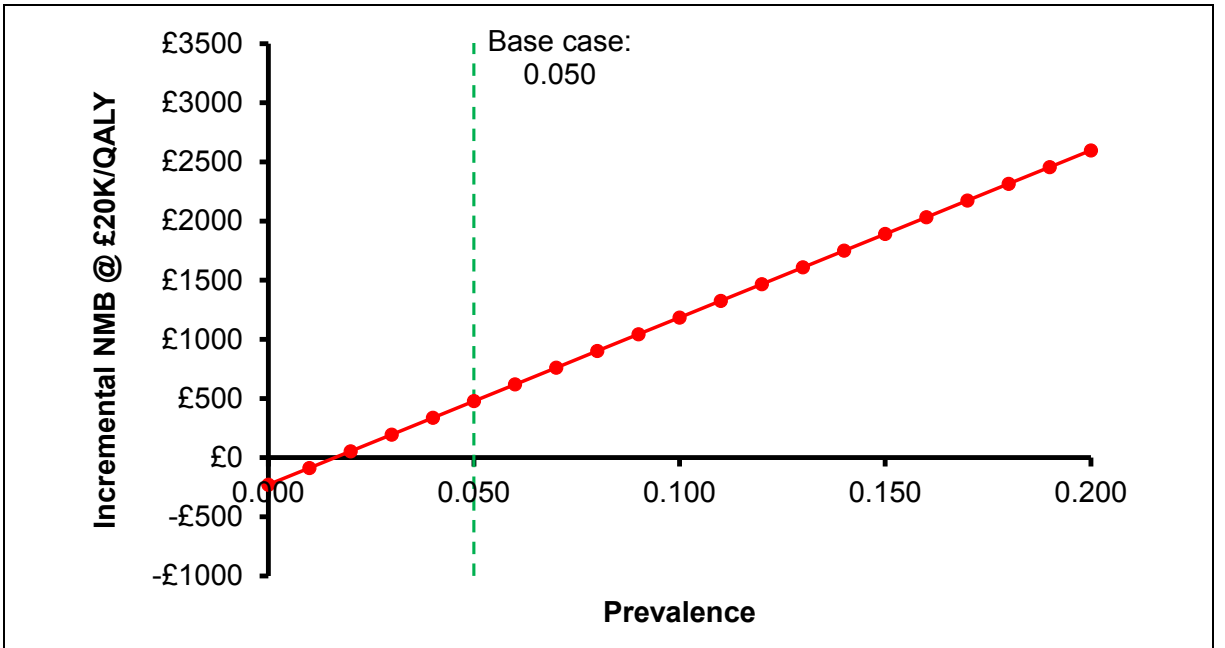
Figure 11: Threshold analysis – Braun (2013) 2-step -v- current practice: prevalence

- Similarly, Braun (2013) >=4 would only represent poor value for money compared with current practice if true prevalence was 1.5% or lower (base case 5%). No other parameter alterations make current practice look optimal (see Figure 12 and Figure 13).



Values less than 0 indicate current practice provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) ≥4
20 most influential parameters only shown

Figure 12: One-way sensitivity analysis – Braun (2013) ≥4 -v- 'current practice'

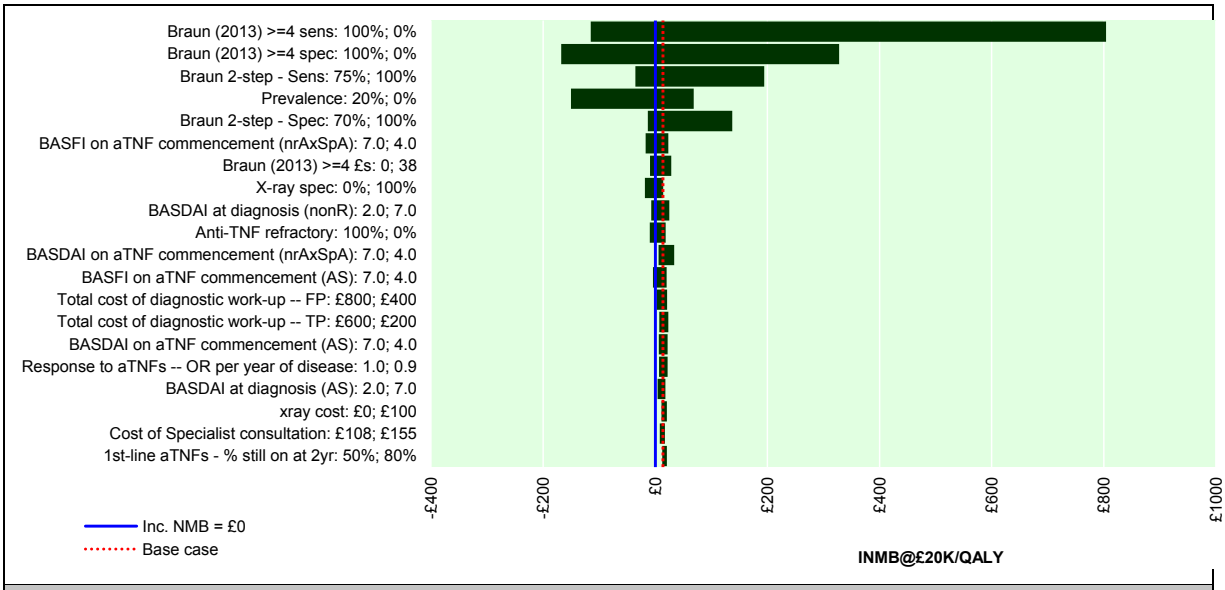


Values less than 0 indicate current practice provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) ≥4

Figure 13: Threshold analysis – Braun (2013) ≥4 -v- 'current practice': prevalence

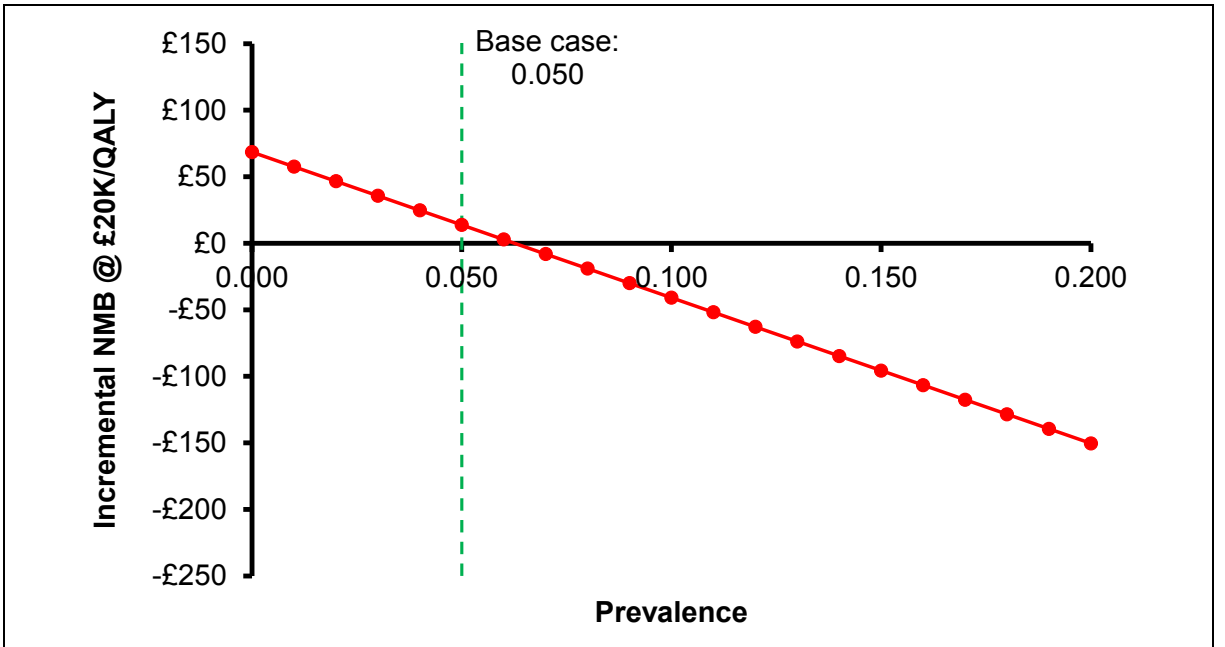
- Braun (2013) >=4 would be preferred to Braun (2013) '2-step' with plausible alterations to several parameters (Figure 14), including

- if true prevalence was above 6% (base case 5%; Figure 15)
- if sensitivity of Braun (2013) '2-step' was less than 79% (base case 80.4%; Figure 16)
- if specificity of Braun (2013) '2-step' was less than 72.5% (base case 75.4%; Figure 17)



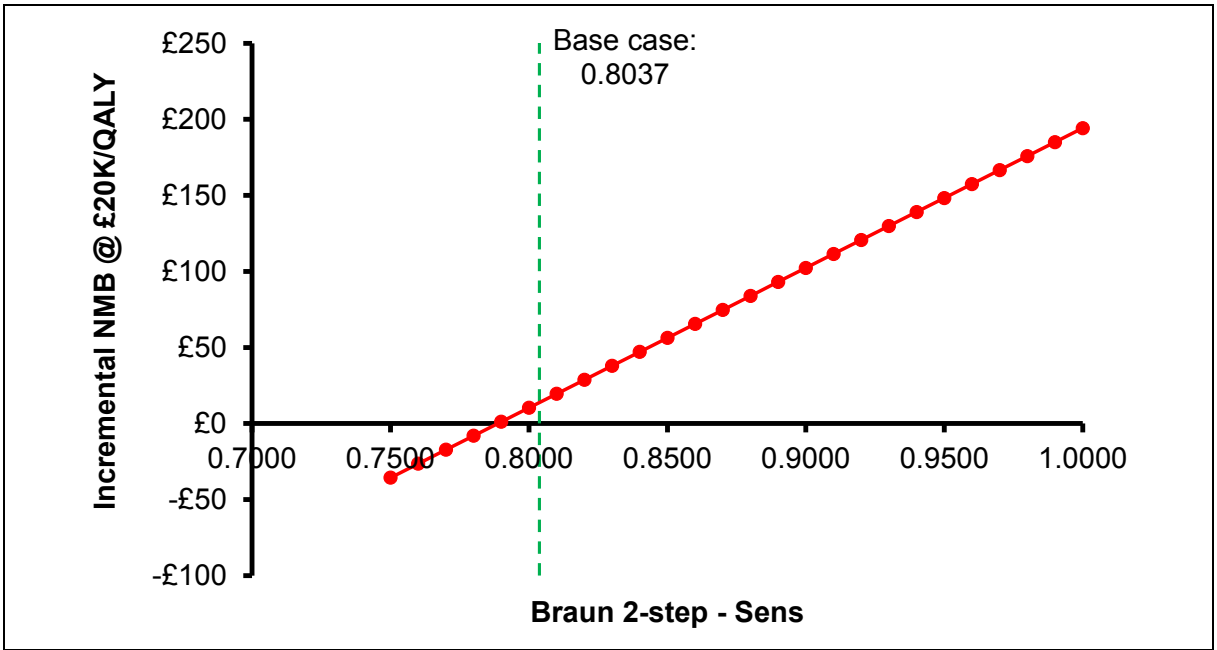
Values less than 0 indicate Braun (2013) ≥4 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step
20 most influential parameters only shown

Figure 14: One-way sensitivity analysis – Braun (2013) 2-step -v- Braun (2013) ≥4



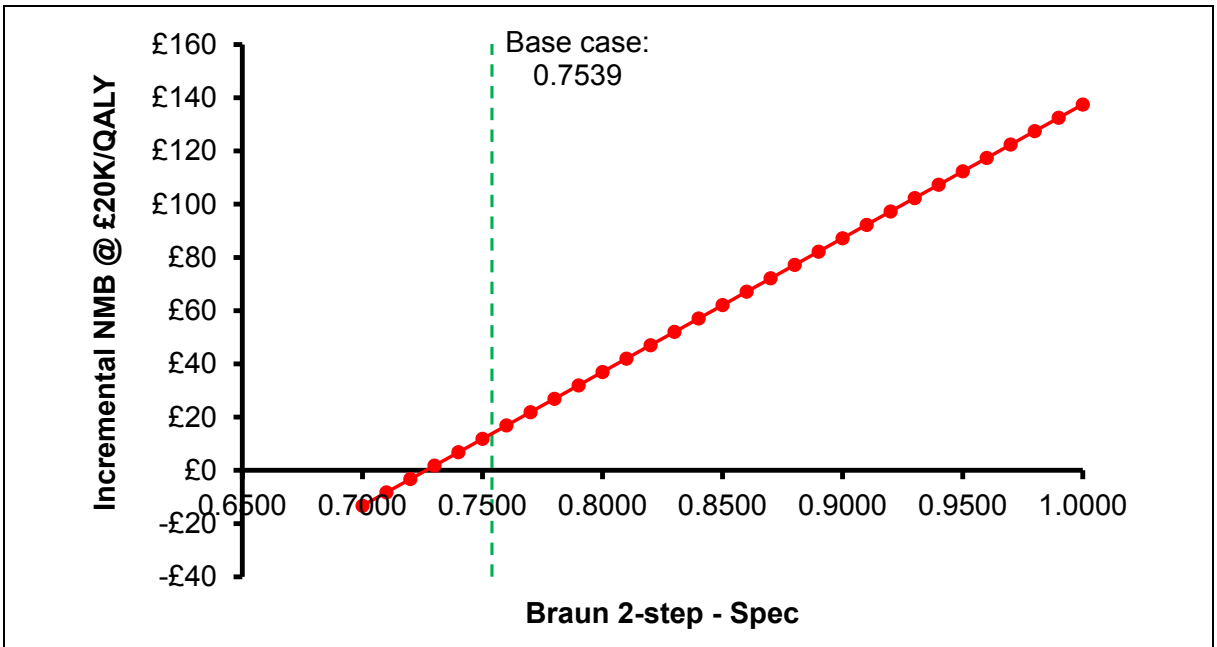
Values less than 0 indicate Braun (2013) ≥4 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step

Figure 15: Threshold analysis – Braun (2013) 2-step -v- Braun (2013) ≥4: prevalence



Values less than 0 indicate Braun (2013) ≥4 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step

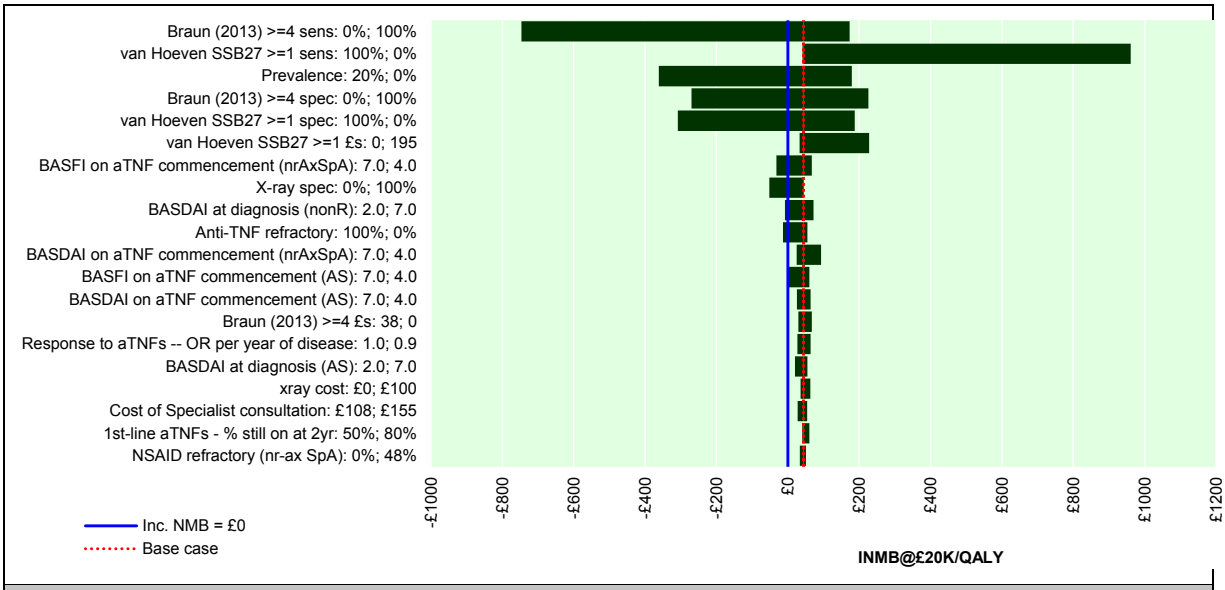
Figure 16: Threshold analysis – Braun (2013) 2-step -v- Braun (2013) ≥4: sensitivity of 2-step strategy



Values less than 0 indicate Braun (2013) ≥4 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step

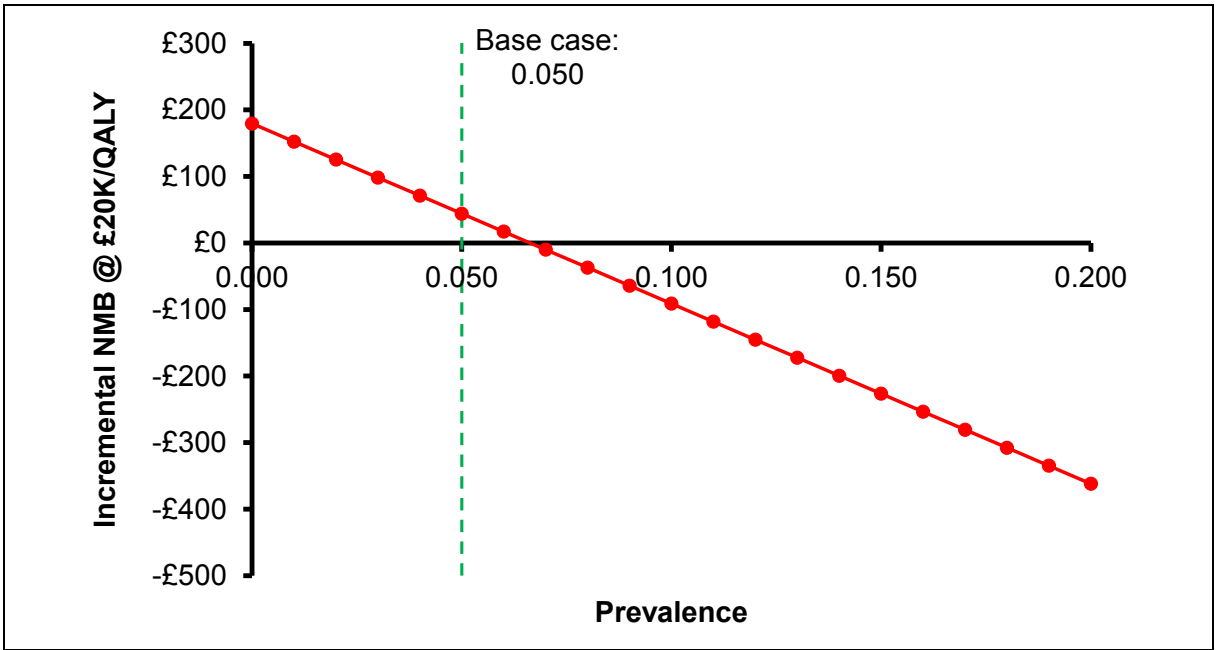
Figure 17: Threshold analysis – Braun (2013) 2-step -v- Braun (2013) ≥4: specificity of 2-step strategy

- Van Hoeven (2014) >=1 would be preferred to Braun (2013) >=4 with plausible alterations to several parameters (Figure 18), including
  - if true prevalence was above 6.5% (base case 5%; Figure 19)
  - if average BASFI scores were assumed to be as high as 7 at the start of biological DMARD therapy (base case 5.3 [AS] / 4.9 [nrAxSpA])



Values less than 0 indicate van Hoeven (SSB27) ≥1 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) ≥4
20 most influential parameters only shown

Figure 18: One-way sensitivity analysis – Braun (2013) ≥4 -v- van Hoeven (SSB27) ≥1



Values less than 0 indicate Braun (2013) ≥4 provides better value for money if QALYs are valued at £20,000; values greater than 0 favour Braun (2013) 2-step

Figure 19: Threshold analysis – Braun (2013) ≥4 -v- van Hoeven (SSB27) ≥1: prevalence

H.5 Discussion

H.5.1 Strengths of the analysis

This is the first economic evaluation of referral strategies for people presenting with signs, symptoms and risk factors suggestive of axial spondyloarthritis. This has been identified as a very significant problem, with an average delay to diagnosis of over 8 years.

The selection of an optimal referral strategy entails a complex interaction of benefits, costs and harms associated with false-positive and false-negative diagnoses. There is always a trade-off between sensitivity and specificity of potential approaches and no simple way to balance the relative desirability of gains in each. Our model provides an objective and robust way to make this assessment and, as such, we consider it a significant step forward that enables rational decision-making. We have also provided simple outputs that enable any future strategies to be assessed without the need to repeat the exercise.

It is a significant strength of our model that it was able to build on recent high-quality work by other researchers. Our simulation of anti-TNF treatment is substantially based on the work undertaken by York University in support of TA383 (Corbett et al. 2016). Our quality of life calculation relies on a new model that – in both its underlying data (over 1,500 observations) and its methods (moving beyond the linear models that have previously been used) – represents a significant improvement in the estimation of health utility associated with spondyloarthritis (Wailoo et al. 2015). We were able to incorporate both sources of evidence into our model in a fully probabilistic way that propagates uncertainty appropriately throughout the model.

H.5.2 Limitation of the analysis

The major weakness of our analysis is in its primary input data, estimating the true prevalence of spondyloarthritis and the accuracy of referral strategies.

As noted above, model outputs are critically dependent on the assumed prevalence of axial spondyloarthritis among people presenting with at least 3 months' low-back pain with age of onset under 45. This is a number that is not known and – because it will always rely on the quantification of occult disease that is, by definition, unknown – can never be conclusively established. We have sought best available data on this parameter and the GDG discussed it at length. We explored the impact of this uncertainty on our model, and we acknowledge that it is significant. However, we have constructed our analysis in such a way that, for as long as our other assumptions hold true, it can trivially incorporate any better estimate of prevalence that becomes available and provide revised results without needing to reconstruct or rerun the model in its entirety (see equation (8), above).

The accuracy of various referral strategies is also uncertain. Methodologically, the highest quality research available came from the CaFaSpA cohort (van Hoeven et al., 2014 & 2015) – a large cohort identified from primary care in which a final diagnosis was established for all participants, not just those who met some referral criteria. However, there are peculiarities in this evidence – above all, a perplexingly low prevalence of HLA-B27 positivity (lower than expected for the general population, never mind people at suspicion of spondyloarthritis). The GDG recommended that a similar piece of research should be undertaken in the UK, to replicate the methodological strength of the study while concentrating on a directly applicable target population. Again, the results of any such research could potentially be analysed using the simple outputs of our model, without recourse to further complex analysis.

In order to provide the most faithful possible simulation of the benefits and harms of case detection, it would be ideal to estimate the effect that diagnosis – leading, ultimately, to immunosuppressive anti-TNF therapy – has on people with serious latent infections. We were able to incorporate tuberculosis in our model due to the availability of recent NICE

modelling, but other infections – notably hepatitis C and HIV – that can be affected by biological therapy are also associated with significant harms and therapeutic costs. It would have been a huge undertaking to model these out in detail and we judged it was not possible to do this without constructing submodels of each disease process, which was not feasible. We think the impact of this factor is likely to be minor, as very few people experience these events – the parameters for the tuberculosis calculation made a negligible difference to our overall results. Nevertheless, we acknowledge that there is, here, a downstream harm of diagnosis that we have not been able to capture.

We initially hoped that we might be able to adapt our model to explore similar trade-offs in the diagnosis of peripheral disease. Unfortunately, this did not prove possible. However, the predominant aetiology of peripheral spondyloarthritis is associated with psoriasis, and most people have a known diagnosis of psoriasis before articular symptoms develop, so there is a much more well developed mechanism for monitoring and referring such people when they develop joint problems (see CG153).

H.5.3 Comparison with other economic evaluations

As this is the first evaluation of its type, there are no direct comparators against which to judge the convergent validity of our model. There are many cost–utility analyses of biological DMARDs for populations comprising people with ankylosing spondylitis and/or people with non-radiographic spondyloarthritis.

We compared the true-positive cases from our model with the outputs of models considered as part of technology appraisal 383 (one developed by York University [Corbett et al. 2016] and manufacturer models submitted by MSD, Pfizer and AbbVie). We found that, in the ankylosing spondylitis population, our model produces somewhat lower costs than the TA models – with the exception of the Pfizer submission, which has a conspicuously low estimate of overall costs, total lifetime discounted costs of £130,000–£210,000 are projected, whereas our model estimates a value of £120,000. This is to be expected, because a proportion of people in our model never receive the most expensive interventions (biological DMARDs), and those who do receive them do so on average a few years into the model (so those costs are discounted to a greater degree than in the TA). Our model estimates somewhat higher lifetime QALYs for people with ankylosing spondylitis than the TA models – 14.6 compared with 8.2–11.7. Again, we think this is at least partially predictable: our cohort starts at a younger age than the population starting anti-TNFs: we simulate an average of 6 years' pre-biological therapy, during which time people accrue around 4.3 discounted QALYs, which is enough to account for the difference on its own. Another potential source of differences is that our analysis benefits from a sophisticated model of health utility that was published after the TA (Wailoo et al. 2015; see H.2.5). It is not immediately obvious whether this model would lead to higher or lower QALY estimates, though it clearly represents a step forward in the estimation of quality of life for people with spondyloarthritis, so we are confident that, one way or another, our estimates are more accurate than those available to the TA developers

Similar differences are apparent when our model's non-radiographic spondyloarthritis true positives are compared with those in the TA models.

On balance, we find this exercise provides a reasonable degree of validation of our findings.

H.5.4 Conclusions

Our model shows unambiguously that many different approaches to referral for suspected axial spondyloarthritis would provide QALY gains at reasonable cost compared with current practice. It is somewhat more difficult to be sure of which strategy is optimal, owing to uncertainties in the inputs – above all, the true prevalence of disease in the presenting population. If it is considered credible that the reported results of the Braun (2013) '2-step'

algorithm can be replicated in NHS practice, then it is likely to be judged the optimal strategy. If this approach is excluded from the decision space, a strategy in which people are referred if they meet 4 or more criteria from Braun et al.'s list of 10 would be considered to represent the best balance of costs and benefits. Future evidence on accuracy of referral strategies and/or the true prevalence of axial spondyloarthritis can be used to re-estimate outputs using equations provided without the need to rerun the model.

H.6 References

Atkinson MD, Brophy S, Siebert S, Gravenor MB, Phillips C, Ford DV, Jones KH, Lyons RA. Protocol for a population-based Ankylosing Spondylitis (PAS) cohort in Wales. BMC musculoskeletal disorders. 2010 Sep 1;11(1):1.

Azevedo VF, Paiva ED, Felippe LR, Moreira RA. Occurrence of fibromyalgia in patients with ankylosing spondylitis. Revista brasileira de reumatologia. 2010 Dec;50(6):646-50.

Boonen A, Van der Heijde D, Landewe R, Guillemin F, Rutten-van Mölken M, Dougados M, Mielants H, De Vlam K, Van Der Tempel H, Boesen S, Spoorenberg A. Direct costs of ankylosing spondylitis and its determinants: an analysis among three European countries. Annals of the rheumatic diseases. 2003 Aug 1;62(8):732-40.

Corbett M, Soares M, Jhuti G, Rice S, Spackman E, Sideris E, Moe-Byrne T, Fox D, Marzo-Ortega H, Kay L, Woolacott N. Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. Health Technology Assessment. 2016 Feb 1;20(9):333.

Dougados M, Béhier JM, Jolchine I, Calin A, van der Heijde D, Olivieri I, Zeidler H, Herman H. 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 & Rheumatism. 2001 Jan 1;44(1):180-5.

Glintborg B, Østergaard M, Krogh NS, Tarp U, Manilo N, Loft AG, Hansen A, Schlemmer A, Fana V, Lindegaard HM, Nordin H. Clinical response, drug survival and predictors thereof in 432 ankylosing spondylitis patients after switching tumour necrosis factor α inhibitor therapy: results from the Danish nationwide DANBIO registry. Annals of the rheumatic diseases. 2013 Jul;72(7):1149.

Kroon FP, van der Burg LR, Ramiro S, Landewé R, Buchbinder R, Falzon L, van der Heijde D. Non-steroidal anti-inflammatory drugs (NSAIDs) for axial spondyloarthritis (ankylosing spondylitis and non-radiographic axial spondyloarthritis). The Cochrane Library. 2015 Jan 1.

Rudwaleit M, Listing J, Brandt J, Braun J, Sieper J. Prediction of a major clinical response (BASDAI 50) to tumour necrosis factor α blockers in ankylosing spondylitis. Annals of the rheumatic diseases. 2004 Jun 1;63(6):665-70.

Rudwaleit M, Haibel H, Baraliakos X, Listing J, Märker-Hermann E, Zeidler H, Braun J, Sieper J. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis & Rheumatism. 2009 Mar 1;60(3):717-27.

Sieper J, Klopsch T, Richter M, Kapelle A, Rudwaleit M, Schwank S, Regourd E, May M. 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. Annals of the rheumatic diseases. 2008 Mar 1;67(3):323-9.

van der Heijde D, Baraf HS, Ramos-Remus C, Calin A, Weaver AL, Schiff M, James M, Markind JE, Reicin AS, Melian A, Dougados M. Evaluation of the efficacy of etoricoxib in

ankylosing spondylitis: results of a fifty-two–week, randomized, controlled study. Arthritis & Rheumatism. 2005 Apr 1;52(4):1205-15.

Wailoo A, Hernández M, Philips C, Brophy S, Siebert S. Modeling health state utility values in ankylosing spondylitis: comparisons of direct and indirect methods. Value in Health. 2015 Jun 30;18(4):425-31.

Whitehurst DG, Bryan S, Lewis M, Hill J, Hay EM. Exploring the cost-utility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. Annals of the rheumatic diseases. 2012 Nov;71(11):1796.

H.7 Economic evidence tables

H.7.1 Switching or augmenting pharmacological interventions for spondyloarthritis

Study, Population, Country and Quality	Data Sources	Other Comments	Incremental			Conclusions	Uncertainty
			Cost	Effect	ICER		
Coates et. al. (2015) RCT of early intervention and tight control of inflammation in patients with early psoriatic arthritis. Tight control vs standard care Directly applicable[a] Very serious limitations[b,c,d,e]	Effects: Primary outcome is the proportion of individuals achieving an American College of Rheumatology 20% response (ACR20). Costs: Within trial resource use measured and costs allocated from national sources. Utilities: Patient completed EQ-5D.	No extrapolation beyond study end-point. Missing costs and EQ-5D data imputed. Funded by Arthritis UK & Pfizer.	48 weeks £2,198	48 weeks 0.042 QALYs	48 weeks £53,948/ QALY	Tight control was not cost-effective when compared to usual care.	In PSA tight control was cost-effective in 7% of iterations compared with standard care (at £20,000/ QALY threshold) Scenario analysis: Total cost reduction of 25% in both tight control and standard care groups & reduction in 3-monthly follow-up visits for individuals achieving the minimum level of disease activity at 2 consecutive consultations. ICER approaches cost-per-QALY threshold however not cost-effective. [£30,632].
RCT: randomised controlled trial			ICER: incremental cost-effectiveness ratio QALY: quality adjusted life year				

a The use of anti-TNFs is included within both strategies for which recommendations on their use in psoriatic arthritis is beyond the scope of this guideline. The protocol however, states to follow NICE guidance on their use for both arms and enables quantification of the impact of faster treatment escalation.

b The time horizon does not capture all relevant benefits and costs as is restricted to the 48 week data collection period.

c Details of parameter uncertainty used in PSA not reported, 25% reduction in costs not justified in detail.

d Multiple imputation used to estimate missing cost and utility values however details of the methods not reported.

e Potential conflict of interest