
Article:
Saramago Goncalves, Pedro Rafael orcid.org/0000-0001-9063-8590, Woods, Bethan Sarah orcid.org/0000-0002-7669-9415, Weatherly, Helen Louise Ann orcid.org/0000-0002-9117-6452 et al. (5 more authors) (2016) Methods for network meta-analysis of continuous outcomes using individual patient data:a case study in acupuncture for chronic pain. BMC Medical Research Methodology. ISSN 1471-2288

Saramago et al. BMC Medical Research Methodology (2016) 16:131
DOI 10.1186/s12874-016-0224-1

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Methods for network meta-analysis of continuous outcomes using individual patient data: a case study in acupuncture for chronic pain

Pedro Saramago[1*], Beth Woods[1], Helen Weatherly[1], Andrea Manca[1], Mark Sculpher[1], Kamran Khan[2], Andrew J. Vickers[3] and Hugh MacPherson[4]

Abstract

Background: Network meta-analysis methods, which are an extension of the standard pair-wise synthesis framework, allow for the simultaneous comparison of multiple interventions and consideration of the entire body of evidence in a single statistical model. There are well-established advantages to using individual patient data to perform network meta-analysis and methods for network meta-analysis of individual patient data have already been developed for dichotomous and time-to-event data. This paper describes appropriate methods for the network meta-analysis of individual patient data on continuous outcomes.

Methods: This paper introduces and describes network meta-analysis of individual patient data models for continuous outcomes using the analysis of covariance framework. Comparisons are made between this approach and change score and final score only approaches, which are frequently used and have been proposed in the methodological literature. A motivating example on the effectiveness of acupuncture for chronic pain is used to demonstrate the methods. Individual patient data on 28 randomised controlled trials were synthesised. Consistency of endpoints across the evidence base was obtained through standardisation and mapping exercises.

Results: Individual patient data availability avoided the use of non-baseline-adjusted models, allowing instead for analysis of covariance models to be applied and thus improving the precision of treatment effect estimates while adjusting for baseline imbalance.

Conclusions: The network meta-analysis of individual patient data using the analysis of covariance approach is advocated to be the most appropriate modelling approach for network meta-analysis of continuous outcomes, particularly in the presence of baseline imbalance. Further methods developments are required to address the challenge of analysing aggregate level data in the presence of baseline imbalance.

Keywords: Evidence synthesis, Network meta-analysis, Mixed treatment comparisons, Individual-patient data, Analysis of covariance, Continuous outcome, Heterogeneity

* Correspondence: pedro.saramago@york.ac.uk
[1]Centre for Health Economics, University of York, York, UK
Full list of author information is available at the end of the article

Background

Evidence synthesis tools are increasingly used to pool estimates of treatment effects from multiple randomised controlled trials (RCTs) to inform assessments of comparative effectiveness generally, and particularly in the context of health technology assessment. One such tool is network meta-analysis (NMA, also known as mixed treatment comparisons), which extends standard pair-wise meta-analysis by allowing the simultaneous synthesis of evidence on multiple treatments [1–4]. Most published work focuses on the pooling of aggregate outcome data (AD), but with the increasing availability of individual patient data (IPD) synthesis methods have recently emerged to utilise IPD [5–10]. The use of IPD allows the consistent use of statistical methods across the body of evidence. It also creates added value by offering the potential to reduce and/or explain network heterogeneity, tackle existing evidence inconsistencies [11], and to examine subgroup effects in patients where interventions might have an effectiveness profile which differs from that of the wider population [10, 12]. Despite its advantages, only a few methodological studies on the synthesis of IPD in NMA are available in the published literature and even fewer examples of its use within cost-effectiveness (CE) analysis exist [13]. Methods for NMA of IPD have focused mainly on a subset of the available types of outcomes, i.e. binary and time-to-event outcomes [10, 14]. Few publications exist dedicated to continuous outcomes [15, 16], an important outcome set in medical applications, as well as in complementary medicine and beyond. Recent publications by Hong et al. [15] and Thom et al. [16] explored and discussed the synthesis of continuous endpoints using IPD in NMA. While the former proposes a framework to pool multiple continuous outcomes under contrast- and arm-based parameterisations, the latter focused mainly on modelling observational evidence available in both IPD and AD formats. Both papers chose the change from baseline as their continuous outcome for synthesis but did not adjust for baseline values of the outcome, apart from when modelling baseline outcome as a treatment-effect modifier [15]. In this paper we present a model for NMA of IPD on continuous outcomes using the analysis of covariance (ANCOVA) approach which does adjust for baseline outcome data.

Analysis of covariance (ANCOVA), where the outcome at follow-up is modelled whilst adjusting for its baseline value, is the preferred method for estimating treatment effects from continuous outcomes [17–19]. Treatment effect estimates based on ANCOVA methods are the most precise estimates and are robust to chance baseline imbalance. As such, these should be the desired outcome measure for synthesis [20–22]. Unfortunately, ANCOVA results are frequently not reported for individual studies and, therefore, ANCOVA is often not used in the synthesis of aggregate evidence. Instead, sub-optimal methods are used [23–25] such as unadjusted differences in change from baseline or final outcome measures.

When IPD is available from each study, the full set of statistical approaches are available to analysts. Riley et al. [22] discuss different approaches to the synthesis of continuous outcome data when IPD is available in a pairwise meta-analysis framework. The authors highlight that availability of IPD is crucial to implement the most appropriate modelling approach, the ANCOVA [19, 22]. To our knowledge, such an ANCOVA synthesis model has not yet been extended to and/or explored in the NMA setting.

In this paper we present a Bayesian NMA model for the synthesis of continuous IPD using the ANCOVA framework. The paper aims to ensure that best practice in the analysis of continuous outcome data within individual trials and pairwise meta-analyses is extended to the NMA context. We also aim to illustrate the differences (and similarities) between NMA of IPD when using ANCOVA, change score and final score only approaches. The method presented is applied to a case study of acupuncture for chronic pain. The paper is structured as follows. Section 2 presents the motivating example for the manuscript, describes the evidence available and outlines the analysis undertaken to obtain outcome data for synthesis. Section 3 describes the NMA ANCOVA model for IPD on continuous outcomes, followed by extensions that incorporate treatment effect–covariate interactions. Results of applying the methods described to the motivating dataset are reported in Section 4, which is followed by some discussion topics and concluding remarks in Section 5.

Evidence on the effectiveness of acupuncture for chronic pain in primary care

There is currently a lack of agreement about the effectiveness of acupuncture as a treatment for chronic pain, as reflected in debates about recent UK guidance surrounding its value [26–32]. Acupuncture received a positive recommendation from the National Institute for Health and Care Excellence (NICE) for its use in back pain [26] and headache/migraine [27], while a negative recommendation was given for its use in osteoarthritis in 2008 and 2014 [28]. The methods in this paper were developed as part of a project to improve evidence regarding the effectiveness and CE of acupuncture for chronic non-specific pain to inform decision making in the UK National Health Service [33].

Dataset

The data used in this study was provided by the Acupuncture Trialists' Collaboration (ATC) who performed a systematic review in which relevant high quality trials were identified and, for a large proportion of trials, IPD was obtained (29 out of 31 studies) [34, 35]. The dataset

available to us comprised 28 out of these 29 RCTs which assessed the effectiveness of acupuncture in three pain conditions: osteoarthritis of the knee (OAK) (7 trials [36–42]), headache, including tension-type headache (TTH) and migraine (6 trials [43–48]) and musculoskeletal conditions, encompassing lower back, shoulder and neck pain (15 trials [49–63]). This dataset comprises 17,512 patients. These studies are summarised in Table 1.

The dataset includes 11 trials comparing acupuncture to sham acupuncture, 8 comparing acupuncture and usual care, and 9 comparing all three comparators. The resulting evidence network is presented in Fig. 1.

Outcomes

One key aspect of the evidence available in this setting is the heterogeneous reporting of relevant outcomes across trials. The ATC dataset varied according to the type of outcomes reported but also on how these were collected across time. To address this issue, two outcome measures are used within this paper. The first is a standardised pain-related outcome, a dimensionless measure of treatment effect usually termed standardised mean difference (SMD) [20, 21, 64, 65]. For this analysis the primary outcome of each study was used to generate patient-level standardised pain estimates. Pain measures varied from days with headache in the headache/migraine pain condition, to visual analogue scale (VAS) pain in the musculoskeletal group or to Western Ontario and McMaster Universities Arthritis Index (WOMAC) pain in the OAK group, as reported in Table 1 (column on the right hand side). Individual-level standardised pain estimates were obtained for each trial by dividing the primary outcome scores by the study-specific standard deviation. Note that while these estimates were used as inputs in the synthesis models, the outputs of the synthesis are in the SMD format, as differences between treatments were estimated within the modelling[1].

While SMDs may be useful for detecting differences between interventions, they are of limited value to decision making as these cannot directly inform estimates of absolute effect or CE modelling, unless they are first transformed [20]. These considerations motivated the second synthesis approach used, which involved translating (or 'mapping') the available patient-reported outcome data from the trials into EuroQol five-dimension (EQ-5D) index values [66]. The EQ-5D is a popular preference-based generic health-related quality of life (HRQoL) measure, typically employed to weight life years gained and thus derive quality-adjusted life-years (QALYs) for use in CE analysis [67]. The EQ-5D preference score was the second outcome explored. Due to its importance in supporting health system decision making processes, the EQ-5D preference score, used in CE analysis, has applications in many jurisdictions worldwide, including the UK [68]. The conventional EQ-5D questionnaire includes five domains, each of which can be at one of three severity levels. Using an algorithm, responses to this questionnaire can be transformed to a numeric value that reflects the preferences of the public for different heath states (here we used values from the UK general public [69]). Values range from −0.594 to 1 (the bounds represent, respectively, the worst imaginable health state and full health, with zero relating to death).

Only a small number of trials ($n = 2$) in the dataset directly provided EQ-5D data [36, 56]. Where such data were not available it was predicted using other generic and disease specific measures[2] (Table 1) through published mapping algorithms. In 50 % ($n = 14$) of the trials, well established published algorithms were used to map from Short Form (SF)-36 dimensions and SF-12 summary scores to EQ-5D[3] [58, 70, 71]. In 10 of the 28 trials, published algorithms which map VAS pain scores [72] and WOMAC scores [73] to EQ-5D were used[4]. For one trial, a double mapping approach was necessary as, to our knowledge, no direct mapping algorithm exists to obtain EQ-5D values from Constant Murley Score (CMS). Thus, an unpublished mapping algorithm (a report describing the derivation of the mapping algorithm is available on request from Kamran Khan: K.A.Khan@warwick.ac.uk [74]) was used to derive VAS pain estimates from CMS, which were then used to obtain individual-level EQ-5D predictions using the Maund et al. [72] algorithm.

For the majority of mapping models used, the proportion of total variation explained (quantified by the coefficient of determination, R^2, in most cases) was low. To account for this additional source of uncertainty, an additional variance component was included[5] [75]. This was achieved by adding to each individual-level EQ-5D prediction a draw from a normal distribution with mean zero and variance equal to the study-specific residual variance, that is, $Var\left[\widehat{EQ5D}\right]\cdot(1-R^2)$, where $\widehat{EQ5D}$ is the predicted (mapped) EQ-5D at individual-level.

HRQoL and standardised pain estimates were obtained at baseline and at the follow-up period closest to 3 months following the start of treatment, as 3 months is the typical end of treatment measurement, though not necessarily the trial's primary end-point. Changes from baseline were obtained by calculating the difference between values for these two time points. Missing data in the ATC dataset (9.3 % ($n = 1{,}622$) and 15.5 % ($n = 2{,}716$) of the total number of patients in the standardised pain and HRQoL outcome, respectively) was assumed to be missing at random (MAR) and a complete-case analysis was conducted.

Additional file 1: Table A1 presents the standardised pain outcome and (mapped/predicted) EQ-5D data. For both outcome measures baseline imbalance can be observed in some trials. The source of this imbalance is not clear, but should be addressed in the synthesis [76, 77].

Table 1 Main characteristics of the data and study outcomes used for analysis

ID	Study 1st author, year	Location	Pain group (type)	Age – mean (SD)	Trial follow-up period / Time point used in the analysis (months)	Treatment	Obser-vations	HRQoL outcome mapped	Pain outcome standardised
1	Diener 2006 [46]	Germany	Headache (migraine)	37.62 (10.4)	6 / 3	Usual care	328	SF-12	Migraine days
						Sham acupuncture	202		
						Acupuncture	305		
2	Endres 2007 [47]	Germany	Headache (TTH)	38.44 (11.77)	6 / 3	Sham acupuncture	200	SF-12	TTH days
						Acupuncture	209		
3	Jena 2008 [48]	Germany	Headache (headache)	43.66 (12.69)	6 / 3	Usual care	1613	SF-36	Headache days
						Acupuncture	1569		
4	Linde 2005 [44]	Germany	Headache (migraine)	42.55 (11.35)	6 / 3	Usual care	76	SF-36	Days of moderate to severe pain
						Sham acupuncture	81		
						Acupuncture	145		
5	Melchart 2005 [45]	Germany	Headache (TTH)	42.68 (13.18)	6 / 3	Usual care	75	SF-36	Headache days
						Sham acupuncture	62		
						Acupuncture	132		
6	Vickers 2004 [43]	UK	Headache (headache)	46.34 (10.39)	12 / 3	Usual care	161	SF-36	Severity score
						Acupuncture	140		
7	Brinkhaus 2006 [55]	Germany	Musculoske-letal (low back)	58.81 (9.13)	12 / 2	Usual care	79	SF-36	VAS pain score
						Sham acupuncture	73		
						Acupuncture	146		
8	Carlsson 2001 [50]	Sweden	Musculoske-letal (low back)	49.84 (15.4)	6 / 3	Sham acupuncture	16	VAS pain	VAS pain score
						Acupuncture	34		
9	Guerra 2004 [53]	Spain	Musculoske-letal (shoulder)	59.19 (11.37)	6 / 3	Sham acupuncture	65	VAS pain	VAS pain score
						Acupuncture	65		
10	Haake 2007 [61]	Germany	Musculoske-letal (low back)	50.15 (14.68)	6 / 3	Usual care	388	SF-12	Von Korff pain intensity score
						Sham acupuncture	387		
						Acupuncture	387		
11	Irnich 2001 [51]	Germany	Musculoske-letal (neck)	NA	3 / 3	Sham acupuncture	61	VAS pain	VAS pain score
						Acupuncture	56		
12	Kennedy 2008 [62]	Northern Ireland	Musculoske-letal (low back)	45.58 (11.1)	3 / 3	Sham acupuncture	24	VAS pain	Roland Morris disability score
						Acupuncture	24		
13	Kerr 2003 [52]	Northern Ireland	Musculoske-letal (low back)	NA	6 / 1	Sham acupuncture	20	VAS pain	VAS pain score
						Acupuncture	26		
14	Kleinhenz 1999 [49]	Germany	Musculoske-letal (shoulder)	NA	3 / 1	Sham acupuncture	27	CMS and predicted VAS pain	CMS
						Acupuncture	25		

Table 1 Main characteristics of the data and study outcomes used for analysis *(Continued)*

15	Salter 2006 [56]	UK	Musculoske-letal (neck)	47.71 (16.51)	3 / 3	Usual care	14	no mapping – EQ-5D available	Northwick Park pain score
						Acupuncture	10		
16	Thomas 2006 [58]	UK	Musculoske-letal (neck)	42.62 (10.71)	24 / 3	Usual care	80	no mapping – EQ-5D available	SF-36 bodily pain score
						Acupuncture	159		
17	Vas 2006 [57]	Spain	Musculoske-letal (neck)	46.73 (13.2)	6 / 1	Sham acupuncture	62	SF-36	VAS pain score
						Acupuncture	61		
18	Vas 2008 [63]	Spain	Musculoske-letal (shoulder)	55.68 (11.37)	12 / 3	Sham acupuncture	220	VAS pain	CMS
						Acupuncture	205		
19	White 2004 [54]	UK	Musculoske-letal (neck)	53.36 (15.61)	12 / 3	Sham acupuncture	65	VAS pain	VAS pain score
						Acupuncture	70		
20	Witt 2006 [59]	Germany	Musculoske-letal (neck)	50.57 (12.93)	6 / 3	Usual care	1698	SF-36	Neck pain and disability score
						Acupuncture	1753		
21	Witt 2006 [60]	Germany	Musculoske-letal (low back)	52.83 (13.33)	6 / 3	Usual care	1390	SF-36	Hanover functional ability score
						Acupuncture	1451		
22	Foster 2007 [41]	UK	Osteoarthritis of the knee	63.23 (8.81)	12 / 1	Usual care	116	VAS pain	WOMAC pain score
						Sham acupuncture	119		
						Acupuncture	117		
23	Berman 2004 [36]	USA	Osteoarthritis of the knee	65.46 (8.62)	6 / 2	Usual care	189	no mapping – EQ-5D available	WOMAC pain score
						Sham acupuncture	191		
						Acupuncture	190		
24	Scharf 2006 [39]	Germany	Osteoarthritis of the knee	62.81 (10.07)	6 / 3	Usual care	316	SF-12	WOMAC total score
						Sham acupuncture	365		
						Acupuncture	326		
25	Vas 2004 [37]	Spain	Osteoarthritis of the knee	67.04 (10.09)	3 / 3	Sham acupuncture	49	WOMAC total	WOMAC total score
						Acupuncture	48		
26	Williamson 2007 [42]	UK	Osteoarthritis of the knee	70.67 (8.94)	3 / 3	Usual care	61	WOMAC total	Oxford knee score
						Acupuncture	60		
27	Witt 2005 [38]	Germany	Osteoarthritis of the knee	64.01 (6.49)	12 / 2	Usual care	70	SF-36	WOMAC total score
						Sham acupuncture	75		
						Acupuncture	149		
28	Witt 2006 [40]	Germany	Osteoarthritis of the knee	61.2 (10.39)	6 / 3	Usual care	310	SF-36	WOMAC total score
						Acupuncture	322		

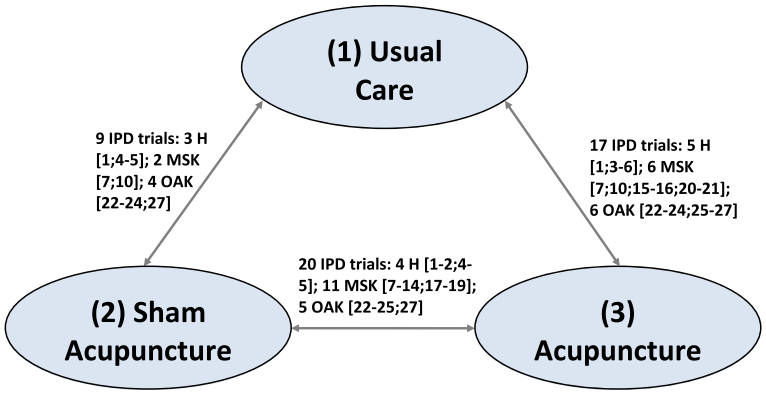

Fig. 1 Network of RCTs. Legend: In the network, a unique treatment category is indicated by a circle. Arrows between circles indicate that these treatments have been compared in a trial (trials are identified using '[]', numbered according to column 'ID' in Table 1. (Pain groups: H – Headache/ migraine; MSK – Musculoskeletal; OAK – Osteoarthritis of the knee)

Methods

Statistical models for the data

All analyses were conducted using Bayesian methods. A contrast-based modelling approach is taken throughout featuring relative treatment effects, in line with the parameterisation used by Lu and Ades [78], Saramago et al. [10] and others. A one-step modelling approach, where the likelihood for data at the IPD level and that of parameter estimates were described simultaneously, was preferred because we intended to explore treatment-by-covariate interactions at the patient-level [7, 79]. Note that all four models described below include pain type interactions which are specific to the current case-study. Table 2 summarises the key characteristics of the four models implemented, highlighting existing differences across these.

ANCOVA analysis (model 1)

The main modelling approach considered (model 1) is a variation of the ANCOVA approach that models the change score adjusting for baseline outcome values and with no stratification variables [19, 22, 80, 81] – such an approach is seen as equivalent to the existing ANCOVA approach.

The model considers a set of J studies for which IPD was available. The set of treatments included in these trials are labelled [A,B,C], where A is the reference treatment and there are K (=3) treatments in total. At baseline, patient i in study j allocated to treatment k provides a baseline measurement Y_{ijk0} (where 0 indicates time t at baseline). Each patient provides a follow-up measurement (the assessment closest to 3 months), Y_{ijk3}. The change from baseline ($Y_{ijk3} - Y_{ijk0}$) is denoted ΔY_{ijk} and is assumed normally distributed with mean θ_{ijk} and study-level variance of V_j.

θ_{ijk}, is assumed to be a function of μ_{jb}, the outcome for treatment b (the lowest indexed treatment in each study) in study j for a patient with a baseline utility of 0, Y_{ijk0}; δ_{jbk}, the study-specific treatment effect for treatment k relative to treatment b; and X_{jp}, p - 1 dummy variables representing pain type p in the j^{th} study. The latter terms were included to allow treatment effects to vary according to pain type (i.e. OAK; headache - including TTH and migraine; and musculoskeletal conditions - including lower back, shoulder and neck pain). There are different ways in which interaction effects can be specified in NMAs [82]. For this example we assumed that pain treatment interaction effects, β_{bkp}, were different for each treatment but exchangeable across treatments. Estimates of β_{bkp} were

Table 2 Summary of key characteristics of implemented models

	Model 1	Model 2	Model 3	Model 4
Outcome type	Continuous	Continuous	Continuous	Continuous
Outcome synthesised	Change from baseline	Change from baseline	Change from baseline	Final score
Approach	ANCOVA (baseline adjustment)	ANCOVA (baseline adjustment)	No baseline adjustment	No baseline adjustment
Pain interactions (case-study specific)	Yes	Yes	Yes	Yes
Further adjustments	None	Patients characteristics as treatment-effect modifiers	None	None

therefore assumed to be drawn from a random distribution with a common mean (B_p) and between treatment variance (σ^2_{Bp}). An exchangeable interaction approach for pain was thought to be the most appropriate as it allowed pain interactions to be different across treatments but related. Pain interaction effects were not included for OAK as this is used as the reference pain indication. Pain interaction terms were specific to the current application and may be excluded if not of interest. However, we emphasise that adjustment for baseline should always be included regardless of the need to model interactions.

A random treatment effect approach was taken due to the expected between-study heterogeneity, the variance of which is described as σ^2.

This model can be written as:

$$\begin{aligned}
&\Delta Y_{ijk} \sim N\left(\theta_{ijk}, V_j\right)\\
&\theta_{ijk} = \{\mu_{jb} + \beta_{0j} Y_{ijk0} \quad if\ \ k = b;\ \ b \in \{A,\ B,\ C, \ldots\}\\
&\qquad \mu_{jb} + \beta_{0j} Y_{ijk0} + \delta_{jbk} + \beta_{bkp} X_{jp} \quad if\ \ k > b\\
&\qquad \delta_{jbk} \sim N\left(d_{bk}, \sigma^2\right) \sim N\left(d_{Ak} - d_{Ab}, \sigma^2\right)\\
&\qquad \beta_{bkp} = \beta_{Akp} - \beta_{Abp} \quad \beta_{Akp} \sim N\left(B_p, \sigma^2_{Bp}\right)\ \ d_{AA}, \beta_{AAp} = 0
\end{aligned} \tag{1}$$

Prior distributions were defined independently as follows: $1/V_j \sim Gamma\ \ (0.001, 0.001)$; $\mu_{jb} \sim N(0, 10^6)$; $\beta_{oj} \sim N(0, 10^6)$; $d_{Ak} \sim N(0, 10^6)$; $\sigma \sim Unif(0, 2)$; $B_p \sim N(0, 10^6)$; $\sigma_{Bp} \sim Unif(0, 2)$. Correlations in the random effects from trials with three or more arms were accounted for using published methodology [3, 64]. In this paper, $k > b$ indicates that k is after b in the alphabet.

Controlling for treatment effect modifying patient-level characteristics (model 2)

For the EQ-5D endpoint, model 1 was extended to include patient-level covariates as potential treatment effect modifiers. Clinical expectations were that older age or higher body max index (BMI) may make patients more difficult to treat and, thus, potentially reduce the effect of treatment. Data on age were available from most studies and it was included as a covariate (centred) in the synthesis model. Again, a range of approaches can be used to incorporate treatment-effect interactions. In this analysis we assumed a common effect across pain types and for both acupuncture and sham acupuncture (i.e. a single interaction term is assumed to apply to all comparisons with usual care) [82] as this was deemed more clinically plausible. A non-linear effect of age was expected *a priori*, and thus squared terms were included for both main effects and treatment interaction effects. BMI data were only available in 10 of the 28 studies and for this reason we did not explore this variable further.

Model 2 thus differs from model 1 in that it considers the effects of the covariate Z (age). Differences to model 1 are shown below:

$$\begin{aligned}
\theta_{ijk} = \{&\mu_{jb} + \beta_{0j} Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 {Z_{ijk}}^2 \ \ if\ k = b;\ b \in \{A, B, C, \ldots\}\\
&\mu_{jb} + \beta_{0j} Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 {Z_{ijk}}^2 + \delta_{jbk}\\
&\quad + \beta_{bkp} X_{jp} \quad if\ k > b \text{ and } b \neq A\\
&\mu_{jb} + \beta_{0j} Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 {Z_{ijk}}^2 + \phi Z_{ijk} + \varphi {Z_{ijk}}^2 + \delta_{jbk}\\
&\quad + \beta_{bkp} X_{jp} \quad if\ k > b \text{ and } b = A Z_{ijk} \sim N\left(m, \sigma^2_Z\right)
\end{aligned} \tag{2}$$

Coefficients on the main covariate effect and the effect squared are represented by ϕ_0 and φ_0. Coefficients on the treatment-by-covariate interaction term and the interaction between treatment and the squared covariate term are represented by ϕ and φ. No interaction term for comparisons of k and b were included when $b \neq A$ because the common regression coefficient cancels out.

Due to the possibility of missing covariate information for some individuals in some studies, Z_{ijk} was assumed to be a normally distributed random variable with mean m and variance σ^2_Z, common across all IPD studies. This represents a Markov chain Monte Carlo (MCMC) multiple imputation technique which generates independent draws of the missing data from its predictive distribution assuming MAR covariate data. Additional priors were required for this model: $\phi_0,\ \phi,\ \varphi_0,\ \varphi \sim N(0, 10^6)$; $m \sim Unif(-50, 50),\ \sigma_z \sim Unif(0, 30)$

Analysis with restricted evidence (model 3 and 4)

Although model 1 is the preferred choice, this model would not be feasible in the absence of outcome information at the individual-level for both baseline and follow-up time points. Sub-optimal models which do not rely on the availability of IPD were therefore run for comparison purposes. Three options are typically available to the analyst when only AD are available [22] – i) in the event of ANCOVA estimates being available, synthesise these using published literature [22]; or ii) model the change score without baseline adjustment (model 3); or iii) model the final outcome score without baseline adjustment (model 4). We note that – though suboptimal - model 3 has also been presented in the context of an NMA of continuous outcomes when IPD were available [15, 16].

Models 3 and 4 are simplifications of model 1 where the baseline outcome variable is omitted and where model 4 considers the final score, rather than the change score, as dependent variable. The synthesis of data using models that ignore baseline outcomes may provide biased treatment effect estimates because of potential baseline imbalances (unless addressed within trials themselves) and due to ignoring potential correlation between the change/final score and the baseline value [77, 83]. It may also reduce

the precision of treatment effect estimates, even if balance at baseline is observed across all synthesised evidence [22].

Calculating the residual deviance

The total residual deviance, *TRD* – a measure of model fit - can be estimated for each of the described models by summing study-level residual deviances, *RD*. Study-level *RDs* are the ratio of the sum across studies of the squared differences between the observed changes from baseline, ΔY_{ijk}, and the estimated mean, θ_{ijk}, divided by the study-level variance, V_j [84]:

$$D_{ijk} = \left(\Delta Y_{ijk} - \theta_{ijk}\right)^2 RD_j = sum\left(D_j\right)/V_j TRD = sum(RD) \quad (3)$$

For a model that fits the data well, it is assumed that the contributions to the *RD* to have a chi-squared distribution with *N* degrees of freedom if a sum over *N* unconstrained data points is made. On this basis, it is expected that the posterior mean of the *TRD* should be close to the number of unconstrained data points if the model predictions are a good fit to the data [20, 84, 85].

Model selection and implementation

Data management was performed in the freely available software package R version 3.0.0 (Copyright © 2013 The R Foundation for Statistical Computing [86]). The NMA analyses were undertaken in WinBUGs [87] version 1.4.3 (Copyright © 2008 Medical Research Council (UK) and Imperial College (UK)), linked to the R software through the packages R2WinBUGS [88] and CodaPkg [89]. Annotated code, sample data and initial values for model 1 are provided in the Additional file 2 to allow readers to adapt it for their own purposes.

In all models the MCMC Gibbs sampler was initially run for 10,000 iterations and these were discarded as 'burn-in'. Models were run for a further 5,000 iterations, on which inferences were based. Chain convergence was checked using autocorrelation and Brooks-Gelman-Rubin diagram [90, 91] diagnostics. Goodness of fit was assessed using the deviance information criterion (DIC) and TRD [84]. Results are presented as EQ-5D preference scores and SMD treatment effect estimates (and associated 95 % credibility intervals, CrIs) and also using the probability of treatment being the 'best' treatment in terms of being the most clinically effective [4].

Results

ANCOVA analysis results (model 1)

Table 2 and Fig. 2 show the evidence from model 1 on relative treatment effect estimates adjusted for baseline and treatment-by-pain interaction effects (medians of the MCMC posterior samples and 95 % CrI shown). Measures of model fit (TRD and DIC) are also shown. The reference category for the pain interaction effects is the OAK pain type.

For both endpoints, model 1 indicates that acupuncture treatment increases the HRQoL of patients and/or reduces pain more than usual care and sham acupuncture treatments, irrespective of pain group. For the EQ-5D endpoint the treatment effect of acupuncture *vs.* usual care in the OAK population is 0.079 (median, 95 % CrI: 0.042 to 0.114), for headache/migraine and musculoskeletal pain patients the comparable treatment effects are 0.056 (median, 95 % CrI: 0.021 to 0.092) and 0.082 (median, 95 % CrI: 0.047 to 0.116), respectively. The results also favour acupuncture over sham acupuncture, although with a greater degree of uncertainty, as reflected by the fact that CrIs include zero for all pain types (OAK: 0.022, 95 % CrI −0.014 to 0.060; headache/migraine: 0.004, 95 % CrI −0.035 to 0.042; and musculoskeletal 0.023, 95 % CrI −0.008 to 0.053). The probability that acupuncture is the best treatment at improving HRQoL is 0.89 for OAK, 0.64 for headache/migraine and 0.95 for musculoskeletal pain.

Results for the SMD endpoint followed a similar pattern. However, in contrast to the EQ-5D analysis, in the latter comparison the CrIs do not include zero in the standardised pain analysis for OAK (0.438, 95 % CrI 0.121 to 0.715) and musculoskeletal (0.527, 95 % CrI 0.323 to 0.735) pain types, though they do for headache/migraine (0.256, 95 % CrI −0.073 to 0.560). The probability that acupuncture is the best treatment at improving standardised pain is 0.96 to 1.00 depending on pain type.

It is interesting to note that sham acupuncture *vs.* usual care treatment effect 95 % CrIs across pain types do not include *0* in the EQ-5D endpoint but they do for SMD, except for the headache/migraine group. These results suggest that sham acupuncture effects may well go beyond pain. Also interesting is the estimated magnitude of the uncertainty over the pain type interactions (not reported) as these, particularly for the EQ-5D endpoint, do not provide strong evidence of a difference between pain types.

Expectations were that some level of heterogeneity existed between-trials. Possibly as a consequence of the mapping work performed, this expectation was not fulfilled for the EQ-5D endpoint (the between-study variance estimate is 0.001). For the SMD endpoint the between study variance was also small relative to the magnitude of the treatment effects (the between-study variance estimate is 0.09). The TRD suggests that the models provide an adequate fit to the data (see Table 3).

Controlling for patient-level characteristics (model 2)

Table 1 provides information on age for each of the trials included in the dataset. The average age was lower in the

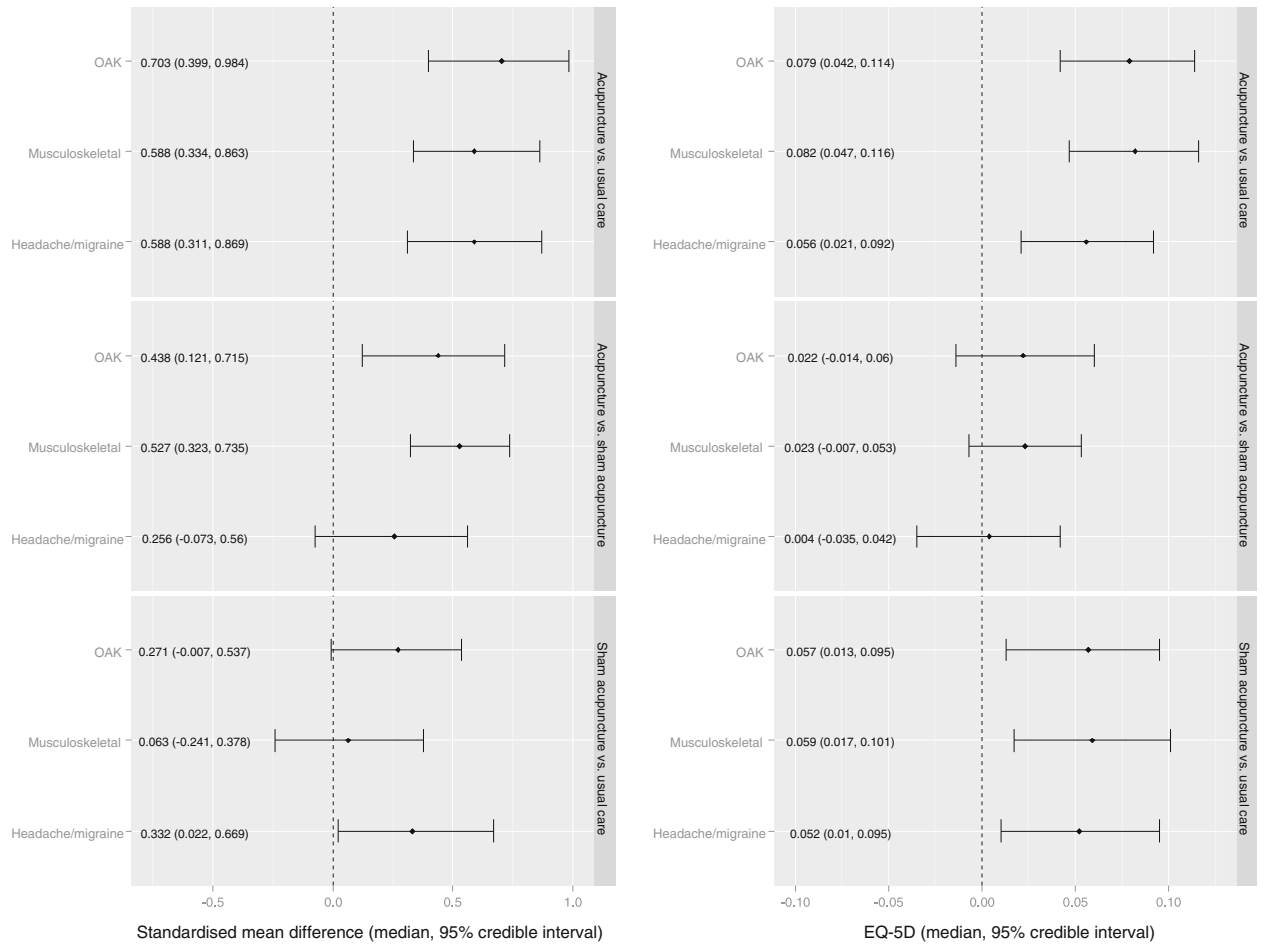

Fig. 2 Forest plot showing network meta-analysis results for standardised pain and EQ-5D outcomes

Table 3 IPD NMA ANCOVA synthesis model results (model 1), EQ-5D preference score and standardised pain endpoints

IPD NMA ANCOVA results: EQ-5D preference score and SMD endpoints[a]			Model 1, ANCOVA, change in outcome score, adjusted for baseline median MCMC posterior sample (95 % CrI)	
			Change EQ-5D	Change standardised pain
Relative treatment effects	*Osteoarthritis of the knee*	*SHAM vs UC*	0.057 (0.013, 0.095)	0.271 (-0.007, 0.537)
		ACU vs UC	0.079 (0.042, 0.114)	0.703 (0.399, 0.984)
		ACU vs SHAM	0.022 (-0.014, 0.060)	0.438 (0.121, 0.715)
	Headache	*SHAM vs UC*	0.052 (0.010, 0.095)	0.332 (0.022, 0.669)
		ACU vs UC	0.056 (0.021, 0.092)	0.588 (0.311, 0.869)
		ACU vs SHAM	0.004 (-0.035, 0.042)	0.256 (-0.073, 0.560)
	Musculoskeletal	*SHAM vs UC*	0.059 (0.017, 0.101)	0.063 (-0.241, 0.378)
		ACU vs UC	0.082 (0.047, 0.116)	0.588 (0.334, 0.863)
		ACU vs SHAM	0.023 (-0.008, 0.053)	0.527 (0.323, 0.735)
	Between-study variance		0.001 (0, 0.003)	0.090 (0.049, 0.170)
	Total residual deviance[b]		15,850 (15,480; 16, 230)	17,060 (16,660; 17,450)
	Deviance information criterion[c]		-6,420.4	37,394.2

[a]*UC* usual care, *SHAM* sham acupuncture, *ACU* acupuncture, *Headache group* headache, migraine and TTH, *Musculoskeletal group* neck, shoulder and low back pain
[b]For the EQ-5D endpoint models used approx. 14800 observations; for the SMD endpoint models used approx. 15900. Models should be preferred when total residual deviance mean posterior is close to the actual number of data points
[c]Deviance information criterion (DIC) is a statistical measure of model fit and model comparison. Models with smaller DIC are preferred

headache/migraine pain group than in the musculoskeletal group, which in turn was lower than the OAK group.

Using the change in EQ-5D as the outcome for synthesis, Table 4 presents the results of applying model 2 (an extension of model 1) to include patient-level information on age – with age considered as a potential treatment effect modifier. The model fit statistics show that the adjusted by age model is marginally better than model 1, providing lower DIC statistics and reduced posterior RD. The results are very similar to model 1 and do not suggest age is a strong effect modifier or that non-linear effects of age on the effect of treatments are present.

Analysis with restricted evidence (model 3 and 4)

Models 3 and 4 model the change score and the final outcome score, respectively. These are seen as simplifications of model 1 where no baseline adjustment is done. Results for models 3 and 4 are presented in Table 5, together with model 1 results for comparison. Generally, all three models convey the same message in relation to which treatment provides higher increases in patients' HRQoL; that is, acupuncture is found to be better than sham and usual care treatments. As expected, models 3 and 4 (model 3 in particular) provide different summary results of treatment effects when compared to model 1. Compared with the ANCOVA model (model 1), model 3, the change score approach, generally inflates the summary treatment effects across pain types, with potential losses in precision (e.g. for OAK the median EQ-5D treatment effect is inflated 19 % in model 3 compared to model 1 for the acupuncture *vs* usual care comparison). Compared to model 1, model 4 summary treatment effects are generally similar or lower; CrIs are however consistently wider in model 4 compared to model 1.

Table 4 IPD NMA ANCOVA synthesis model (model 2) results with adjustments, EQ-5D preference endpoint

IPD NMA results: EQ-5D preference scores endpoint[a]			Model 2, ANCOVA, with adjustment for baseline score, age and treatment-by-age interactions, median MCMC posterior sample (95 % CrI)
Relative treatment effects	*Osteoarthritis of the knee*	*SHAM vs UC*	0.040 (-0.006, 0.084)
		ACU vs UC	0.066 (0.025, 0.105)
		ACU vs SHAM	0.026 (-0.012, 0.066)
	Headache	*SHAM vs UC*	0.056 (0.012, 0.098)
		ACU vs UC	0.060 (0.023, 0.095)
		ACU vs SHAM	0.004 (-0.036, 0.043)
	Musculoskeletal	*SHAM vs UC*	0.045 (-0.001, 0.094)
		ACU vs UC	0.074 (0.038, 0.109)
		ACU vs SHAM	0.029 (-0.009, 0.067)
Main effects	*Age*		-0.002 (-0.002, -0.001)
	Age^2		0.000 (0.000, 0.000)
Age common interactions	*Age*		0.000 (0.000, 0.001)
	Age^2		0.000 (0.000, 0.000)
	Between-study variance		0.001 (0.000,0.003)
	Total residual deviance[b]		15,590 (15,210; 15,970)
	Deviance information criterion[c]		-6,462.0

[a]*UC* usual care, *SHAM* sham acupuncture, *ACU* acupuncture, *Headache group* headache, migraine and TTH, *Musculoskeletal group* neck, shoulder and low back pain

[b]Compare to approx. 14, 800 observations

[c]Deviance information criterion (DIC) is a statistical measure of model fit and model comparison. Models should be preferred with smaller DIC

Discussion

This study presents methods for conducting NMA of IPD on continuous outcomes, building on previous work on ANCOVA models for pairwise meta-analysis [22]. IPD availability avoided the use of non-baseline-adjusted models, allowing for ANCOVA models to be applied, thus improving precision of treatment effect estimates while adjusting for baseline imbalance [22]. Our results generalise the findings from Riley et al. [22] to the NMA setting and reinforce the idea that different approaches to the synthesis of continuous outcomes will produce different results. The ANCOVA approach is advocated to be the most appropriate modelling approach. Due to limited reporting of ANCOVA results in trial publications, IPD will typically be required to facilitate implementation of the ANCOVA NMA approach. The appropriate analysis of continuous endpoints therefore provides a further rationale for obtaining access to IPD, in addition to those well documented in the NMA literature [10, 12, 15, 92].

Recent work by Hong et al. [15] and Thom et al. [16] presented and discussed IPD NMA models for continuous outcomes. While Hong and colleagues [15] introduced contrast-based and arm-based models for multiple outcomes, Thom et al. [16] synthesised AD and IPD, some of which was observational rather than RCT data. They also considered interactions between treatment effects and covariates. The existence of ecological bias was explored in Hong et al. [15] by partitioning within- and across-study interactions [10]. Both publications used the change from baseline as their continuous outcome measure. In both publications models were presented that did not incorporate an adjustment for baseline outcome values, and in Hong et al. [15] adjustment for baseline outcome values was only considered in the context of modelling baseline outcomes as a treatment effect modifier. Thom et al. [16] recognised that the approach taken was not the recommended one, but noted that an ANCOVA-type approach was not possible

Table 5 IPD NMA results for models (1), (3) and (4), EQ-5D preference score endpoint

IPD NMA results: EQ-5D preference scores endpoint[a]			Model 1, ANCOVA, change in EQ-5D scores, adjusted for baseline	Model 3, change in EQ-5D scores, without baseline adjustment	Model 4, follow-up EQ-5D score, without baseline adjustment
			Median MCMC posterior sample (95 % CrI)	Median MCMC posterior sample (95 % CrI)	Median MCMC posterior sample (95 % CrI)
Relative treatment effects	*Osteoarthritis of the knee*	*SHAM vs UC*	0.057 (0.013, 0.095)	0.077 (0.033, 0.118)	0.051 (0.008, 0.094)
		ACU vs UC	0.079 (0.042, 0.114)	0.093 (0.054, 0.129)	0.074 (0.035, 0.113)
		ACU vs SHAM	0.022 (-0.014, 0.060)	0.016 (-0.022, 0.054)	0.023 (-0.014, 0.065)
	Headache	*SHAM vs UC*	0.052 (0.010, 0.095)	0.044 (0.002, 0.086)	0.052 (0.007, 0.098)
		ACU vs UC	0.056 (0.021, 0.092)	0.057 (0.023, 0.090)	0.054 (0.016, 0.092)
		ACU vs SHAM	0.004 (-0.035, 0.042)	0.013 (-0.025, 0.051)	0.002 (-0.038, 0.040)
	Musculoskeletal	*SHAM vs UC*	0.059 (0.017, 0.101)	0.062 (0.019, 0.104)	0.054 (0.010, 0.099)
		ACU vs UC	0.082 (0.047, 0.116)	0.084 (0.048, 0.119)	0.080 (0.044, 0.118)
		ACU vs SHAM	0.023 (-0.008, 0.053)	0.022 (-0.011, 0.055)	0.026 (-0.006, 0.056)
	Between-study variance		0.001 (0, 0.003)	0.001 (0, 0.003)	0.001 (0, 0.003)
	Total residual deviance[b]		15,850 (15,480; 16,230)	16,990 (16,570; 17,420)	15,370 (15,010; 15,730)
	Deviance information criterion[c]		-6,420.4	-69.9	-3,823.7

[a]*UC* usual care, *SHAM* sham acupuncture, *ACU* acupuncture, *Headache group* headache, migraine and TTH, *Musculoskeletal group* neck, shoulder and low back pain, *OAK* osteoarthritis of the knee
[b]Compare to approx. 14,800 observations
[c]Deviance information criterion (DIC) is a statistical measure of model fit and model comparison. Models should be preferred with smaller DIC

as, for most studies in their motivating example, only AD was accessible to them. Our work emphasises that where IPD is available, all models of continuous outcomes should include adjustment for the baseline outcome, and unadjusted models should not be presented.

Analyses in this paper were conducted to explore the implications of using non-ANCOVA models in a NMA framework, as other methods have been used in the literature [15] to analyse continuous outcome IPD, and these methods are often necessary in the absence of IPD. The results showed some differences with the ANCOVA results. Modelling final scores or change scores without baseline adjustment produced estimates of treatment effect which differed by up to 19 % compared to the baseline adjusted model. By explicitly accounting for correlation between the change score and the baseline score in the presence of baseline imbalance, the tested ANCOVA model (model 1) avoids bias in the pooled treatment effect estimates. These results emphasise how important it is to adjust for baseline to adequately synthesise evidence in this setting; tasks very much facilitated with the availability of IPD. We hope that by highlighting the consequence of using suboptimal model(s) may encourage readers to obtain IPD so that the most appropriate methods may be implemented. When IPD is available ANCOVA should always be used. There has been a discussion in the literature about the fact that final or change score analyses may 'bound' the true relative effect estimate. Although this may be true for a single trial, it may not hold for NMA models [18]. This emphasises the importance of conducting appropriate analyses as the potential direction of bias is difficult to predict. Any bias in treatment effect or impact on precision could lead to inappropriate decisions regarding adoption and further research.

The motivating example related to the effectiveness of acupuncture for the treatment of chronic pain. The analyses found acupuncture to be more effective than usual care with respect to reducing pain and improving EQ-5D preference scores in patients with chronic pain of OAK, musculoskeletal and headache/migraine origin. The benefits of acupuncture over sham acupuncture are smaller than when compared to usual care. The methods used provided outputs in a format that can be used to directly inform CE considerations once the full set of relevant comparators are considered.

A recent study by Vickers et al. [35] also explored the effectiveness of acupuncture for chronic pain. This study performed an IPD pair-wise meta-analysis using the same data plus data from an additional trial [93] – data which, due to lack of consent, was not available to be used in the current analysis. Using study-specific primary outcome measures and the ANCOVA methodology, the Vickers et al. [35] study conducted meta-analyses separately for comparisons of acupuncture with sham acupuncture and usual care, and within each pain type. Despite the methodological differences, and differences for some trials in choice of primary outcome measure and/or primary end point, the authors' findings are similar.

The instruments used to measure health outcomes differed between trials. Standardisation and mapping

approaches were used to derive, pain-related outcomes and EQ-5D, respectively. Analysis of the pain outcome required development of methods for conducting standardised mean difference analysis with IPD. Analysis of the EQ-5D data required an extensive mapping exercise whereby separate mapping functions were applied to each study, with choice of mapping dependent on the available outcome data. Access to IPD in this context also avoided the use of any assumptions regarding the distribution of HRQoL instrument scores – thus allowing the observed distributions to be adequately reflected in the mapped EQ-5D estimates.

This study has a number of limitations. The applicability of these methods is conditional on the access to IPD. If IPD is not available or is partially available, other methods need to be used and limitations stressed. Often a mixture of IPD and AD is available – anecdotally a 50 % success rate of obtaining IPD is attained in the academic world, lower success rates may be achieved elsewhere, where it is common to have, for instance, only a company's own RCT data and not that for competitor interventions. In the context of continuous outcomes the advantages of access to IPD are significant and efforts to share data should be pursued. As access to IPD for all studies in all NMAs is likely to be unrealistic in the medium-term, it would be useful to have available a methodology which had the advantages of the ANCOVA approach but could be used when only some (or even no) studies in the database were available in IPD form.

Additionally there are a series of limitations related to the case study. Firstly, the synthesis of heterogeneous outcomes relied on imperfect standardisation processes (which assume that any differences in within trial outcome variability are due to the use of different instruments) and mappings which are typically able to explain only a minority of variation in EQ-5D. The availability of key outcomes across trials would have reduced these concerns, as would the collection of generic preference based measures of HRQoL in all trials. Also, the outcome data closest to 3 months were selected for synthesis. For some trials, the nearest reported outcome data were at only months 1 or 2. If the effect of acupuncture increases gradually, these effects may underestimate 3 month outcomes. Furthermore, some of the trials show increased benefits of acupuncture over comparators at 12 [94] and 24 months [95] compared to 3 months. This evidence may be an indication of the long-term clinical benefits of acupuncture and has implications for estimating long-term HRQoL and CE. Collection of trial data for more than 3 months is therefore warranted together with further work analysing repeated outcome measurements in a NMA to evaluate the importance of these effects.

A complete-case analysis was conducted. This approach to missing data has been thoroughly documented in the methods literature as not being optimal as it can lead to bias if observations with missing values systematically differ from the complete cases and may inflate standard errors due to the reduction in sample size. Some recent work has been done in this area [96], although it does not consider the case where IPD is available. Finally, another potential issue for future exploration is that the impact of each pain condition on treatment effects was assumed to be exchangeable [82]; this assumption could be explored further by comparing different assumptions over the inclusion of the interaction effects, or even with the inclusion of no interaction effects. In summary, a worthwhile extension to this work would be to develop a multivariate ANCOVA modelling framework considering both multiple endpoints and time points, missing data and which enables relevant aggregate data to be included, building on recent work [15, 97–101].

Conclusion

In conclusion, this paper has reiterated the importance of accessing and analysing IPD and presented methods to fully exploit the benefits of access to this data in the context of continuous outcomes. Methods for conducting ANCOVA IPD NMA of continuous outcomes are presented and discussed. The methods developed are applicable to contexts in which endpoints are reposted consistently and to contexts in which outcome measures differ across trials. Given the demonstrable benefits of access to IPD, we suggest that more effort should be made to share and develop repositories for data in this format [102].

Endnotes

[1]Considering s^{t}_{tx} as the standardised value of the pain measurement p made at the time point t in patients under treatment tx, it can be demonstrated that $(S^{t1}_{tx1} - S^{t0}_{tx1}) - (S^{t1}_{tx0} - S^{t0}_{tx0}) = (S^{t1}_{tx1} - S^{t1}_{tx0}) - (S^{t0}_{tx1} - S^{t0}_{tx0}) = \Delta SMD$

[2]The selection of the outcome to be mapped was not at random. Preference was given to generic preference-based instruments (i.e. SF-12 and SF-36) and, in its absence, to condition-specific ones (i.e. WOMAC, VAS pain and CMS), conditional on the existence of a valid and published algorithm. WOMAC was used in preference to VAS pain and CMS as it covers a broader definition of HRQoL.

[3]A random effects generalised least squares algorithm considering dimensions, dimensions squared and interactions from Rowen et al. [71] was used. A multinomial logit using PCS and MCS summary scores, summary scores squared and interaction terms from Gray et al. [70] was used.

[4]An OLS including total WOMAC score, total WOMAC squared, age and gender as covariates from Barton et al. [73] was used. An OLS including VAS pain and VAS pain squared as covariates from Maund et al. [72] was used.

[5]A mapping process involves additional sources of uncertainty - the uncertainty in the mapping function regression coefficients and the structure of the mapping model. These additional sources of uncertainty are not accounted for in this analysis.

Additional files

Additional file 1: Provides summary information of mapped EQ-5D and standardised pain data by study. (DOCX 105 kb)

Additional file 2: Provides WinBUGS modelling code for models 1 and 2 as described in the main text. (DOCX 41 kb)

Abbreviations

AD: Aggregate data; ANCOVA: Analysis of covariance; ATC: Acupuncture Trialists' Collaboration; BMI: Body max index; CE: Cost-effectiveness; CMS: Constant Murley Score; CrI: Credibility interval; DIC: Deviance information criterion; EQ-5D: EuroQol five-dimensional; HRQoL: Health-related quality of life; IPD: Individual patient data; MAR: Missing at random; MCMC: Markov chain Monte Carlo; MCS: Mental component summary; NICE: National Institute for Health and Care Excellence; NIHR: National Institute for Health Research; NMA: Network meta-analysis; OAK: Osteoarthritis of the knee; OLS: Ordinary least squares; PCS: Physical component summary; QALY: Quality-adjusted life-year; RD: Residual deviance; SF-12: Short Form 12 dimensions; SF-36: Short Form 36 dimensions; SMD: Standardised mean difference; TRD: Total residual deviance; TTH: Tension-type headache; UK: United Kingdom; VAS: Visual analogue scale; WOMAC: Western Ontario and McMaster Universities Arthritis Index

Acknowledgements

The views expressed in this presentation are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The ATC is funded by an R21 [AT004189I] from the National Center for Complementary and Alternative Medicine (NCCAM) at the National Institutes of Health (NIH) to AV and by a grant from the Samueli Institute. AM's contribution was made under the terms of a Career Development research training fellowship issued by the NIHR [grant CDF-2009-02-21]. The study sponsors had no role in the study design; collection, analysis and interpretation of the data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Funding

This work was supported by the National Institute for Health Research (NIHR) under Programme Grants for Applied Research [Grant No. RP-PG-0707-10186].

Availability of data and materials

This study represents a meta-analysis of data collected by authors other than those named on the authorship list. Where this data is publically available it is included in the supplementary material. The authors are unable to give unrestricted access to data that they did not collect. In particular, some of the trialists that gave data did not do so with ethical approvals that allowed unrestricted data sharing.
The WinBUGS model code relating to the newly developed modelling approach is provided in the additional supporting files.

Authors' contributions

All authors made a substantial contribution to this work. PS, lead analyst and main author of this work, extensively contributed to all aspects of the statistical methods development, analysis and write up. BW, contributed to all aspects of the methods development, analysis and write. HW, co-ordinated the research and contributed to the analysis and write up. AM, co-ordinated the research and advised on aspects of the methods development and write up. KK, contributed to the analysis and write up. MS, co-ordinated the research and advised on aspects of the methods development and write up. AJV, advised on aspects of the methods development and write up. HM, co-ordinated the research and advised on aspects of the analysis and write up. All authors read and approved the final manuscript.

Competing interests

HM reports and states that he practices acupuncture part-time and is a Trustee of the not-for-profit organisation, the Northern College of Acupuncture. None of the other authors have competing interests for this work.

Consent to publish

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

[1]Centre for Health Economics, University of York, York, UK. [2]Warwick Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK. [3]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, USA. [4]Department of Health Sciences, University of York, York, UK.

Received: 20 May 2016 Accepted: 9 September 2016
Published online: 06 October 2016

References

1. Lu G, Ades A, Sutton A, Cooper N, Briggs A, Caldwell D. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med. 2007;26(20):3681–99.
2. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002;21(16):2313–24.
3. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23(20):3105–24.
4. Ades A, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu G. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics. 2006;24(1):1–19.
5. Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. J Clin Epidemiol. 2007;60(5):431–9.
6. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med. 2008;27(29):6111–36.
7. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutitie F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27(11):1870–93.
8. Sutton A, Kendrick D, Coupland C. Meta-analysis of individual- and aggregate-level data. Stat Med. 2008;27(5):651–69.
9. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. Brit Med J. 2010;340:c221.
10. Saramago P, Sutton A, Cooper N, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Stat Med. 2012;31(28):3516–36.
11. Dias S, Welton N, Caldwell D, Ades A. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29(7–8):932–44.
12. Donegan S, Williamson P, D'Alessandro U, Smith CT. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. Stat Med. 2012;31(29):3840–57.
13. Saramago P, Manca A, Sutton A. Deriving input parameters for cost-effectiveness modeling: Taxonomy of data types and approaches to their statistical synthesis. Value Health. 2012;15(5):639–49.
14. Saramago P, Chuang LH, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. BMC Med Res Methodol. 2014;14:105.
15. Hong H, Fu H, Price KL, Carlin BP. Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. Stat Med. 2015;34(20):2794–819.
16. Thom HH, Capkun G, Cerulli A, Nixon RM, Howard LS. Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. BMC Med Res Methodol. 2015;15:34.
17. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.0.1 [updated

September 2008]. The Cochrane Collaboration; 2008. Available from www.handbook.cochrane.org.
18. Fu R, Vandermeer BW, Shamliyan TA, O'Neil ME, Yazdi F, Fox SH, Morton SC. Handling Continuous Outcomes. In: Quantitative Synthesis. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-2007-10057-I.) AHRQ Publication No. 13-EHC103-EF. Rockville, MD: Agency for Healthcare Research and Quality. July 2013. Available from www.effectivehealthcare.ahrq.gov/reports/final.cfm
19. Vickers A, Altman D. Analysing controlled trials with baseline and follow up measurements. Brit Med J. 2001;323(7321):1123–4.
20. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. NICE; 2011. Available from http://www.nicedsu.org.uk/.
21. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions, Version 5.1.0 [updated March 2011]. edn. Oxford: Wiley-Blackwell; 2011.
22. Riley R, Kauser I, Bland M, Thijs L, Staessen J, Wang J, Gueyffier F, Deeks J. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. Stat Med. 2013;32(16):2747–66.
23. Scott D, Boye K, Timlin L, Clark J, Best J. A network meta-analysis to compare glycaemic control in patients with type 2 diabetes treated with exenatide once weekly or liraglutide once daily in comparison with insulin glargine, exenatide twice daily or placebo. Diabetes Obes Metab. 2013;15(3):213–23.
24. Cooper N, Sutton A, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. Arch Intern Med. 2006;166(12):1269–75.
25. Nikolakopoulou A, Chaimani A, Veroniki A, Vasiliadis H, Schmid C, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. Plos One. 2014;9(1):e86754.
26. NICE. NICE guideline on low back pain: early management of persistent non-specific low back pain. 2009.
27. NICE. NICE guideline on diagnosis and management of headaches in young people and adults. 2012.
28. NICE. NICE guideline on osteoarthritis: the care and management of osteoarthritis in adults. 2014.
29. Latimer N. NICE guideline on osteoarthritis: is it fair to acupuncture? Yes. Acupunct Med. 2009;27(2):72–5.
30. Latimer NR, Bhanu AC, Whitehurst DG. Inconsistencies in NICE guidance for acupuncture: reanalysis and discussion. Acupunct Med. 2012;30(3):182–6.
31. Cummings M. Why recommend acupuncture for low back pain but not for osteoarthritis? A commentary on recent NICE guidelines. Acupunct Med. 2009;27(3):128–9.
32. White A. NICE guideline on osteoarthritis: is it fair to acupuncture? No. Acupunct Med. 2009;27(2):70–2.
33. MacPherson H, Vickers A, Bland B, Torgerson D, Corbett M, Spackman E, Saramago P, Woods B, Sculpher M, Manca A, et al. Acupuncture for chronic pain and depression in primary care: a programme of research. NIHR J. 2016. (in press).
34. Vickers AJ, Cronin AM, Maschino AC, Lewith G, Macpherson H, Victor N, Sherman KJ, Witt C, Linde K. Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Trialists' Collaboration. Trials. 2010;11:90.
35. Vickers AJ, Cronin AM, Maschino AC, Lewith G, MacPherson H, Foster NE, Sherman KJ, Witt CM, Linde K. Acupuncture for chronic pain: individual patient data meta-analysis. Arch Intern Med. 2012;172(19):1444–53.
36. Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. Ann Intern Med. 2004;141(12):901–10.
37. Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM, Borge MA, Gaspar O, Sanchez-Rodriguez F, Aguilar I, et al. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. BMJ. 2004;329(7476):1216.
38. Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, Hummelsberger J, Walther HU, Melchart D, Willich SN. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. Lancet. 2005;366(9480):136–43.
39. Scharf HP, Mansmann U, Streitberger K, Witte S, Kramer J, Maier C, Trampisch HJ, Victor N. Acupuncture and knee osteoarthritis: a three-armed randomized trial. Ann Intern Med. 2006;145(1):12–20.
40. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. Arthritis Rheum. 2006;54(11):3485–93.
41. Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, Hay EM. Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomised controlled trial. BMJ. 2007;335(7617):436.
42. Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. Rheumatology (Oxford). 2007;46(9):1445–9.
43. Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, Fisher P, Van Haselen R, Wonderling D, Grieve R. Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis. Health Technol Assess. 2004;8(48):iii. 1–35.
44. Linde K, Streng A, Jurgens S, Hoppe A, Brinkhaus B, Witt C, Wagenpfeil S, Pfaffenrath V, Hammes MG, Weidenhammer W, et al. Acupuncture for patients with migraine: a randomized controlled trial. JAMA. 2005;293(17):2118–25.
45. Melchart D, Streng A, Hoppe A, Brinkhaus B, Witt C, Wagenpfeil S, Pfaffenrath V, Hammes M, Hummelsberger J, Irnich D, et al. Acupuncture in patients with tension-type headache: randomised controlled trial. BMJ. 2005;331(7513):376–82.
46. Diener HC, Kronfeld K, Boewing G, Lungenhausen M, Maier C, Molsberger A, Tegenthoff M, Trampisch HJ, Zenz M, Meinert R. Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial. Lancet Neurol. 2006;5(4):310–6.
47. Endres HG, Bowing G, Diener HC, Lange S, Maier C, Molsberger A, Zenz M, Vickers AJ, Tegenthoff M. Acupuncture for tension-type headache: a multicentre, sham-controlled, patient-and observer-blinded, randomised trial. J Headache Pain. 2007;8(5):306–14.
48. Jena S, Witt CM, Brinkhaus B, Wegscheider K, Willich SN. Acupuncture in patients with headache. Cephalalgia. 2008;28(9):969–79.
49. Kleinhenz J, Streitberger K, Windeler J, Gussbacher A, Mavridis G, Martin E. Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendinitis. Pain. 1999;83(2):235–41.
50. Carlsson CP, Sjolund BH. Acupuncture for chronic low back pain: a randomized placebo-controlled study with long-term follow-up. Clin J Pain. 2001;17(4):296–305.
51. Irnich D, Behrens N, Molzen H, Konig A, Gleditsch J, Krauss M, Natalis M, Senn E, Beyer A, Schops P. Randomised trial of acupuncture compared with conventional massage and "sham" laser acupuncture for treatment of chronic neck pain. BMJ. 2001;322(7302):1574–8.
52. Kerr DP, Walsh DM, Baxter D. Acupuncture in the management of chronic low back pain: a blinded randomized controlled trial. Clin J Pain. 2003;19(6):364–70.
53. de Hoyos JA G, Andres Martin Mdel C, Bassas y Baena de Leon E, Vigara Lopez M, Molina Lopez T, Verdugo Morilla FA, Gonzalez Moreno MJ. Randomised trial of long term effect of acupuncture for shoulder pain. Pain. 2004;112(3):289–98.
54. White P, Lewith G, Prescott P, Conway J. Acupuncture versus placebo for the treatment of chronic mechanical neck pain: a randomized, controlled trial. Ann Intern Med. 2004;141(12):911–9.
55. Brinkhaus B, Witt CM, Jena S, Linde K, Streng A, Wagenpfeil S, Irnich D, Walther HU, Melchart D, Willich SN. Acupuncture in patients with chronic low back pain: a randomized controlled trial. Arch Intern Med. 2006;166(4):450–7.
56. Salter GC, Roman M, Bland MJ, MacPherson H. Acupuncture for chronic neck pain: a pilot for a randomised controlled trial. BMC Musculoskelet Disord. 2006;7:99.
57. Vas J, Perea-Milla E, Mendez C, Sanchez Navarro C, Leon Rubio JM, Brioso M, Garcia Obrero I. Efficacy and safety of acupuncture for chronic uncomplicated neck pain: a randomised controlled study. Pain. 2006;126(1–3):245–55.
58. Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, Fitter M, Roman M, Walters S, Nicholl JP. Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. Health Technol Assess. 2005;9(32):iii–+.
59. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture for patients with chronic neck pain. Pain. 2006;125(1–2):98–106.
60. Witt CM, Jena S, Selim D, Brinkhaus B, Reinhold T, Wruck K, Liecker B, Linde K, Wegscheider K, Willich SN. Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. Am J Epidemiol. 2006;164(5):487–96.

61. Haake M, Muller HH, Schade-Brittinger C, Basler HD, Schafer H, Maier C, Endres HG, Trampisch HJ, Molsberger A. German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. Arch Intern Med. 2007;167(17):1892–8.
62. Kennedy S, Baxter GD, Kerr DP, Bradbury I, Park J, McDonough SM. Acupuncture for acute non-specific low back pain: a pilot randomised non-penetrating sham controlled trial. Complement Ther Med. 2008;16(3):139–46.
63. Vas J, Ortega C, Olmo V, Perez-Fernandez F, Hernandez L, Medina I, Seminario JM, Herrera A, Luna F, Perea-Milla E, et al. Single-point acupuncture and physiotherapy for the treatment of painful shoulder: a multicentre randomized controlled trial. Rheumatology (Oxford). 2008;47(6):887–93.
64. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making. 2013;33(5):607–17.
65. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? BMC Med Res Methodol. 2014;14:30.
66. Brooks R, De Charro F. EuroQol: The current state of play. Health Policy. 1996;37(1):53–72.
67. Williams A. Euroqol - a New Facility for the Measurement of Health-Related Quality-of-Life. Health Policy. 1990;16(3):199–208.
68. NICE. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence, Institute's Decision Support Unit; 2013.
69. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. In: Centre for Health Economics. 1999.
70. Gray A, Rivero-Arias O, Clarke P. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Mak. 2006;26(1):18–29.
71. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health and Quality of Life Outcomes. 2009;7(1):1–9.
72. Maund E, Craig D, Suekarran S, Neilson A, Wright K, Brealey S, Dennis L, Goodchild L, Hanchard N, Rangan A, et al. Management of frozen shoulder: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2012;16(11):i-xvi+1-243.
73. Barton G, Sach T, Jenkinson C, Avery A, Doherty M, Muir K. 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.
74. Khan K. Mapping outcomes data to estimate health state utilities: An analysis of individual patient level data from 15 acupuncture trials. Department of Economics and Related Studies. UK: University of York, 2011.
75. 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. Med Decis Making. 2014;34(7):863–72.
76. Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. J Clin Epidemiol. 2007;60(12):1229–33.
77. Aiello F, Attanasio M, Tine F. Assessing covariate imbalance in meta-analysis studies. Stat Med. 2011;30(22):2671–82.
78. Lu GB, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc. 2006;101(474):447–59.
79. Higgins J, Whitehead A, Turner R, Omar R, Thompson S. Meta-analysis of continuous outcome data from individual patients. Stat Med. 2001;20(15):2219–41.
80. Senn S. Change from baseline and analysis of covariance revisited. Stat Med. 2006;25(24):4334–44.
81. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol. 2006;59(9):920–5.
82. Cooper N, Sutton A, Morris D, Ades A, Welton N. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009;28(14):1861–81.
83. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009;172(1):21–47.
84. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. J Roy Stat Soc B. 2002;64:583–616.
85. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE. Evidence synthesis for decision making in healthcare. First Edition, 2012. John Wiley & Sons, Ltd.
86. Team RC. In: Computing RFfS, editor. R: a language and environment for statistical computing. Austria; 2014.
87. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10(4):325–37.
88. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. J Stat Softw. 2005;12(3):1–16.
89. Plummer M, Nicky B, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6(1):7.
90. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7(4):457–72.
91. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7(4):434–55.
92. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. Stat Med. 2013;32(6):914–30.
93. Cherkin D, Eisenberg D, Sherman K, Barlow W, Kaptchuk T, Street J, Deyo R. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. Arch Intern Med. 2001;161(8):1081–8.
94. Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, Fisher P, Van Haselen R. Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial. BMJ. 2004;328(7442):744.
95. Thomas K, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell M, Roman M, Walters S, Nicholl J. Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. Brit Med J. 2006;333(7569):623–6.
96. Mavridis D, White I, Higgins J, Cipriani A, Salanti G. Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. Stat Med. 2014;34(5):721–41.
97. Bujkiewicz S, Thompson J, Sutton A, Cooper N, Harrison M, Symmons D, Abrams K. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. Stat Med. 2013;32(22):3926–43.
98. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DP, Abrams KR. Use of Bayesian multivariate meta-analysis to estimate the HAQ for mapping onto the EQ-5D questionnaire in rheumatoid arthritis. Value Health. 2014;17(1):109–15.
99. Achana FA, Cooper NJ, Bujkiewicz S, Hubbard SJ, Kendrick D, Jones DR, Sutton AJ. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. BMC Med Res Methodol. 2014;14:92.
100. Bujkiewicz S, Thompson JR, Riley RD, Abrams KR. Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. Stat Med. 2016;35(7):1063-89.
101. Jenkins D, Martina R, Bujkiewicz S, Dequen P, Abrams K. Network Meta-Analysis of Biological Response Modifiers in Rheumatoid arthritis Including real World Evidence at Multiple time Points. Value Health. 2015;18(7):A343.
102. Tudor Smith C, Dwan K, Altman D, Clarke M, Riley R, Williamson PR. Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. Plos One. 2014;9(5):e97886.