National Institute for Health and Care Excellence

Final

Osteoarthritis in over 16s: diagnosis and management

Cost-utility analysis: Electroacupuncture for the management of osteoarthritis

NICE guideline NG226

Economic analysis report

October 2022

Final

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-4740-9

Contents

1 Introduction

A systematic review of the published clinical and economic evidence was undertaken wherein acupuncture was compared with both sham acupuncture and usual care. The clinical evidence showed an unclear clinical benefit with acupuncture compared to no treatment and no clinically difference in quality of life, pain and physical function compared with sham acupuncture. Electroacupuncture showed a short-term clinically important benefit in pain and physical function compared with no treatment and a clinically important benefit in the short term for pain compared with sham acupuncture.

There were four economic evaluations identified during the review, one taking a German perspective and the other three taking a UK perspective (see section 1.1.7 of the evidence review). All compared acupuncture to usual care. The German study was based on a 3-month trial with costs and outcomes extrapolated from 3 months to 12 months. The three UK-based studies had time horizons that ranged from 8 weeks to 12 months. One study was based on a pooled estimate of effects from a network meta-analysis, one study was based on a single trial and the final study was an analysis of three separate trials.

The results from all four evaluations were consistent with acupuncture showing cost effectiveness compared to usual care at a cost per QALY threshold of £20,000 (range across studies varied between £3,889 and £17,381).

Nonetheless, none of the evaluations assessed electroacupuncture on its own versus usual care, and the committee were interested in exploring this question further. For this reason, this question was prioritised for new economic modelling.

2 Methods

2.1 Model overview

A cost-utility analysis was undertaken where quality-adjusted life years (QALYs) and costs over 1 year from a current UK NHS and personal social services perspective were considered. The analysis followed the standard assumptions of the NICE reference case for interventions with health outcomes in an NHS setting, however discounting for costs and health effects was not required as the time horizon was only 1 year.[20] An incremental analysis was undertaken.

2.1.1 Comparators

The committee discussed the correct comparator for evaluations of electroacupuncture. In economic evaluations, comparisons with no treatment or usual care or an alternative active comparator are usually considered most relevant for assessing the real-world impact of an intervention on resource use and QALYs. It was decided that for electroacupuncture to be recommended there should be evidence of:

- a clinical benefit compared with both sham and usual care, and
- cost effectiveness compared with usual care.

Comparing electroacupuncture to usual care is the most common approach to assess its cost-effectiveness and this approach has been taken on the NICE guidelines on low back pain (NG59) and primary chronic pain (NG193), which both assessed acupuncture.

The following comparators were therefore included in the analysis:

1. Electroacupuncture
2. Usual care

It was assumed that both groups otherwise received the same care. Electroacupuncture is an adjunctive treatment to usual care and does not replace any treatment options, therefore it was appropriate to compare it to usual care in the model despite the fact that the majority of trials compared electroacupuncture to an active treatment.

The effectiveness data came from studies that compared electroacupuncture with usual care.

2.1.2 Population

The population of the analysis was adults aged 16 and over with osteoarthritis (OA) of the knee.

2.2 Approach to modelling

2.2.1 Model structure

A simple model was used to simulate costs and QALYs over a 52-week time horizon, which was deemed sufficiently long to capture the treatment effects. An area under the curve analysis was considered sufficient to capture differences in quality of life between electroacupuncture and usual care.

Mortality is not impacted by treatment and given the short time horizon of the model, so it does not need to be modelled. Differences in QALYs between electroacupuncture and usual care in the model would therefore be driven by differences in quality of life alone. The model was run for 12 cycles, each cycle representing 1 month of life. There were no serious adverse events associated with treatment modelled since none were reported in the clinical review. A comparison between the results of electroacupuncture and usual care allowed us to identify the most cost-effective strategy. More details on the model structure are described in section 2.2.1. To account for uncertainty, a probabilistic analysis was undertaken (see section 2.2.2 for further details).

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline. The clinical evidence reported that electroacupuncture showed benefits for pain and physical function when compared to usual care and improved quality of life compared to sham acupuncture. In economic evaluation, a particular measure of QoL is required known as a utility. The analysis is therefore based on studies from the clinical review that reported health outcomes data via either the WOMAC score, SF-12 or the SF-36, which were subsequently mapped to EQ-5D utility scores (see section 0 for more detail). The available data on the difference in utility between electroacupuncture and usual care were combined with assumptions about what was likely to happen to treatment effect beyond the follow-up in the trials (follow-up times of the trials are detailed further down in Table 2), to calculate the average QALY gain with electroacupuncture compared to usual care. This is described in detail in section 2.3.3. An alternate base case did not extrapolate beyond the trial data.

The key difference in costs was agreed to be those related to delivering an electroacupuncture programme. No other costs were incorporated in the analysis. The average resource use from the interventions in each study was identified and costed, and an overall weighted average cost calculated, weighting by the number of participants analysed in each study. This is described in detail in section 2.3.4.

Costs and QALYs were combined to derive the overall cost effectiveness of electroacupuncture in an OA of the knee population.

Pooling acupuncture studies

It was acknowledged that the intervention was delivered differently across different studies both in terms of frequency and duration of treatment and this may have different costs, and it was agreed that there would be two separate base cases; one using pooled costs of the interventions in the clinical studies along with the pooled treatment effects and another using separate trial data, thereby evaluating separate cost per QALYs from each trial. This approach was taken as the committee were interested in the effect of treatment frequency and duration on the cost per QALY.

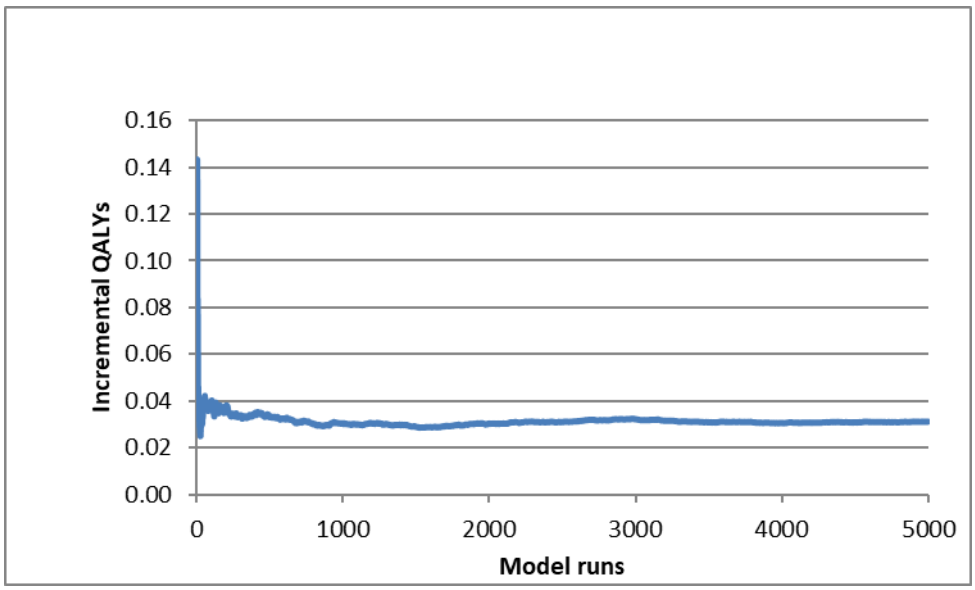
2.2.2 Uncertainty

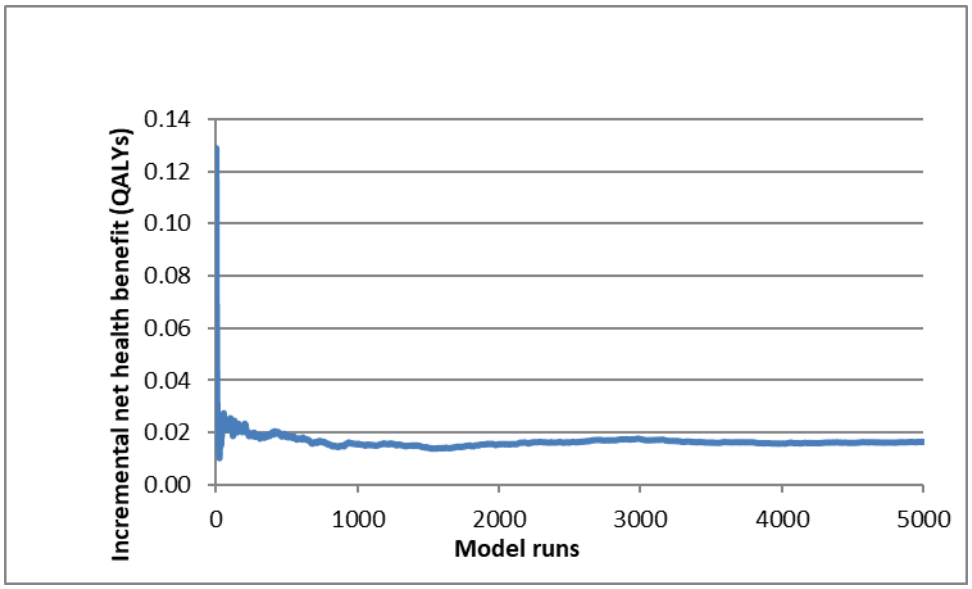
The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for certain model input parameters. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean costs and mean QALYs were calculated using these values. The model was run repeatedly – 5,000 times for the base case and for each sensitivity analysis – and results were summarised.

When running the probabilistic analysis, multiple runs are required to take into account random variation in sampling. To ensure the number of model runs were sufficient in the probabilistic analysis we checked for convergence in the incremental QALYs and net health benefit at a threshold of £20,000 per QALY gained for electroacupuncture versus usual care. This was done by plotting the number of runs against the mean outcome at that point (see

example in Figure 1) for the base-case analysis. Convergence was assessed visually, and all had stabilised before 5000 runs.

Figure 1: Checking for convergence: incremental QALYs and incremental net health benefit (electroacupuncture vs usual care)

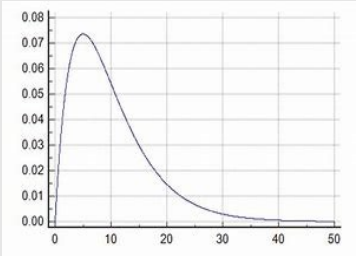
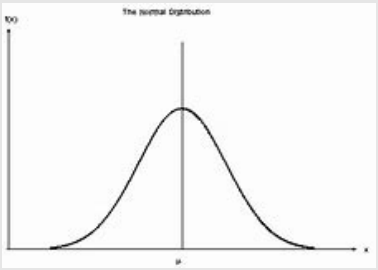




The way in which distributions are defined reflects the nature of the data. The variables that were probabilistic in the model and their distributional parameters are detailed in

Table 1. Probability distributions in the analysis were parameterised using error estimates from data sources.

Table 1: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
Quality of life scores (e.g., PCS and MCS from the SF-12 and SF-36, WOMAC score) where a tobit or mixture model were used to map to EQ-5D.	Gamma 	Bounded at 0, positively skewed. Derived from mean and its standard error. Alpha and beta values were calculated as follows: • Alpha = $(mean/SE)^2$ • Beta = $SE^2/Mean$ Note: SE determined based on the standard deviation across the studies.
Regression coefficients to map from SF-12, SF-36 or WOMAC score to EQ-5D.	Normal 	The normal distribution is symmetric. Derived from mean and its standard error.
Utilities, specifically for EQ-5D scores where an OLS regression model were used to map from SF-12, Sf-36 or WOMAC scores to EQ-5D.	Beta	Bounded between 0 and 1. Derived from mean and its standard error, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = $mean^2 \times [(1-mean)/SE^2] - mean$ Beta = $alpha \times [(1-mean)/mean]$

Abbreviations: EQ=5D= EuroQol 5 dimensions; MCS= mental component score; OLS= ordinary least squares; PCS= physical component score; SE = standard error; SF-12= short from health survey, 12 items; SF-36= short form health survey, 36 items; WOMAC= Western Ontario and McMaster Universities Osteoarthritis Index

The following variables were left deterministic (that is, they were not varied in the probabilistic analysis):

- the cost-effectiveness threshold (which was deemed to be fixed by NICE),
- the resources, including time and cost of staff, required to implement acupuncture from each study,
- drug prices (based on drug tariff which is known).

In addition, various deterministic and probabilistic sensitivity analyses were undertaken to test the robustness of model assumptions. In these, one or more inputs were changed, and the analysis rerun to evaluate the impact on results and whether conclusions on which intervention should be recommended would change. Details of the sensitivity analyses undertaken can be found in methods section 2.5 Sensitivity analyses.

2.3 Model inputs

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline (see Appendix E.5 of the evidence review), supplemented by additional data sources as required. Model inputs were validated with clinical members of the guideline committee. More details about sources, calculations and rationale for selection can be found in the sections below.

2.3.1 Clinical studies used in analysis

The analysis is based on studies from the clinical review that reported utilities (EQ-5D), or SF-36 that could be mapped to EQ-5D, or other scales that could be mapped to EQ-5D. Where a study reported more than one type of outcome, then the following hierarchy was used: EQ-5D, mapped SF-36, then other mapped scales. The basis for this being that direct measurement of utilities was preferred over mapped measures, and where mapping was the only option then mapping SF-36 was preferred over mapping from other scales, as the SF-36 is more established and more widely used.

11 studies were included in the clinical review. Of these, 4 were with usual care comparisons. Of these 4: 1 study reported SF-12 summary scores, 1 study reported SF-36 summary scores (both of which could be mapped to the EQ-5D-3L); and 2 studies reported WOMAC scores that could also be mapped to the EQ-5D-3L.

The 4 studies are summarised in Table 2 below.

Table 2: Summary of studies

Study	Number of sessions (per week)	Length of sessions (minutes)	Duration of sessions (weeks)	Voltage	Population	Number of participants	Comparator	Outcome indicators used	Timepoints health outcomes measured
Berman 1999[4]	Two	20	8	2.5-4.0 Hz	Knee OA	73	Standard care: oral therapy	WOMAC total, pain and function scores	Baseline, 4, 8 and 12 weeks
Dunning 2018[8]	One to two (up to 10 sessions in total)	20-30	6	2.0 Hz	Knee OA	222	Standard care: manual therapy and exercise	WOMAC total, pain, stiffness and function scores	Baseline, 2, 6 and 12 weeks
Mavrommatis 2012[18]	Two	20	8	2.0 - 6.0 Hz	Knee OA	80	Pharmacological therapy	WOMAC total, pain, stiffness and function scores, SF-36 (PCS, MCS) [(a)]	WOMAC: Baseline, 4, 8 and 12 weeks. SF-36: Baseline, 8 weeks
Suarez-Almazor 2010[26]	Two	30	6	15.0 Hz	Knee OA	225	Waiting list (study did not specify if this group received any other form of therapy)	WOMAC pain and function, SF-12 (PCS, MCS) [(b)]	Baseline, 4, 6 and 12 weeks

Abbreviations: Hz= Hertz; MCS= mental component score; OA= osteoarthritis; PCS= physical component score; SF-12= short-form survey 12 items; SF-36= short-form survey 36 items; WOMAC= Western Ontario and McMasters Universities Osteoarthritis Index

(a) WOMAC scores were not reported using the scale whose range is between 0-92 and therefore could not be mapped to EQ-5D. SF-36 PCS and MCS scores were used instead for analysis.

(b) Neither the total WOMAC not the WOMAC stiffness scores were reported resulting in an incomplete set of WOMAC values. Subsequently, WOMAC scores to EQ-5D mapping algorithms could not be utilised. SF-12 PCS and MCS scores were used instead for analysis.

There was heterogeneity observed between the studies in terms of the type of electroacupuncture device used, the voltage given as well as the pressure points selected for needling. Presenting results according to individual trials does address this limitation, however two of the trials reported small patient numbers: N=73 in Berman 1999 and N=80 in Mavrommatis 2012. Therefore, results in the base case will be based on individual trials as well as pooled estimates of effect using both a weighted average and an unweighted average.

2.3.2 Calculating the difference in QALYs

For studies that reported SF-12 and SF-36 data, the mean for each summary score along with the standard deviation were extracted for the baseline and any subsequent time points, for both the intervention and control groups.

For studies that reported WOMAC scores, both the total WOMAC score and the subscale scores along with the standard deviations were extracted for the baseline and any subsequent time points, for both the intervention and control groups. One study did not report the WOMAC stiffness subscale score[4], so it was instead calculated by subtracting the other subscale scores from the total WOMAC score. The standard deviation was assumed to be 20% of the subscale score.

Mapping SF-12/36 to EQ-5D

An algorithm by Price 2019 was chosen in the base case to convert from SF-36 physical and mental component scores (PCS and MCS) to EQ-5D.[23] It focuses on an OA population and uses a UK population tariff for EQ-5D scores, thereby meeting the NICE reference case. It has other advantages too, such as many observations underpinning the results (N=19,410 observations from 2,201 individuals) as well as the availability of a co-variance matrix, which is useful when making its algorithm probabilistic. The advantage of a co-variance matrix is that it enables interrelation between individual variables.

A scenario analysis will use an algorithm by Lawrence 2004.[13] Although this algorithm uses a general US population tariff to derive EQ-5D values, it also has many observations underpinning the results (14,580 individuals, of which 7,313 were selected randomly for the analytic sample and 7,267 were reserved for validation of the mapped scores) as well as a co-variance matrix. Three regression models were used for mapping. Although the 6-variable model had the best reported goodness of fit (R^2=0.628) and predictive ability (Mallow's C_p= 22.1) of the three, the 2- and 3-variable models performed better in predicting the EQ-5D scores across a range of disease areas. Since predicted scores were virtually identical between the 2- and 3- variable models, the authors used the 2-variable model for subsequent analysis, and this was the only model for which the variance-covariance matrix was published. For these reasons, the 2-variable model was used for mapping.

A second scenario analysis was chosen which used an algorithm by Maund 2012.[17] This was a systematic review and cost-effectiveness analysis, that derived QoL needed for the cost utility analysis by creating a regression to map from the SF-36 summary scores to the EQ-5D UK tariff. The dataset used to generate the regression was the SAPPHIRE trial (2008),[28] which was a trial in a UK population with rotator cuff disease (N=200). The algorithm was based on a regression model using individual-level data at 1, 3 and 12 months. This dataset was preferred to the 3-month dataset as the explanatory power and fit was better. There were five models to choose from, of which, 3 were ordinary least squares (OLS) models, one was a tobit model and one was a censored least absolute deviation (CLAD) model. Of the OLS models, model 3 had the highest explanatory fit (adjusted R^2=0.4284) as well as the closest predicted EQ-5D score to the actual EQ-5D score and was therefore chosen for mapping.

It should be noted that although the Price and Lawrence algorithms were intended to convert SF-12 summary scores to EQ-5D, it is possible to map from either the SF-12 or SF-36 summary scores to EQ-5D-3L since both utilise the same summary scores on the same scale.

Mapping WOMAC to EQ-5D

WOMAC scores were mapped to the EQ-5D-3L (UK tariff) using the regression model from Wailoo 2014 in the base case[27], while a scenario analysis explored the effect of using regression model E from Barton 2008.[1] In Barton 2008, five mapping models were developed with each model taking on different baseline WOMAC scores as the independent variable(s). For example, model A used the total WOMAC score, model B used WOMAC pain, stiffness, and functioning scores etc. Model E was chosen since it reported the lowest mean absolute error (MAE) and the highest adjusted R^2. The Wailoo 2014[27] model was based on 7,072 observations from 1,768 patients recruited in a registry study from 15 hospitals across Spain who were either scheduled to undergo primary joint replacement surgery due to knee/hip OA or had received postoperative management. Of the available models, the five-class mixture model was preferred due to its superior summary measures of fit (mean absolute error, root mean squared error, Akaike information criterion and Bayesian information criterion). This model used the distribution of the EQ-5D UK value set to predict EQ-5D as a function of the WOMAC subscale scores.

The study by Barton 2008 mapped total WOMAC scores to EQ-5D-3L UK tariff using responses from individuals taking part in the Lifestyle Interventions for Knee Pain (LIKP) study. Inclusion criteria for the LIKP study were knee pain on most days over the past month, age greater than or equal to 45 and a BMI greater than 28 kg/m^2. The EQ-5D and WOMAC scores were completed at baseline by 348 individuals and 259 individuals further completed responses at 6,12 and 24 months. Five models were developed, of which model E had the highest adjusted R^2 (0.313) and the lowest mean absolute error and root mean squared error (0.129 and 0.180, respectively) and was therefore selected for mapping purposes.

A second scenario analysis will therefore be run using an algorithm by Price 2019[23] mapping total WOMAC scores to assess the impact of the above-mentioned algorithms on the cost effectiveness results. This algorithm is based on a trial on patients with chronic pain of the knee similar to OA in the UK (N=261) and uses a UK tariff for conversions to EQ-5D. However, this algorithm has not been externally validated.

Accounting for uncertainty in the regression weights

The coefficients in the mapping algorithms were themselves made probabilistic to account for uncertainties in the mapping equations. Various methods were used but they all included drawing from a normal distribution.

Standard errors for the coefficients were reported with the Maund algorithm, so these were used to make the values probabilistic. They were not reported with the Barton algorithm, so were assumed to be 20% of the deterministic point estimates. The Lawrence, Price SF-12 and Price WOMAC algorithms provided variance/covariance matrices, and the Cholesky decomposition method was utilised to make the point estimates probabilistic.

Finally, the Wailoo algorithm listed p values for all the model coefficients, and these were used to calculate the standard errors for each coefficient.

Accounting for uncertainty in mapping algorithms

Some publications have suggested that there is a problem with underestimation of uncertainty of utilities derived from mapping algorithms.[7,2,14] The most obvious explanation for the variance underestimation of derived utilities is that there are important unmeasured predictors in most mapping algorithms. This leads to a relatively high degree of unexplained variance of utilities. In OLS based mapping algorithms, this is reflected as a relatively low R squared.[6]

There were three OLS based mapping algorithms used during the analysis and a high level of unexplained variation was reported in all (that is, a relatively low R squared). To account for this source of uncertainty in the mapping process, an additional variance component was included in the EQ-5D predictions.

Chan 2014 [6] suggests methods that could be used to estimate the variance of mapped values, by accounting for a low R squared in OLS-based mapping algorithms. Multiple methods are suggested, but some are only possible if patient-level data is available. One simple method that could be used to account for an artificially low variance of utilities because of a low R squared, is to inflate the variance of the derived utilities by a factor of 1/R squared. This estimator helps account for a low R squared but does not account for the uncertainty of the regression coefficients. This adjustment has also been used in other studies using a mapping algorithm for pain.[16]

This adjustment factor was applied to the variance of the mapped EQ-5D values for both utilities mapped from WOMAC using the Barton algorithm (adjusted R squared = 0.313), and utilities mapped from the SF-12/SF-36 using the Lawrence algorithm (R squared = 0.612) and Maund algorithm (adjusted R squared = 0.428). See Appendix B for details of the variance before and after the adjustment was made. Where adjusted R squared were reported, these were converted to R squared by rearranging the formula:

Adjusted R squared $= \frac{(1\text{-}R^2)(n\text{-}1)}{n-k-1}$	Where: *n*=number of points in data sample *K*=number of independent regressors

The resulting EQ-5D scores (adjusted at baseline to electroacupuncture) to be used in the base case are presented in Table 3 below. Unadjusted EQ-5D scores are presented in Appendix A:

Table 3: EQ-5D-3L mapped over time by randomised trial

Trial	Timeframe (weeks)	EQ-5D (usual care)	EQ-5D (electroacupuncture)[(a)]	Incremental change in EQ-5D	Mapping algorithm used
Berman 1999[4]	0	0.468	0.468	0.000	Wailoo 2014
	4	0.486	0.638	0.152	
	8	0.485	0.678	0.193	
	12	0.483	0.653	0.171	
Dunning 2018[8]	0	0.553	0.553	0.000	Wailoo 2014
	2	0.629	0.701	0.072	
	6	0.673	0.780	0.107	
	12	0.668	0.815	0.147	
Mavrommatis 2012[18]	0	0.298	0.298	0.000	Price 2019
	8	0.471	0.634	0.163	
Suarez-Almazor 2010[26]	0	0.463	0.463	0.000	Price 2019
	4	NR	0.532	NR	
	6	NR	0.716	NR	
	12	0.452	0.550	0.098	

NR= not reported, a linear change in EQ-5D was assumed and calculated between week 0 and week 12
(a) Adjusted for differences at baseline

A second base case analysis utilised pooled utility scores from all four trials. This was done by first calculating the utilities associated with each quarter of the yearly time horizon in each individual and then calculating the quality adjusted life years by taking an average of all four quarters. There were two different pooled utility scores used in the model:

1. A weighted average calculated by including the number of participants in each trial into the calculation
2. An unweighted average that took the simple average for the quality adjusted life year of all four trials

The quarterly utility scores for individual trials are presented in Table 4 below, while the quality adjusted life years for individual trials and pooled trials are presented in Table 5.

Table 4. Quarterly utility scores for individual trials

Trial	Timeframe (quarterly years)	EQ-5D (usual care)	EQ-5D (electroacupuncture)[a]	Incremental change in EQ-5D	Mapping algorithm used
Berman 1999[4]	Q1	0.482	0.626	0.144	Wailoo 2014
	Q2	0.475	0.561	0.085	
	Q3	0.468	0.468	0.000	
	Q4	0.468	0.468	0.000	
Dunning 2018[8]	Q1	0.651	0.750	0.099	Wailoo 2014
	Q2	0.610	0.684	0.074	
	Q3	0.553	0.553	0.000	
	Q4	0.553	0.553	0.000	
Mavrommatis 2012[18]	Q1	0.413	0.522	0.109	Price 2019
	Q2	0.384	0.466	0.082	
	Q3	0.298	0.298	0.000	
	Q4	0.298	0.298	0.000	
Suarez-Almazor 2010[26]	Q1	0.457	0.534	0.076	Price 2019
	Q2	0.457	0.506	0.049	
	Q3	0.463	0.463	0.000	
	Q4	0.463	0.463	0.000	

Table 5. Quality adjusted life years for individual trials and pooled trials

Trial	Unit	EQ-5D (usual care)	EQ-5D (electroacupuncture)[a]	Incremental change in EQ-5D	Mapping algorithm used
Berman 1999[4]	QALY	0.473	0.531	0.057	Wailoo 2014
Dunning 2018[8]	QALY	0.592	0.635	0.043	Wailoo 2014
Mavrommatis 2012[18]	QALY	0.348	0.396	0.048	Price 2019
Suarez-Almazor 2010[26]	QALY	0.460	0.492	0.031	Price 2019
Pooled trials	Weighted QALY	0.496	0.537	0.041	-
	Unweighted QALY	0.468	0.513	0.045	

2.3.3 Duration of treatment benefit

During the guideline economic review of acupuncture in OA, there were 3 economic evaluations identified.[12, 24, 29] In Latimer 2012, the model from the previous OA guideline, the treatment effect was assumed to end at the end of the trial follow-up period. The same

assumption was made in Whitehurst 2011. In Reinhold 2008 however, the time horizon was 12 months, of which the duration of treatment was 3 months. There was usual care effect data beyond 3 months, so data were instead extrapolated up to 12 months, the result being that utility gradually began declining back to baseline at 12 months.

Acupuncture for OA was also evaluated in a detailed NIHR report.[15] The NIHR economic model assumed that the benefits of acupuncture and other non-pharmacological treatments lasted only for 8 weeks due to limited evidence regarding whether or not the treatment effects continued beyond the treatment period.

In contrast to this, the NICE chronic pain guideline model had trial data beyond the acupuncture treatment period and, explicitly assumed that the treatment effect linearly increased from baseline and then linearly diminished after the acupuncture sessions had ceased. However, since the treatment benefits beyond the trial period were uncertain, two base cases were modelled: one where the time horizon of the model was at the end of trial data (12 weeks) and another where the treatment effect was extrapolated beyond the trial data (up to 36 weeks).

The guideline committee considered this all during discussions and agreed that it was likely the treatment effect continued after the trial concluded, albeit waning over time. However, in the absence of trial data, it was unclear how long the treatment effect would persist. While the assumption from Reinhold 2008 utility would decline back to baseline at 12 months was considered by the committee to be too lenient, they agreed that a gradual decline in treatment effect back to baseline at 36 weeks as observed in the chronic pain model was more plausible. However, the committee still felt that 36 weeks as lenient so settled on 26 weeks or half a year. The committee therefore decided that the model would include treatment effect until 26 weeks. Figure 2 depicts the modelled treatment effect extrapolated up to 26 weeks using data from the pooled trials, while Figure 3 shows the treatment effect with no further extrapolation. The area under the curve represents the incremental QALY gain with an 8-week course of electroacupuncture compared with usual care. The solid line up to 12 weeks represents the treatment effect observed during the trial period, while the dotted line from week 12 onwards represents a linear extrapolation, or a lack thereof.

Figure 2. Treatment effect extrapolated up to 26 weeks (pooled trials)

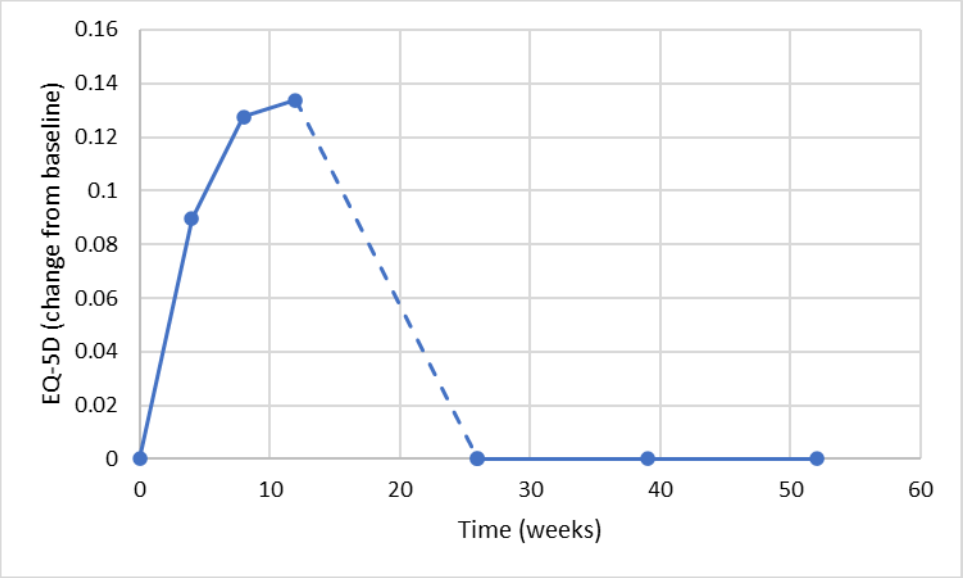
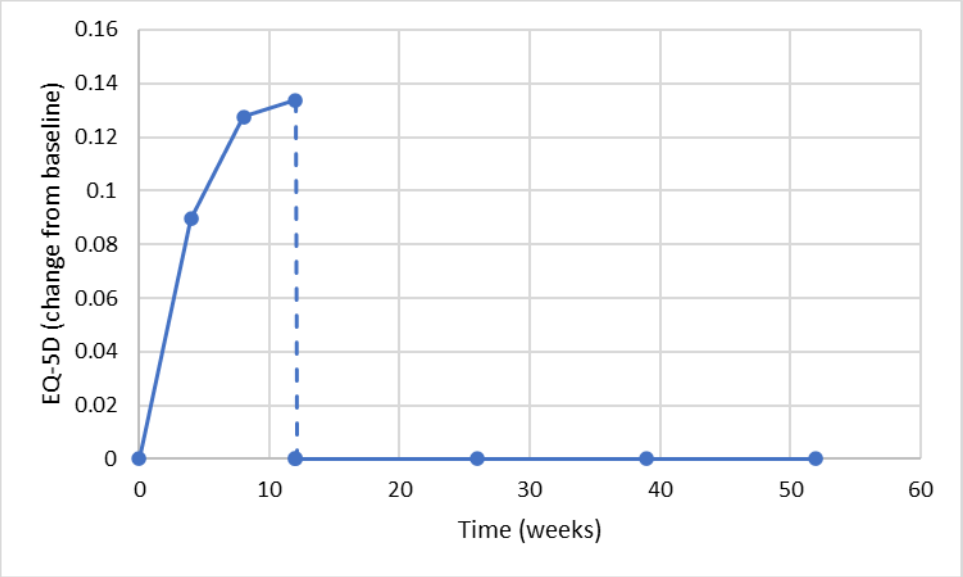


Figure 3. Treatment effect over 12 weeks with no further extrapolation (pooled trials)



2.3.4 Resource use and costs

The main intervention resource use costed in other economic evaluations is the NHS staff time. MacPherson 2017 assumed that sessions were given weekly lasting for around 40 minutes over 8 weeks by a hospital physiotherapist. This was based on feedback from clinical consultants and published literature.

The chronic pain review used a weighted average of the available trials to calculate resource use. Unit cost was based on a band 6 NHS staff (cost of around £65/hour, which was decided by the committee).

From the three studies identified during the guideline economic review,[12, 24, 29] Reinhold 2008 reported that resource use consisted of 10-15 sessions over 3 months. The duration of the sessions was not reported.

In Whitehurst 2011, 6 sessions of acupuncture were delivered over 6 weeks, with each session was estimated to last 45 minutes. The cost of a session was based on a unit cost associated with NHS community physical therapy as it was thought physical therapists were the largest group of healthcare professionals providing acupuncture in the NHS. During each follow-up, additional OA-related resource use data such as GP or hospital consultations, prescribed medications or any over-the-counter purchases were collected from self-reported postal questionnaires. As a result, the total healthcare costs per individual in the acupuncture arm was reported as being £312 over 12 months, which was £78 more that the control arm (who were receiving advice and exercise sessions only).

Finally, Latimer 2012 assumed that each session lasted for 30 minutes and was delivered by a physiotherapist.[12] This followed the assumptions from the original NICE analysis on which it was based (CG59).

The guideline committee agreed that the approach of using a weighted average that was taken in the chronic pain review was appropriate since the resource use would then reflect the clinical outcomes in the model. The resource use from the individual trials and from pooled trials is shown in Table 6. The frequency, duration and length of sessions are broadly similar to those reported in MacPherson 2017 and Whitehurst 2011, both of which incidentally took a UK NHS perspective. This suggests that resource use modelled for electroacupuncture is reflective of UK clinical practice.

Table 6. Resource use and costs

Study	Frequency (per week)	Duration of intervention (weeks)	Length of intervention (minutes)	Total minutes	Total hours	Acupuncture needle type used	Number of pressure points needled	Electrostimulator used
Berman 1999[4]	2	8	20	320	5.3	22mm/34G	9	NR
Dunning 2018[8]	1 or 2 (up to 10 sessions in total)	6	20-30	300	5.0	0.25mm x 30mm 0.30mm x 40mm 0.30mm x 50mm	9	ES-160 electrostimulator
Mavrommatis (2012)[18]	2	8	20	320	5.3	30mm/30G	10	ES-160 electrostimulator
Suarez-Almazor (2010)[26]	2	6	30	360	6.0	Filiform needles 34G, 1-1.5cun and 36G 0.5cun for ear-knee	6	TENS equipment
Pooled trials	1.92	7	25	336	5.6	Copper handled acupuncture needles	10	ES-160 electrostimulator

Abbreviations: cun= unit of measurement used in acupuncture; G= guage; mm= millimetres; NR= not reported

2.3.4.1 Costs

The costs of different bands of staff used in the analysis are presented in Table 7.

Table 7: Staff costs

Staff type	Cost per hour	Source
Base case		
Band 6 community physiotherapist	£50	PSSRU 2020 [(a)]
Sensitivity analysis		
Band 5 community physiotherapist	£38	PSSRU 2020 [(a)]
Band 7 community physiotherapist	£60	PSSRU 2020 [(a)]
GP	£153	PSSRU 2020 [(a)]
Band 4 community healthcare assistant	£33	PSSRU 2020 [(a)] Band 4 community-based scientific and professional staff used as a proxy.

(a) Costs include qualification costs (excluding individual and productivity costs), taken from PSSRU 2020, section V.18.

The band of staff that would deliver the intervention was discussed extensively with the committee. Theoretically, a band 5 could also deliver the intervention, but would require a lot of managerial support. More generally it was thought a band 6 or above would be more typical. However, this might be the case because of career structure (e.g., more senior staff looking for a new field to train in) rather than a certain grade being a prerequisite for delivering the intervention. The needling itself is a skill that can come with practice. There are also the contextual effects associated with acupuncture, in terms of the way the clinician interacts with the patient for example, and a higher-grade individual might provide more of a contextual effect. After discussing all these points, the committee felt that a band 6 staff member should be used in the base case, and a higher and lower band tested in sensitivity analyses.

Total session staff costs

In the base case, it is assumed there are 1-1 sessions between the patient and the healthcare professional. The length of sessions as well as their frequency were taken from the clinical trials (see Table 6). The total cost of sessions was calculated by multiplying the total hours by the hourly staff cost.

Device costs

In the base case, the ES-160 electroacupuncture device was chosen to deliver electroacupuncture since it was used in two of the four clinical trials. A scenario analysis explored the effect of using the AS-super 4 electroacupuncture device, which is a popular alternative in clinical practice. It was assumed that both devices had a lifespan of 5 years. The cost of each device along with their necessary accessories are presented in

Table 8.

Table 8. Unit cost of electroacupuncture devices

Device details	Device cost[(a)]	Cost of crocodile clips[(a)]	Cost of lead cables[(a)]
ES-160	£395[9]	£39.50[10] [(b)]	£41.20[11] [(b)]
AS-super 4	£240	£23[(c)]	£0

(a) Taken from online sources, excluding VAT.
(b) Cost of 10 units as the base case assumption is that 10 needles are utilised per session
(c) Clips and cables sold together

The durability of the devices and their accessories were assumed to be 5 years, and 6 months, respectively. Therefore, the device cost per year was calculated by annuitization, using a discount rate of 3.5%. This annuitized cost was divided by the expected number of sessions in a year to give an average cost per session. We assumed that 3 sessions were delivered a week, each session lasting 3.5 hours, which gave a total of 546 hours per year. This was then divided by the average length of an individual session to give the number of sessions per year. To account for missed appointments, the total number of sessions per year was multiplied by an attendance rate of 95%.[11] The final device cost per session was calculated at £0.07 or £0.10, depending on the length of the session.

The ES-160 device is battery operated, and therefore 4 batteries, at a total cost of £1.38 were also included.[5] It was assumed that the batteries had a lifespan of 18 hours.[25] The battery cost per session was calculated by taking the average cost of the batteries per hour (£1.38/18) and multiplying it by the length of the session.

Consumable costs

There were certain consumables that are required to safely deliver electroacupuncture sessions and their costs are listed in Table 9 below.

Table 9. Consumable costs per patient per session

Consumable (quantity)	Cost per patient per session
Needles (10)	£0.80
Disinfectant swab (1)	£0.30
Examination gloves (1)	£0.12

A guideline committee member remarked that they would expect to see copper handled acupuncture needles in practice. The cost per individual copper handled needle was therefore taken from the NHS supply chain, which was £0.08.[5] The number of needles needed per session were discussed with the committee. The assumption was made to use 10 needles per session since this was the maximum number of needles used in the electroacupuncture trials. The cost of the needles is small in comparison to the staff costs.

The costs of disinfectant swabs and examination gloves were based on an average unit costs taken from the NHS supply chain July 2020.[5] It was assumed that a single unit of each would be required per patient per session.

Booster sessions

The base case depicted the treatment schedule used in the clinical trials, in which patients were given electroacupuncture between 6-8 weeks only. Any benefit associated with electroacupuncture would disappear by 6 months. A separate scenario explored the effect of a maintenance schedule during which patients were given a booster session every two months once their initial programme had concluded. This meant patients would have 5

booster sessions each year. As a result of these booster sessions, the treatment benefit resulting from electroacupuncture was assumed to remain steady throughout the year.

2.4 Computations

The model was constructed in Microsoft Excel 2010 and was evaluated on an individual patient basis.

The QoL difference between electroacupuncture and usual care (taking into account baseline differences) was the treatment effect. This was based on studies in the clinical review where reported outcomes had been mapped to EQ-5D. QoL differences were based on a meta-analysis of change from baseline scores from the acupuncture group compared to the usual care group. Both the pooled and individual study EQ-5D difference at each time point were extracted and a linear trend line fitted to the points. Treatment effect was extrapolated beyond the trial data at 12 weeks using a downward-sloping line that assumed there was no difference in treatment effect between both arms after 26 weeks.

The area beneath the trend line was considered the area under the curve for calculating QALY gain. Only the incremental QALYs and costs were calculated. Costs were calculated based on average resource use from the trials and were pooled using a weighted average based on the number of participants analysed in the study. Costs and QALYs were not discounted because the model time horizon was set to 1 year.

The incremental cost and QALYs accrued by the patient were used to calculate a cost per QALY for acupuncture.

2.5 Sensitivity analyses

All the sensitivity analyses were undertaken probabilistically and deterministically to both the pooled and individual study base cases.

SA1: 3-month time horizon

A 3-month time horizon was chosen for sensitivity analysis to reflect the durations observed in three of the four trials. The Mavrommatis 2012 trial reported outcomes data up to 8 weeks and it was therefore assumed that treatment effects remained steady between weeks 8-12. This approach was chosen to keep model assumptions simple since the other three trials did not provide a clear indication treatment effect post-intervention; Berman 1999 reported a decline in effect, Dunning 2018 reported an increase and Suarez Almazor 2010 reported a steady state. By applying a 3-month time horizon, the treatment effect would be confined to this duration only.

SA2: Booster sessions included

The committee remarked that the trial settings did not reflect real-world practice since patients would not initiate a course of electroacupuncture and then cease follow-up once the initial course had concluded. Booster sessions would be required during the maintenance phase, which could vary between once per month and once per year. It is also unclear what proportion of patients would require the booster sessions and therefore it was decided the inclusion of booster sessions would be explored in a sensitivity analysis. Booster sessions would be therefore given every two months, which would result in a sustained treatment effect over the 1-year time horizon.

SA3: Group sessions

The trials utilised a 1-1 meeting between the patient and the healthcare professional, and this type of session was therefore used in the base case analysis. However, it is common in practice for patients to be seen together as a group, as reported in Berkovitz 2009.[3] In this study, sessions were conducted by a healthcare professional with acupuncture training who was supported by a non-acupuncturist assistant. Subsequently, it was decided that the sensitivity analysis would utilise a team consisting of a band 6 physiotherapist and a band 4 healthcare assistant. The committee reported that in their experience clinics would run up to three sessions per week lasting 3.5 hours each, so this was modelled. Furthermore, four patients would be seen at any time together and their treatment session would last up to 30 minutes. This meant that 28 patients could feasibly be treated during a treatment clinic lasting 3.5 hours under the supervision of a team of two. A description of the resource use and cost associated with group session is presented in Table 10.

Table 10. Resource use and cost associated with group sessions

Description	Cost /resource use
Staff	
Physiotherapist (band 6)	£50 (per hour)
Healthcare assistant (band 4)	£33 (per hour)
Session details	
Number of sessions	3 (per week)
Number of patients seen per session	4
Length of session	0.5 (hour)
Did not attend rate	5%
Consumables	
Needles (10)	£0.80 (per patient per session)
Disinfectant swab (1)	£0.30 (per patient per session)
Examination gloves (1)	£0.12 (per patient per session)

SA4: Using the AS-super 4 device

The committee commented that the AS-super 4 device is commonly used in UK practice. Since it was cheaper to obtain than the ES-160 device that was used in the base case and expected to provide similar levels of clinical effectiveness, it was decided that a scenario analysis would explore the effect of substituting the device used to deliver electroacupuncture. The life cycle of the AS-super 4 device was assumed to be 5 years, the same as the ES-160 device (see

Table 8).

SA5/SA6/SA7: using band 5/7 physiotherapists and GPs

In the base case, the committee agreed that a band 6 staff member might be a typical grade of physiotherapist that would deliver acupuncture. However, it could be a higher band, or even a lower band such as a band 5, providing they had adequate support.

The cost of a band 5 physiotherapist was used in a sensitivity analysis (SA5), while the cost of a band 7 physiotherapist was explored in a separate analysis (SA6). The committee were also interested in the cost of a GP providing the treatment (SA7).

SA8/SA9/SA10/SA11: Using alternative mapping algorithms (Barton/Price for WOMAC and Lawrence/Maund for SF-12/36)

Alternative mapping algorithms were chosen to convert SF-12, SF-36 and WOMAC scores to EQ-5D. See section 0 for further details.

2.6 Model validation

The model was developed in consultation with the committee; model structure, inputs and results were presented to and discussed with the committee for clinical validation and interpretation.

The model was systematically checked by the health economist undertaking the analysis; this included inputting null and extreme values and checking that results were plausible given inputs. The model was peer reviewed by a second experienced health economist from the NGC; this included systematic checking of many of the model calculations.

2.7 Estimation of cost effectiveness

The widely used cost-effectiveness metric is the incremental cost-effectiveness ratio (ICER). This is calculated by dividing the difference in costs associated with 2 alternatives by the difference in QALYs. The decision rule then applied is that if the ICER falls below a given cost per QALY threshold the result is considered to be cost effective. If both costs are lower and QALYs are higher the option is said to dominate and an ICER is not calculated.

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$

Where: Costs(A) = total costs for option A; QALYs(A) = total QALYs for option A

Cost effective if:
- ICER < Threshold

2.8 Interpreting results

NICE sets out the principles that committees should consider when judging whether an intervention offers good value for money.[20-22] In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention costs less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

3 Results

The base case results are presented in

Table 11 and Table 12. The costs and QALYs for the pooled trials were calculated in the model as both a weighted average, according to trial sample size, and a simple unweighted average. The base case results below reflect the weighted average scores. The unweighted results are presented in the sensitivity analysis and are not significantly different from the weighted results.

The base case results for pooled trials showed that both the probabilistic and deterministic costs per QALY for electroacupuncture versus usual care were below the NICE cost effectiveness threshold of £20,000 per QALY gained (£7,504 and £7,209). The probability of cost effectiveness at £20,000 was 97%, and this increased to 98% when the threshold was raised to £30,000.

When the individual trials were scrutinised, all four trials showed that electroacupuncture was cost effective versus usual care. This trend was also observed in the results of the sensitivity analyses where electroacupuncture was cost effective versus usual care, except when it was delivered by a GP. The results appeared robust. In the analysis of pooled trials, the cost per QALY gained ranged between £5,258 and £12,581, except with a GP where the cost per QALY gained was £21,833.

Table 11: Base case results - by arm (probabilistic)

Base case	Treatment arm	*Staff costs*	*Equipment costs*	*Consumable costs*	Total costs	QALYs
Pooled trials	EA	£273	£8	£15	£296	0.523
	UC	£0	£0	£0	£0	0.380
Berman 1999	EA	£267	£9	£20	£295	0.546
	UC	£0	£0	£0	£0	0.507
Dunning 2018	EA	£250	£6	£12	£268	0.607
	UC	£0	£0	£0	£0	0.580
Mavrommatis 2012	EA	£267	£9	£20	£295	0.396
	UC	£0	£0	£0	£0	0.298
Suarez Almazor 2010	EA	£300	£8	£15	£323	0.491
	UC	£0	£0	£0	£0	0.460

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality-adjusted life-years

Table 12. Base case results– incremental EA vs. UC (probabilistic and deterministic)

Base case	Analysis	Incremental cost	Incremental QALYs	Cost per QALY gained	Probability cost effective at £20k	Probability cost effective at £30k
Pooled trials	Probabilistic	£296	0.039	£7,504	97%	99%
	Deterministic	£296	0.041	£7,209	NA	NA
Berman 1999	Probabilistic	£295	0.039	£7,641	62%	71%
	Deterministic	£295	0.057	£5,163	NA	NA
Dunning 2018	Probabilistic	£268	0.027	£10,098	59%	68%
	Deterministic	£268	0.043	£6,217	NA	NA
Mavrommatis 2012	Probabilistic	£295	0.098	£3,010	99%	99%
	Deterministic	£295	0.048	£6,204	NA	NA
Suarez Almazor 2010	Probabilistic	£323	0.031	£10,267	99%	99%
	Deterministic	£323	0.031	£10,314	NA	NA

Abbreviations: EA=electroacupuncture; NA=not applicable; NT=Usual care; QALYs=quality adjusted life years

3.1 Sensitivity analyses

Table 13. Pooled trials (costs and QALYs calculated based on a weighted average)

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£296	0.039	£7,504	97%	99%
Time horizon					
SA1 3-month time horizon	£296	0.024	£12,581	86%	97%
SA2 Booster sessions	£497	0.094	£5,259	99%	99%
Costs					
SA3 Group sessions	£207	0.039	£5,258	99%	99%
SA4 Using AS-super 4 device	£289	0.039	£7,353	98%	99%
SA4 Band 5 physiotherapist	£230	0.040	£5,801	98%	99%
SA6 Band 7 physiotherapist	£350	0.040	£8,830	97%	99%
SA7 GP	£858	0.039	£21,833	34%	75%
Utilities					
SA8 Alternative utilities (Barton)	£296	0.041	£7,298	79%	84%
SA9 Alternative utilities (Lawrence)	£296	0.050	£5,857	100%	100%
SA10 Alternative utilities (Price)	£296	0.031	£9,604	86%	94%
SA11 Alternative utilities (Maund)	£296	0.031	£9,420	56%	58%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

Table 14. Pooled trials (costs and QALYs calculated based on an unweighted average)

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£296	0.049	£6,068	79%	85%
Time horizon					
SA1 3-month time horizon	£296	0.029	£10,222	67%	79%
SA2 Booster sessions	£492	0.116	£4,245	85%	88%
Costs					
SA3 Group sessions	£213	0.049	£4,383	83%	87%
SA4 Using AS-super 4 device	£288	0.048	£5,946	79%	85%
SA4 Band 5 physiotherapist	£231	0.049	£4,714	83%	87%
SA6 Band 7 physiotherapist	£350	0.049	£7,148	76%	83%
SA7 GP	£853	0.048	£17,612	41%	59%
Utilities					
SA8 Alternative utilities (Barton)	£296	0.052	£5,723	79%	81%
SA9 Alternative utilities (Lawrence)	£296	0.060	£4,934	99%	100%
SA10 Alternative utilities (Price)	£296	0.033	£8,847	75%	82%
SA11 Alternative utilities (Maund)	£296	0.036	£8,189	57%	62%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

Table 15. Berman 1999 trial

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£295	0.039	£7,641	62%	72%
Time horizon					
SA1 3-month time horizon	£295	0.025	£11,989	52%	63%
SA2 Booster sessions	£483	0.098	£4,934	72%	77%
Costs					
SA3 Group sessions	£227	0.039	£5,851	68%	75%
SA4 Using AS-super 4 device	£287	0.038	£7,598	63%	71%
SA4 Band 5 physiotherapist	£231	0.039	£5,940	68%	74%
SA6 Band 7 physiotherapist	£349	0.039	£8,901	58%	68%
SA7 GP	£845	0.038	£22,292	34%	46%
Utilities					
SA8 Alternative utilities (Barton)	£295	0.052	£5,716	65%	67%
SA9 Alternative utilities (Lawrence)	£295	0.061	£4,861	100%	100%
SA10 Alternative utilities (Price)	£295	0.038	£7,858	62%	70%
SA11 Alternative utilities (Maund)	£295	0.039	£7,664	63%	71%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

Table 16. Dunning 2018 trial

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£268	0.027	£10,098	59%	68%
Time horizon					
SA1 3-month time horizon	£268	0.016	£16,824	44%	56%
SA2 Booster sessions	£468	0.064	£7,289	67%	75%
Costs					
SA3 Group sessions	£194	0.027	£7,320	66%	74%
SA4 Using AS-super 4 device	£263	0.026	£9,936	58%	68%
SA4 Band 5 physiotherapist	£208	0.027	£7,612	65%	73%
SA6 Band 7 physiotherapist	£318	0.027	£11,681	54%	65%
SA7 GP	£783	0.027	£29,184	27%	40%
Utilities					
SA8 Alternative utilities (Barton)	£268	0.025	£10,656	54%	57%
SA9 Alternative utilities (Lawrence)	£268	0.049	£5,430	100%	100%
SA10 Alternative utilities (Price)	£268	0.026	£10,230	58%	67%
SA11 Alternative utilities (Maund)	£268	0.027	£9,898	59%	69%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

Table 17. Mavrommatis 2012 trial

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£295	0.098	£3,010	100%	100%
Time horizon					
SA1 3-month time horizon	£295	0.056	£5,272	100%	100%
SA2 Booster sessions	£483	0.225	£2,149	100%	100%
Costs					
SA3 Group sessions	£227	0.098	£2,308	100%	100%
SA4 Using AS-super 4 device	£287	0.098	£2,926	100%	100%
SA4 Band 5 physiotherapist	£231	0.098	£2,364	100%	100%
SA6 Band 7 physiotherapist	£349	0.098	£3,563	100%	100%
SA7 GP	£845	0.098	£8,632	100%	100%
Utilities					
SA8 Alternative utilities (Barton)	£295	0.098	£3,011	100%	100%
SA9 Alternative utilities (Lawrence)	£295	0.098	£3,010	100%	100%
SA10 Alternative utilities (Price)	£295	0.040	£7,392	95%	98%
SA11 Alternative utilities (Maund)	£295	0.053	£5,563	59%	60%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

Table 18. Suarez Almazor 2010 trial

Analysis	Mean difference (EA-UC)		ICER (Cost per QALY gained)	Probability cost effective at £20k	Probability cost effective at £30k
	Inc. cost	Inc. QALY			
Basecase results	£323	0.031	£10,267	96%	99%
Time horizon					
SA1 3-month time horizon	£323	0.019	£16,970	72%	96%
SA2 Booster sessions	£534	0.077	£6,959	100%	100%
Costs					
SA3 Group sessions	£206	0.031	£6,585	99%	100%
SA4 Using AS-super 4 device	£316	0.032	£10,019	96%	99%
SA4 Band 5 physiotherapist	£251	0.031	£7,998	98%	100%
SA6 Band 7 physiotherapist	£383	0.031	£12,194	92%	98%
SA7 GP	£941	0.031	£30,117	3%	49%
Utilities					
SA8 Alternative utilities (Barton)	£323	0.032	£10,231	96%	99%
SA9 Alternative utilities (Lawrence)	£323	0.031	£10,324	95%	99%
SA10 Alternative utilities (Price)	£323	0.030	£10,827	83%	91%
SA11 Alternative utilities (Maund)	£323	0.026	£12,619	48%	49%

Abbreviations: EA=electroacupuncture; UC=Usual care; QALYs=quality adjusted life years

4 Discussion

4.1 Summary of results

One cost utility analysis reported that acupuncture was cost effective compared with usual care (cost per QALY gained of £13,944). This analysis was assessed as partially applicable with minor limitations.

One cost utility analysis reported that acupuncture was cost effective compared with usual care (cost per QALY gained of £3,889). This analysis was assessed as directly applicable with minor limitations.

One cost utility analysis of three separate trials reported that acupuncture was cost effective compared with usual care (cost per QALY gained ranged between £6,911 and £17,381). This analysis was assessed as directly applicable with potentially serious limitations.

One cost utility analysis reported that acupuncture was cost effective compared with usual care (cost per QALY gained of £12,786). In a full incremental analysis versus other intervention (braces, heat treatment, insoles, interferential therapy, laser/light therapy, manual therapy, neuromuscular electrical stimulation, pulsed electrical stimulation, pulsed electromagnetic field, static magnets and transcutaneous electrical nerve stimulation) where the analysis was confined to trials with adequate allocation concealment and also trials with adequate allocation concealment with an endpoint between 3-13 weeks, acupuncture was the most cost-effective strategy with costs per QALY gained of £13,502 and £14,275, respectively.

One original cost utility analysis reported that electroacupuncture was cost effective compared with usual care with a cost per QALY gained of £7,504. This analysis was assessed as directly applicable with potentially serious limitations.

The committee agreed that based on the available evidence, acupuncture and electroacupuncture were both cost effective versus usual care. However, they were unconvinced by the clinical evidence of the magnitude of the treatment benefit with both acupuncture and electroacupuncture, which was a particularly important consideration given the large resource implication for the NHS if either intervention were to be recommended.

4.2 Limitations and interpretation

This analysis and model attempted to assess the cost effectiveness of electroacupuncture versus usual care in people with OA. However, there are a number of limitations that should be taken into account when interpreting this analysis.

The analysis was based on four studies in total with a usual care comparison, which is a small number. There was heterogeneity between the studies in terms of the type of electroacupuncture device used, the voltage given as well as the pressure points selected for needling. Presenting the results according to individual trials does address this limitation, however two of the trials reported small patient numbers: N=73 in Berman 1999 and N=80 in Mavrommatis 2012, which made it difficult to make meaningful analysis based on these solely. Consequently, results based on a pooled estimate of effect were also presented to the committee, using both a weighted average and an unweighted average. There were also different lengths of interventions and timeframes that outcomes were reported between studies. The populations in the studies however were felt to be representative of the OA population.

The long-term effects of electroacupuncture were not recorded in the clinical trials, and therefore treatment effects beyond the trial follow-up period were extrapolated based on

clinical expert judgement. These assumptions were tested during sensitivity analyses; however, it is not a substitute for data from clinical trials. It is therefore unclear presently whether the base case model assumption was a fair reflection of actual treatment effects with electroacupuncture.

Lastly, direct valuations of EQ-5D were not reported in the clinical trials, and EQ-5D were mapped instead from the SF-12/36 summary scores and WOMAC scores. This process will have led to additional uncertainty in the model results.

4.3 Generalisability to other populations or settings

The populations reflected in the trials used for treatment effect in this analysis are people with OA of the knee. The committee agreed that these populations are likely to be generalisable to the wider OA of the knee population.

4.4 Comparisons with published studies

There were no studies identified that evaluated the cost effectiveness of electroacupuncture versus usual care.

The four studies identified during the economic review all reported that acupuncture was cost effective compared to usual care (see section 4.1). Acupuncture was also evaluated in the NICE chronic pain in over 16s guideline [NG193].[19] The cost per QALY versus no acupuncture was reported as £5,710 in a lifetime analysis with treatment effects extrapolated up to 36 weeks and £14,552 when the data was not extrapolated beyond the trial period.

This analysis differed from those mentioned above in that it specifically compared electroacupuncture to usual care. The incremental costs and QALYs reported here are similar to those reported with acupuncture versus usual care, which is not surprising since when the additional cost for electrical devices needed to deliver electroacupuncture are distributed over the total number of sessions during the device's life, the additional cost becomes minuscule.

4.5 Conclusions

This economic evaluation demonstrated that electroacupuncture is cost effective compared to usual care in people with osteoarthritis.

The conclusions of this analysis are robust to most the assumptions used, except when the service was delivered by a GP. Previous research evaluating acupuncture in osteoarthritis also found it cost effective compared to usual care.

4.6 Implications for future research

Further research is warranted in specific sub-populations of osteoarthritis in who acupuncture or electroacupuncture could be clinically and cost effective, in line with the guideline research recommendation. Longer follow-up of people receiving acupuncture or electroacupuncture is also warranted to adequately model their treatment effects. Direct EQ-5D valuations should also be utilised in these trials

References

1. Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? Health and quality of life outcomes. 2008; 6(1):1-11

2. Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? Health Qual Life Outcomes. 2008; 6:51

3. Berkovitz S, Cummings M, Perrin C, Ito R. High volume acupuncture clinic (HVAC) for chronic knee pain–audit of a possible model for delivery of acupuncture in the National Health Service. Acupuncture in Medicine. 2008; 26(1):46-50

4. Berman B, Singh B, Lao L, Langenberg P, Li H, Hadhazy V et al. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. Rheumatology (Oxford, England). 1999; 38(4):346-354

5. Catalogue. NSC. NHS Supply Chain, 2020. 2020. Available from: Available from: http://www.supplychain.nhs.uk/ Last accessed:

6. Chan KK, Willan AR, Gupta M, Pullenayegum E. Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies. Medical Decision Making. 2014; 34(7):863-872

7. Chuang LH, Whitehead SJ. Mapping for economic evaluation. British Medical Bulletin. 2012; 101:1-15

8. Dunning J, Butts R, Young I, Mourad F, Galante V, Bliton P et al. Periosteal electrical dry needling as an adjunct to exercise and manual therapy for knee osteoarthritis: a multicenter randomized clinical trial. The Clinical journal of pain. 2018; 34(12):1149

9. Harmony M. Electroacupuncture machine ES-160. 2022. Available from: https://www.harmonymedical.co.uk/electro-acupuncture-machine-es-160.html Last accessed: 22/02/2022.

10. Harmony M. Electroacupuncture machine ES-160 crocodlie clips. 2022. Available from: https://www.harmonymedical.co.uk/crocodile-clip.html Last accessed: 22/02/2022.

11. Harmony M. Electroacupuncture machine ES-160 lead cables. 2022. Available from: https://www.harmonymedical.co.uk/acupuncture/electrotherapy/replacement-leads/crocodile-clip.html Last accessed: 22/02/2022.

12. Latimer NR, Bhanu AC, Whitehurst DG. Inconsistencies in NICE guidance for acupuncture: reanalysis and discussion. Acupuncture in Medicine. 2012; 30(3):182-186

13. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. Medical Decision Making. 2004; 24(2):160-169

14. Longworth L, Rowen D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. 2011. Available from: http://www.nicedsu.org.uk

15. MacPherson H, Vickers A, Bland JM, Torgerson DJ, Corbett MS, Spackman E et al. Acupuncture for chronic pain and depression in primary care:: a programme of research. Programme Grants for Applied Research. 2017:1-342

16. MacPherson H, Vickers A, Bland M, Torgerson D, Corbett M, Spackman E et al. Acupuncture for chronic pain and depression in primary care: a programme of research. NIHR Journals Library Programme Grants for Applied Research. 2017; 5(3):342

17. Maund E, Craig D, Suekarran S, Neilson A, Wright K, Brealey S et al. Management of frozen shoulder: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2012; 16(11):1-264

18. Mavrommatis CI, Argyra E, Vadalouka A, Vasilakos DG. Acupuncture as an adjunctive therapy to pharmacological treatment in patients with chronic pain due to osteoarthritis of the knee: a 3-armed, randomized, placebo-controlled trial. PAIN®. 2012; 153(8):1720-1726

19. National Institute for H, Clinical E. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. 2021;

20. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2020]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

21. National Institute for Health and Care Excellence. The NICE Charter. 2020. Available from: https://www.nice.org.uk/about/who-we-are/our-charter Last accessed: 10/02/2022.

22. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for Health and Clinical Excellence, 2008. Available from: https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf

23. Price A, Smith J, Dakin H, Kang S, Eibich P, Cook J et al. The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: development and economic modelling. Health Technology Assessment (Winchester, England). 2019; 23(32):1-216

24. Reinhold T, Witt CM, Jena S, Brinkhaus B, Willich SN. Quality of life and cost-effectiveness of acupuncture treatment in patients with osteoarthritis pain. The European Journal of Health Economics. 2008; 9(3):209-219

25. Rs. A Complete C Battery Guide. 2022. Available from: https://uk.rs-online.com/web/generalDisplay.html?id=ideas-and-advice/c-batteries-guide Last accessed: 22/02/2022.

26. Suarez-Almazor ME, Looney C, Liu Y, Cox V, Pietz K, Marcus DM et al. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. Arthritis Care & Research. 2010; 62(9):1229-1236

27. Wailoo A, Alava MH, Martinez AE. Modelling the relationship between the WOMAC osteoarthritis index and EQ-5D. Health and quality of life outcomes. 2014; 12(1):1-6

28. Watson J, Helliwell P, Morton V, Adebajo A, Dickson J, Russell I et al. Shoulder acute pain in primary healthcare: is retraining effective for GP principals? SAPPHIRE--a randomized controlled trial. Rheumatology. 2008; 47(12):1795-1802

29. Whitehurst DG, Bryan S, Hay EM, Thomas E, Young J, Foster NE. Cost-effectiveness of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee. Physical therapy. 2011; 91(5):630-641

Appendices

Appendix A: Raw trial data

Table 19. SF-12 raw data

Intervention	Measurement timeframe		SF-12 summary score		EQ-5D Mapped from SF-12 (base case: Price algorithm)	EQ-5D Mapped from SF-12 (Lawrence algorithm)	EQ-5D Mapped from SF-12 (Maund algorithm)
			Physical component score	Mental component score			
Suarez Almazor 2010							
EA	Baseline	Mean	35.0	52.3	0.463	0.628	0.627
		Standard deviation	9.9	9.4			
	4 weeks	Mean	38.5	53.9	0.532	0.692	0.671
		Standard deviation	10.0	8.3			
	6 weeks	Mean	40.5	53.4	0.560	0.716	0.695
		Standard deviation	10.0	7.9			
	12 weeks	Mean	39.5	54.1	0.550	0.708	0.683
		Standard deviation	9.7	8.2			
UC	Baseline	Mean	35.3	53.7	0.483	0.645	0.632
		Standard deviation	8.4	10.7			
	4 weeks	Mean					
		Standard deviation					
	6 weeks	Mean					
		Standard deviation					
	12 weeks	Mean	35.8	51.6	0.472	0.632	0.636
		Standard deviation	8.9	9.8			

Table 20. SF-36 raw data

Intervention	Measurement timeframe		SF-36 summary score		EQ-5D Mapped from SF-36 (base case: Price algorithm)	EQ-5D Mapped from SF-36 (Lawrence algorithm)	EQ-5D Mapped from SF-36 (Maund algorithm)
			Physical component score	Mental component score			
Mavrommatis 2012							
EA	Baseline	Mean	29.3	45.9	0.298	0.485	0.518
		Standard deviation	5.2	8.5			
	8 weeks	Mean	45.8	52.2	0.634	0.779	0.754
		Standard deviation	6.9	8.0			
UC	Baseline	Mean	28.0	45.2	0.265	0.460	0.492
		Standard deviation	5.7	8.0			
	8 weeks	Mean	35.3	50.7	0.438	0.616	0.628
		Standard deviation	4.5	7.4			

Table 21. WOMAC scores raw data

Intervention	Measurement timeframe		WOMAC subscale score			Total WOMAC score	EQ-5D Mapped from WOMAC (base case: Wailoo algorithm)	EQ-5D Mapped from WOMAC (Barton algorithm)	EQ-5D Mapped from WOMAC (Price algorithm)
			Pain score	Stiffness score	Physical function score				
Berman 1999									
EA	Baseline	Mean	9.58		34.56	48.69	0.468	0.518	0.557
		Standard deviation	3.26		12.20	16.23			
	4 weeks	Mean	6.25		24.11	33.36	0.638	0.663	0.722
		Standard deviation	3.46		13.17	17.66			
	8 weeks	Mean	5.34		20.31	28.08	0.678	0.702	0.779
		Standard deviation	3.62		13.26	17.96			
	12 weeks	Mean	5.56		23.17	31.58	0.953	0.677	0.741
		Standard deviation	3.44		13.92	18.27			
UC	Baseline	Mean	9.78		36.19	50.87	0.435	0.482	0.546
		Standard deviation	2.83		9.22	12.30			
	4 weeks	Mean	9.46		36.11	50.05	0.453	0.491	0.556
		Standard deviation	3.50		10.04	14.03			
	8 weeks	Mean	9.46		36.14	50.11	0.452	0.491	0.555
		Standard deviation	3.56		10.55	14.52			
	12 weeks	Mean	9.51		36.78	50.43	0.450	0.487	0.551
		Standard deviation	3.01		10.71	14.10			
Dunning 2018									
EA	Baseline	Mean	8.70	4.00	28.90	57.10	0.553	0.559	0.630
		Standard deviation	3.20	1.60	10.60	13.20			
	2 weeks	Mean	5.40	2.50	17.10	25.00	0.701	0.689	0.808

Intervention	Measurement timeframe		WOMAC subscale score			Total WOMAC score	EQ-5D Mapped from WOMAC (base case: Wailoo algorithm)	EQ-5D Mapped from WOMAC (Barton algorithm)	EQ-5D Mapped from WOMAC (Price algorithm)
			Pain score	Stiffness score	Physical function score				
		Standard deviation	3.20	1.40	10.60	14.30			
	6 weeks	Mean	3.40	1.70	12.10	17.20	0.780	0.729	0.902
		Standard deviation	2.60	1.40	9.80	13.10			
	12 weeks	Mean	2.80	1.30	10.10	14.20	0.815	0.741	0.942
		Standard deviation	2.50	1.30	9.30	12.50			
UC	Baseline	Mean	8.00	3.80	28.10	39.90	0.575	0.580	0.647
		Standard deviation	3.30	1.40	11.10	14.60			
	2 weeks	Mean	6.10	3.00	22.30	31.40	0.652	0.651	0.737
		Standard deviation	3.00	1.50	11.60	15.10			
	6 weeks	Mean	4.80	2.40	18.70	25.90	0.696	0.688	0.797
		Standard deviation	2.80	1.50	10.90	14.30			
	12 weeks	Mean	5.20	2.40	18.70	26.40	0.690	0.685	0.791
		Standard deviation	3.20	1.50	10.90	15.60			

Appendix B: Adjusted standard deviations for mapping uncertainty

						Unadjusted SD's				Adjusted SD's			
Study	**Intervention**	**EQ-5D baseline mean**	**EQ-5D mean - outcome point 1**	**EQ-5D mean - outcome point 2**	**EQ-5D mean - outcome point 3**	**Baseline SD**	**Outcome point 1 SD**	**Outcome point 2 SD**	**Outcome point 3 SD**	**Baseline SD**	**Outcome point 1 SD**	**Outcome point 2 SD**	**Outcome point 3 SD**
Barton 2008 algorithm													
Berman 1999	EA	0.516	0.661	0.699	0.675	0.564	0.474	0.457	0.468	0.165	0.139	0.134	0.137
	UC	0.481	0.490	0.489	0.485	0.573	0.560	0.569	0.569	0.165	0.161	0.164	0.164
Dunning 2018	EA	0.558	0.688	0.728	0.740	0.854	0.763	0.751	0.757	0.142	0.127	0.125	0.126
	UC	0.579	0.650	0.687	0.684	0.838	0.794	0.778	0.775	0.140	0.132	0.130	0.129
Lawrence 2004 algorithm													
Mavrommatis 2012	EA	0.485	0.779			0.168	0.127			0.034	0.026		
	UC	0.460	0.616			0.114	0.099			0.023	0.020		
Suarez Almazor 2010	EA	0.628	0.692	0.716	0.708	0.167	0.163	0.161	0.159	0.034	0.017	0.017	0.016
	UC	0.645			0.632	0.158			0.158	0.024			0.024
Maund 2012 algorithm													
Mavrommatis 2012	EA	0.421	0.531			2.889	3.052			0.688	0.727		
	UC	-0.520	-0.324			1.532	1.984			0.365	0.472		
Suarez Almazor 2010	EA	0.460	0.474	0.488	0.478	6.073	6.198	6.145	6.254	0.740	0.755	0.748	0.762
	UC	0.455			0.469	4.257			4.136	0.756			0.734