BMC
Medical Research Methodology

RESEARCH ARTICLE

Open Access

Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data

Pedro Saramago[1*], Ling-Hsiang Chuang[2] and Marta O Soares[1]

Abstract

Background: Network meta-analysis methods extend the standard pair-wise framework to allow simultaneous comparison of multiple interventions in a single statistical model. Despite published work on network meta-analysis mainly focussing on the synthesis of aggregate data, methods have been developed that allow the use of individual patient-level data specifically when outcomes are dichotomous or continuous. This paper focuses on the synthesis of individual patient-level and summary time to event data, motivated by a real data example looking at the effectiveness of high compression treatments on the healing of venous leg ulcers.

Methods: This paper introduces a novel network meta-analysis modelling approach that allows individual patient-level (time to event with censoring) and summary-level data (event count for a given follow-up time) to be synthesised jointly by assuming an underlying, common, distribution of time to healing. Alternative model assumptions were tested within the motivating example. Model fit and adequacy measures were used to compare and select models.

Results: Due to the availability of individual patient-level data in our example we were able to use a Weibull distribution to describe time to healing; otherwise, we would have been limited to specifying a uniparametric distribution. Absolute effectiveness estimates were more sensitive than relative effectiveness estimates to a range of alternative specifications for the model.

Conclusions: The synthesis of time to event data considering individual patient-level data provides modelling flexibility, and can be particularly important when absolute effectiveness estimates, and not just relative effect estimates, are of interest.

Keywords: Evidence synthesis, Time to event data, Survival data, Individual patient data, Aggregate data, Meta-analysis, Mixed treatment comparisons, Network-meta-analysis, Treatment-effect modifiers

Background

Treatment decisions in medicine, whether at the patient- or policy level, should consider all relevant alternative health care technologies potentially capable of delivering the change being sought. Such informed decision making on the use of competing treatments requires evidence of relative effects from randomised controlled trials (RCTs) (which may be further used to inform estimates of cost-effectiveness). However, trying to synthesise evidence from several different pair-wise comparisons relevant to a decision problem can be challenging. A piecemeal approach can be avoided if all RCTs evaluating interventions relevant to the treatment decision are considered collectively, for example, with the use of network meta-analysis (NMA). NMA is a well-established statistical technique that extends standard pairwise meta-analysis framework to allow simultaneous comparison of multiple interventions in a single statistical model [1,2]. This approach then produces relative effect estimates (and associated descriptions of uncertainty) for all treatments connected by the network of evidence [3] – even where head-to-head trials for comparisons do not exist.

NMA using individual patient data

Published work on NMA mainly focuses on the synthesis of aggregate data (AD) (sometimes called summary data, e.g. group means and standard errors available from study reports) [4,5]; however, methods have been developed that allow use of individual patient-level data (IPD) in NMA,

* Correspondence: pedro.saramago@york.ac.uk
[1]Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK
Full list of author information is available at the end of the article

specifically for binary and continuous data [6-8]. The appeal of including IPD in a NMA is that it is likely to reduce statistical heterogeneity across the network (and in this way help resolve possible inconsistencies); it may also allow subgroup effects to be estimated which in turn could guide more personalised treatment decisions [6]. The use of IPD, alone or in combination with AD, has been shown to improve inference in NMAs where the outcome of interest is binary (dichotomous) by aiding convergence, and by providing unbiased treatment–covariate interactions (that would otherwise be affected by ecological bias [9]). For continuous outcomes, IPD is likely to produce more precise estimates of treatment effects, even in the absence of treatment–covariate interactions [10].

NMA using time to event related outcome data

When undertaking an NMA, aggregate time to event data presented as hazard ratios (HR), and some measure of uncertainty around these estimates, can be pooled directly using standard methods (analogous to pooling odds ratios or relative risks) [11]. However, other AD outputs such as median/mean time to event [12] and cumulative counts of patients having the outcome event in a period of time [13] have also been meta-analysed in a network. In modelling these data, an underlying time to event distribution can be specified which will then allow HRs to be generated from the original AD [5,14,15]. To date, evidence synthesis that includes IPD on time to event has been limited to pairwise analysis; additionally there has been little methodological exploration how outcomes in the form of IPD and AD might be considered together in the same NMA [13].

Developing a NMA combining AD and IPD data to synthesise time to event related outcomes

This paper describes a modelling framework that combines AD and IPD in the synthesis of time to event related outcomes within an NMA. It extends the work of Saramago et al. [6] and Sutton et al. [16] on the synthesis of AD and IPD for binary data to consider time to event outcomes. And extends the work on NMAs for time to event AD outcomes by Soares et al. [15] and Woods et al. [14] (the general modelling framework is also described in Dias et al. [5]). Our work was motivated by an NMA on treatments for healing venous leg ulcers, for which we had data from multiple RCTs (more information provide in the next section). A proportion of the RCT data on ulcer healing was available in IPD format (time to healing and time to censoring), with the remaining data available as AD, i.e. count data (proportion of patients healed at end of follow-up which differs across studies). To maximally draw from available data we aimed to jointly synthesise

the available AD and IPD, and, in this way, generating better estimates and providing fuller characterisations of uncertainty to best inform decisions on the use of the treatments of interest.

Motivating example: high compression treatments for venous leg ulcers

The case study relates to compression systems aiming to deliver high compression (classed as ≥40 mmHg compression at the ankle) to promote venous leg ulcer healing. Available standardised systems are: two layer hosiery (HH), the four layer bandage (4LB), the short stretch bandage (SSB), the zinc paste bandage (ZINC), and the two layer bandage system (2LB). A detailed description is provided in Additional file 1 with further details of these systems presented elsewhere [17].

Effectiveness evidence from RCTs was obtained from the most recent update of the relevant Cochrane review available to us, and from a recent multicentre RCT which compared 4LB with HH. All available RCT evidence was assessed for inclusion in the current NMA: a detailed process that has been reported elsewhere [17]. The final NMA contained data from 16 RCTs on the relative effectiveness of high compression systems for the treatment of venous leg ulcers. Data for two of the 16 included RCTs (VenUS I and VenUS IV, hereby denominated studies 1 and 2) had full IPD data available (841 participants) which included time to healing or censoring for each participant, together with other individual-level characteristics such as treatment centre, ulcer duration and size and also patient mobility. For the remaining RCTs (1105 participants), aggregate data on the number of healed ulcers were extracted from the source review alongside information regarding treatment type, number of participants allocated to each treatment group, mean duration of follow-up (if this was not stated, trial duration was used), mean ulcer duration and size.

The 16 included RCTs described nine unique high compression treatments: the five standard treatments (4LB, SSB, ZINC, HH and the 2LB) and four *ad hoc* systems [17]. The *ad hoc* group consisted of treatments deemed irrelevant to current clinical practice, and are not reported further (results can be provided upon request). These studies were, however, included in the NMA as their data was potentially relevant, for example, in describing determinants of healing.

Table 1 describes the data available and Figure 1 presents the treatment network formed by the evidence [17-32]. The most populated comparison was the 4LB vs. SSB, informed by seven RCTs: six with healing data available as AD [18-22,31] and one as IPD [32]. The link between the 2LB and 4LB was informed by two RCTs and each of the remaining six comparisons in the NMA were informed

Table 1 Analytic dataset

ID	Study	Treatment	Follow up (weeks)	Number patients	Mean duration (months)	Mean size (cm²)	Number healed	Evidence format available
16	Duby *et al.* 1993 [18]	4LB	12	25	20.5	11.9	11	AD
		SSB	12	25	26.7	13.1	10	
17	Scriven *et al.* 1998 [19]	4LB	52	32	13	13.3	17.6	AD
		SSB	52	32	21	8.3	18.24	
18	Partsch *et al.* 2001 [20]	4LB	16	53	1.25	1.5	33	AD
		SSB	16	59	1	1.9	43	
19	Ukat *et al.* 2003 [21]	4LB	12	44	–	17.7	13	AD
		SSB	12	45	–	12.2	10	
20	Franks *et al.* 2004 [22]	4LB	24	74	2	5	59	AD
		SSB	24	82	2	3.5	62	
21	Junger *et al.* 2004b [23]	SSB	12	60	5.57	5.95	19	AD
		HH	12	61	4.14	5.62	29	
22	Kralj *et al.* 1996 [24]	4LB	24	20	7.9	18.6	7	AD
		Ad hoc: Ba	24	20	6.9	17.2	8	
23	Polignano *et al.* 2004b [25]	4LB	24	39	–	10.1	29	AD
		ZINC	24	29	–	9.3	19	
24	Wilkinson *et al.* 1997 [26]	4LB	12	17	–	11.2	8	AD
		Ad hoc: BHeH	12	18	–	8.6	8	
25	Colgan *et al.* 1995 [27]	4LB	12	10	9.3	27.5	6	AD
		Ad hoc: BzeaH	12	10	66.5	48.5	7	
26	Blecken *et al.* 2005 [28]	4LB	12	12	–	50.08	4	AD
		Ad hoc: HV	12	12	–	48.98	4	
27	Moffatt *et al.* 2008 [29]	4LB	4	42	48.8	5.7	3	AD
		2LB	4	39	46.6	11.8	6	
28	Szewczyk *et al.* 2010 [30]	4LB	12	15	–	6	9	AD
		2LB	12	16	–	5.3	10	
29	Wong *et al.* 2012 [31]	4LB	24	107	–	–	72	AD
		SSB	24	107	–	–	77	
30	Iglesias *et al.* 2004 [32]	4LB	52	195	3	3.81	107	IPD
		SSB	52	192	3	3.82	86	
14	Ashby *et al.* 2013 [17]	4LB	52	224	12.29	9.30	157	IPD
		HH	52	230	10.82	9.41	163	

AD – aggregate-level data; IPD – individual patient data; 4LB, SSB, HH, Zinc paste,2LB and the *ad hoc* systems Ba, BHeH, BzeaH, HV as described in Additional file 1.
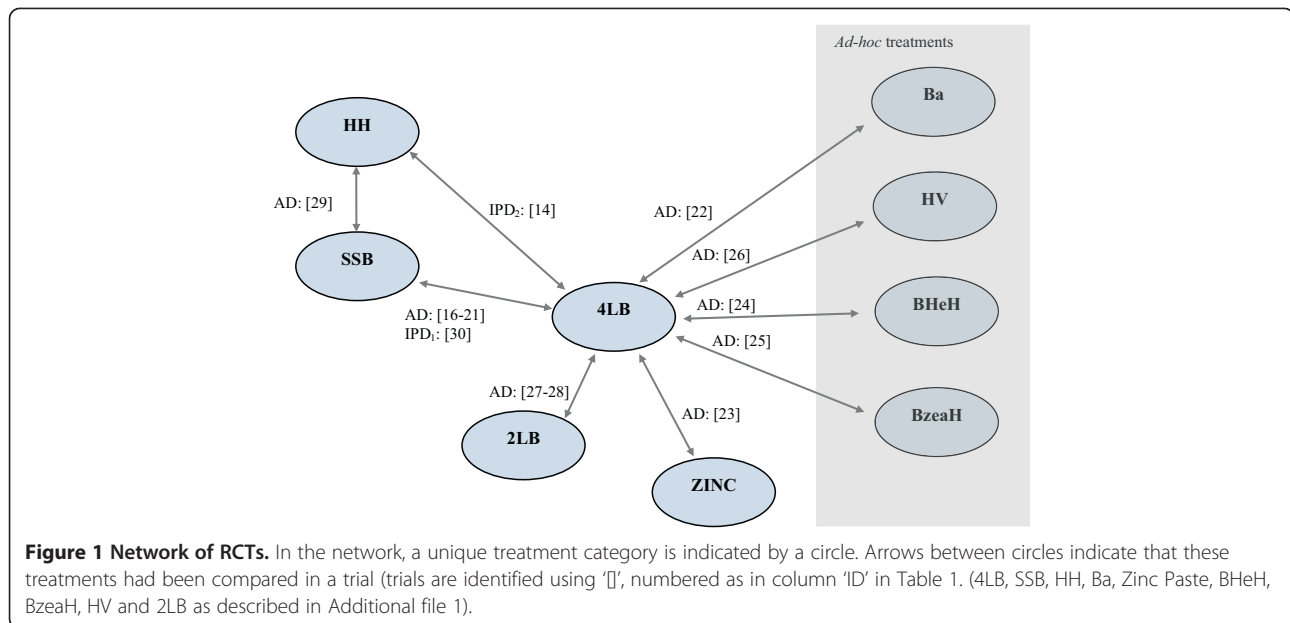
by AD extracted from one RCT for each comparison (Table 1).

Methods

We first describe the modelling framework for our main analysis, *model A*. This model brings together IPD and AD into the same synthesis model, with both types of evidence contributing to the estimation of all key models parameters (i.e. absolute and relative effectiveness estimates). We then detail the evaluation of alternative assumptions applied to *model A*, thus highlighting and

challenging specific assumptions of the modelling framework proposed. All synthesis was conducted in a Bayesian framework.

Suppose there are two sets of studies, one set for which IPD are available and one set for which AD are available. Consider the set of studies to be T = {1,…, number of IPD studies (NS_{IPD}),…, NS_{IPD} + number of AD studies (NS_{AD})} where the total number of studies is *J*; consider the set of treatments to be {1,2,3,…}, where the total number of treatments is *K*, and treatment 1 the reference.

Figure 1 Network of RCTs. In the network, a unique treatment category is indicated by a circle. Arrows between circles indicate that these treatments had been compared in a trial (trials are identified using '[]', numbered as in column 'ID' in Table 1. (4LB, SSB, HH, Ba, Zinc Paste, BHeH, BzeaH, HV and 2LB as described in Additional file 1).

Statistical model for the data

We describe *model A* in two interrelated parts: *part I* describes the modelling of the IPD and *part II* the modelling of the AD.

$$Model\ A, part\ I-\ modelling\ the\ IPD\ studies,$$
$$controlling\ for\ baseline\ covariates$$

$$t_{ijk} \sim Weilbull; (s, \lambda_{ijk}) I\left(t_{ijk}^c\right)$$

$$log\left(\lambda_{ijk}\right) = \begin{cases} \mu_j^{IPD} + \gamma_j^c + \beta_{0j} x_{ijk} & if\ k = b \\ \mu_j^{IPD} + d_{bk} + \gamma_j^c + \beta_0 x_{ijk} & if\ k > b \end{cases}$$

$$x_{ijk} \sim N(m,p) \qquad \gamma_j^c \sim N(0, \pi)$$

(A1)

Time to ulcer healing (t_{ijk}) of the ith participant in the jth study (where $j =1,..., NS_{IPD}$) and in the kth treatment arm was assumed to be Weibull distributed [33] with shape[a.] parameter, s, and scale, λ_{ijk}. For some participants, time to event was not observed, and these observations were censored at the time the participant last had trial data recorded, t_{ijk}^c. The linear predictor $log(\lambda_{ijk})$, was modelled as a function of the log-hazard of an event for the baseline treatment b, μ_j^{IPD}, of the log of the treatment effects, d_{bk} and of a study-specific individual-level regression term, $\beta_{0j} x_{ijk}$, where β_{0j} exemplifies a covariate effect, i.e. the difference in the log hazard ratio per unit increase in the covariate x_{ijk} a patient-level covariate for the ith patient in the jth trial on treatment k available in the IPD data sets [17]). The effect of each covariate on the hazard of healing was assumed to be equal in both IPD studies. Due to the possibility of missing covariate

information existing for some individuals, x_{ijk} was represented as a Normally distributed random variable with mean m and precision p, common across all IPD studies. This is essentially a multiple imputation technique through MCMC[b.] and assumes that the missingness mechanism was at random[c.]. Additionally, to account for centre variability within each IPD study, γ_j^c was defined for each centre, c, in the jth study, these were combined using a common frailty effect described by a normal distribution with mean zero and precision τ.

The treatment effects, d_{bk}, were log-hazard ratios for treatment k relative to the study-specific baseline treatment b, partitioned further as $d_{1k} - d_{1b}$. This equation defines the set of functional parameters, that is the log effects of each treatment in relation to the reference treatment 1. Note that $d_{11} = 0$, where treatment 1 was 4LB, arbitrarily chosen as reference treatment.

Vague prior distributions were specified for $[\mu_b^{IPD} (\sim N(0, 10^6))]$ for the regression coefficient $\beta_{0j} (\sim N(0, 10^6))$ for $m(\sim N(0, 10^6))$ and $p(\sim Gamma(0.01, 0.01))$ for the shared centre effect $\tau(\sim Gamma(0.01, 0.01))$, for $s(\sim Gamma(0.01, 0.01))$ and for $d_{1k} \sim N(0, 10^6)$.

$$Model\ A,\ part\ II - modelling\ the\ AD\ studies$$

$$r_{jk} \sim Bin\left(p_{jk}, n_{jk}\right)$$

$$p_{jk} = 1 - exp\left(-\lambda_{jk}^{AD} \cdot \left(t_{jk}^{AD}\right)^s\right)$$

$$log\left(\lambda_{jk}^{AD}\right) = \begin{cases} \mu_j^{AD} & if\ k = b \\ \mu_j^{AD} + d_{bk} & if\ k > b \end{cases}$$

(A2)

Within the AD studies, the observed number of participants with a healed ulcer within each study, r_{jk}, out of the

total number of individuals in the jth trial [(where $j = NS_{IPD}+1,..., (NS_{IPD}+ NS_{AD})$)] and k^{th} treatment (intention to treat), n_{jk}, was assumed to be Binomially distributed. The underlying probabilities of an event for each arm and in each trial were represented by p_{jk}. In turn, p_{jk} was expressed as a function of the scale parameter, λ_{jk}^{AD}, of follow-up time, t_{jk}^{AD}, and the shape parameter, s, of the Weibull distribution. The linear predictor, $\log\left(\lambda_{jk}^{AD}\right)$ was a function of the baseline log-hazard of an event for treatment b in study j, μ_{jb}^{AD}, and by the log-hazard ratio for treatment k and baseline treatment b, d_{bk} ($=d_{1k} - d_{1b}$) Note that there are parameters common to both model parts (equations A1 and A2), namely the log-hazard ratios and the shape parameter of the time to healing distribution. Vague prior distributions were specified for $\mu_{jb}^{AD}\left(\sim N\left(0, 10^6\right)\right)$.

Alternative modelling assumptions
Underlying *model A* are a set of assumptions that are detailed below. Such assumptions have been relaxed in *models B to D*.

Exploring between-study variation
Model A assumed that each included RCT aimed to measure a common treatment effect (fixed-effect); however, it is likely that there was between-study variation. *Model B* included a random effect to characterise between-study heterogeneity, where d_{bk} was replaced by a study specific δ_{jbk}, that is then described by a Normal distribution with mean d_{bk} and precision, *prec* – this is common to both *parts I* (eq. A1) and *II* (eq. A2).

Time to healing distributions
Model A used the Weibull distribution to describe time to healing. Our choice of survival distribution was limited as distributions such as the Log-Logistic or the Log-Normal do not allow the probability of healing over time to be expressed in a closed form, and hence impede the approach proposed here for the joint synthesis of IPD and AD. Other distributions, such as the Gompertz, were not readily defined within the software used in this work (WinBugs/OpenBugs), specifically under censoring. Nonetheless, the goodness of fit could still be assessed in each IPD data source individually. To do this, we applied parametric regression survival-time models [33] to both IPD data sources [17,32] independently (covariates and frailty effect considered, as in *model A*).

Distributional shape parameter
Model A assumes that the Weibull shape parameter of the hazard of healing was common to both IPD data sources.

It is possible that this parameter differed between studies, in which case HRs could be affected. Thus, we implemented two alternative NMA models to ascertain the impact of this assumption on the relative effectiveness estimates: *model C1* used the shape parameter from the first IPD study to describe the AD studies and *model C2* used the shape parameter from the second IPD study to describe these same studies. Because *models C1* and *C2* represent simple modifications of *model A* we do not present these algebraically.

Treatment-covariate associations
Model A uses baseline covariates to adjust for clinical heterogeneity in the IPD. To further explore the impact of covariates on the relative treatment effects (i.e. whether they were effect modifiers), and potentially help explain between-study heterogeneity, we also included interaction terms between alternative treatments and baseline ulcer area and duration— as described by Cooper *et al.* 2009 [34] and Saramago *et al.* 2012 [6]. *Model D* assumed a regression (slope) coefficient for the interaction terms, this effect was common across treatments and thus common to *parts I* (eq. D1) and *II* (eq. D2). This assumption was data driven, as this was the only option we were able to implement with the data available (i.e. compared to assuming 'exchangeability' or 'independence'). Note that interaction estimates obtained were influenced by the full evidence base for which study mean covariate(s) values were available, including trials considering *ad hoc* treatments.

Model D, part I- modelling the IPD studies, considering common treatment-by-covariate interactions

Model D, *part I– modelling the IPD studies*, *considering common treatment–by–covariate interactions*

$$t_{ijk}\sim Weibull;\ (s, \lambda_{ijk})I\left(t_{ijk}^c\right)$$

$$log(\lambda_{ijk}) = \begin{cases} \mu_b^{IPD} + \gamma_j^c + \beta_{0j}x_{ijk} & if\ k = b \\ \mu_b^{IPD} + \gamma_j^c + \beta_{0j}x_{ijk} + d_{bk} + \beta x_{ijk} & if\ k > b \end{cases}$$

$$x_{ijk}\sim N(m,p) \qquad \gamma_j^c\sim N(0,\tau)$$

(D1)

Time to ulcer healing was modelled in the same way as in *model A* (equation A1). A treatment-by-covariate interaction regression term, βx_{ijk}, was defined, where β is the association effect, assumed common across studies and the same regardless of treatment (excluding control). We included no interaction term for each comparison of k versus b when $b \neq 1$ because the common regression coefficient cancels out; but, as $\beta_{11} = 0$, we included an interaction term for each comparison of k versus treatment *1*. For the remainder of the parameters of

interest, vague prior distributions were assigned as in *model A*.

$$r_{jk} \sim Bin\left(p_{jk}, n_{jk}\right)$$

$$p_{jk} = 1 - exp\left(-\lambda_{jk}^{AD} \cdot \left(t_{jk}^{AD}\right)^s\right)$$

$$log\left(\lambda_{jk}^{AD}\right) = \begin{cases} \mu_{jb}^{AD} & \text{if } k = \text{b} \\ \mu_{jb}^{AD} + d_{bk} + \beta\bar{x}_j & \text{if } k > \text{b} \end{cases}$$

$$(D2)$$

In part two of *model D*, $\bar{x}_j$ represented the mean covariate value for the jth trial ((where $j = NS_{IPD}+1,..., (NS_{IPD}+NS_{AD})$). Both the IPD and AD contributed to the estimation of this interaction term (i.e. same β term in equations D1 and D2). All other components of the model were as described for *model A*.

Model selection and implementation

The NMA analyses were undertaken in the WinBUGs software [35]. In all models the MCMC sampler was run for 10 000 iterations and these were discarded as 'burn-in'. Models were run for a further 5000 iterations, on which inferences were based. Chain convergence was assessed by running the model for two disparate sets of initials values, and by then checking that the Brooks-Gelman-Rubin diagnostic converges to 1.0. The WinBUGS code is included for reference in Additional file 2. Within the NMA, goodness of fit was assessed using the deviance information criterion (DIC) [36]. Results were presented using hazard ratio estimates (and associated credibility intervals, CrIs) and also using the probability of each compression system being the 'best' treatment in terms of being the most clinically effective [37].

The statistical software STATA [38] was used to fit alternative time to event distributions to the IPD datasets individually. Goodness of fit was assessed with the Akaike Information Criteria (AIC) statistic [39].

Results

Table 2 shows parameter estimates obtained for *model A* (first column) and alternative models relaxing its assumptions (*models B* to *D*, second to fifth columns). *Model A* considers the baseline information for the following covariates: the log of the ulcer area (in cm^2) and duration (in months) - both centred to its sample mean; and dummies generated from a categorical variable on participants' mobility with three categories ('immobile', 'walks with difficulty' and 'walks freely', the latter being the reference category). The results for *model A* highlight that the modelling framework proposed is feasible. A graphical

representation of *model A* results is shown in Figure 2. This plots the probability of healing over time for the five main high compression ulcer treatments, but considers only the uncertainty over relative treatment effects. The results of testing the assumptions are described next.

Exploring between-study variation

Despite estimates of HRs from the random effect model (*model B*) being associated with wider CrIs than those from *model A* (as expected), point estimates were found to be fairly similar except for the comparison between HH *vs.* 4LB: HH is estimated to be more effective in *model B* (HR 1.63, 95% CrI 0.76-3.53) compared to *model A* (HR: 1.05, 95% CrI 0.85 to 1.29), although the CrI of the former includes the later. The treatment with the greatest estimated probability of healing was HH in *model B* (59%), rather than 2LB (72%) as in *model A*. Differences may be explained by any existing variation between studies of SSB *vs.* 4LB indirectly impacting on the evidence loop 4LB *vs.* SSB *vs.* HH. Baseline covariate effect estimates remained similar. However, note that the gain in quality fitting of the random-effects model compared to the fixed-effects is null (DIC: 5396.21 and 5396.22, respectively). Previous published work assessing evidence on the SSB *vs.* 4LB comparison [40] similarly found no evidence of between-study heterogeneity.

Time to healing distributions

The Weibull was used to describe time to healing in *model A*. Whilst we were limited in the use of other distributions, goodness of fit was explored by applying alternative time to event distributions to the IPD studies individually. Table 3 shows results of such analysis (AIC statistic) and Additional file 3 graphically presents the fitted curves and the Kaplan Meier counterpart for control arms. The best fitting distributions for both studies were the Log-Logistic and Log-Normal. Of the remaining, the Weibull and Gompertz distributions provided better fit than the Exponential; this was expected given the flexibility of these distributions in assuming increasing, decreasing or constant hazards over time. The Weibull was best in IPD study 2 and the Gompertz best in IPD study 1.

Distributional Weibull shape parameter of healing hazard

The Weibull shape parameters estimated within *models C1* and *C2* indicate that in IPD study 1 [32] the hazard of healing was expected to decrease over time ($s_1 = 0.93$, 95% CrI 0.86-1.01), while in IPD study 2 it is expected to increase ($s_2 = 1.27$, 95% CrI 1.17-1.38). Note that there is no overlap in the CrIs. However, results show that relative effectiveness estimates are robust to the range of assumptions tested: the estimated HRs did not differ between *models C1* and *C2*, and did not substantially differ from *model A*.

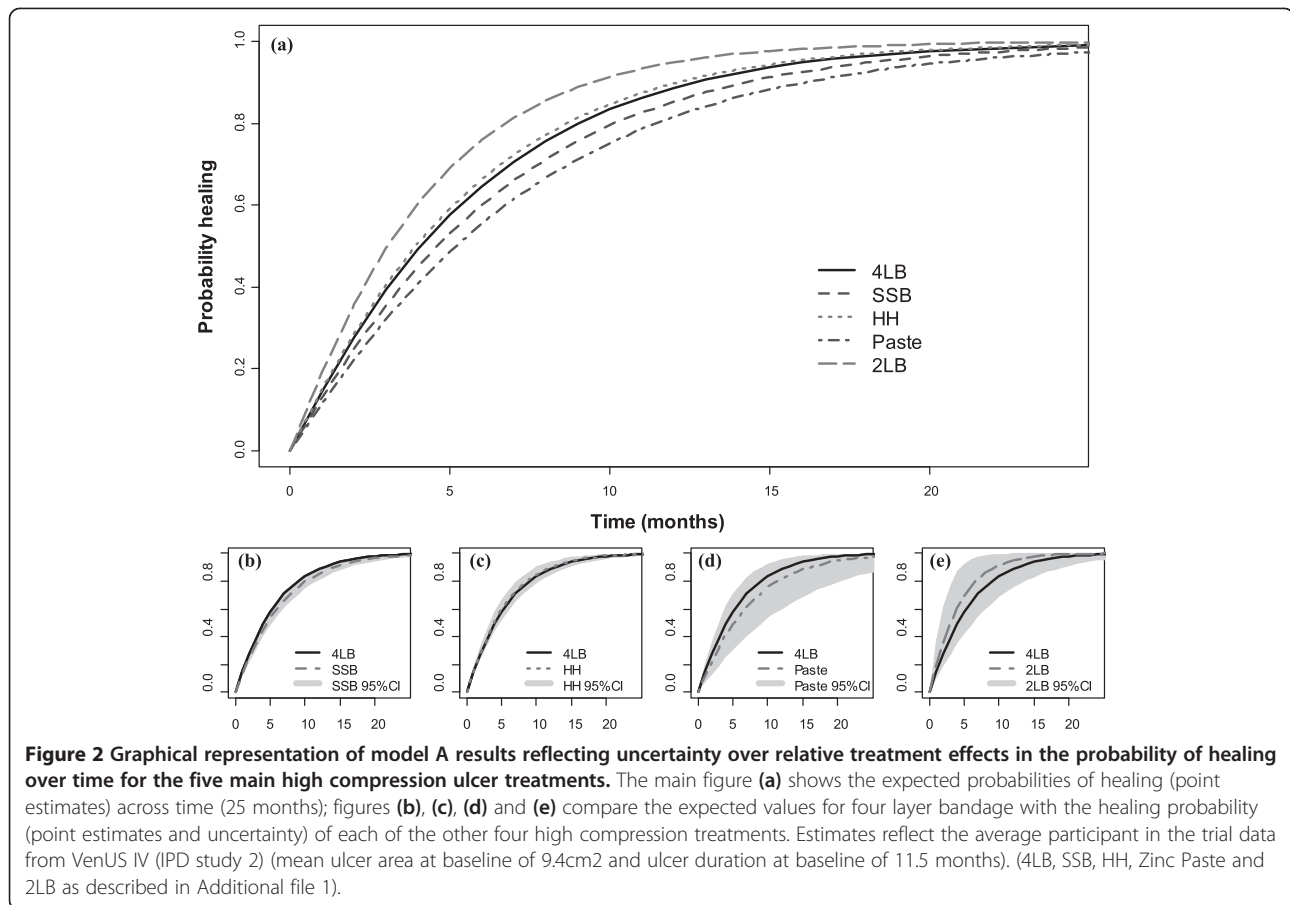**Table 2 Parameter estimates from the alternative MTC synthesis models**

Hazard ratios		Model A			Model B			Model C1			Model C2			Model D		
		HR	Median (95% CrI)	P	HR	Median (95% CrI)	P	HR	Median (95% CrI)	P	HR	Median (95% CrI)	P	HR	Median (95% CrI)	P
Treatment effects	4LB	—	—	5.5	—	—	1.4	—	—	6.2	—	—	5.7	—	—	4.3
	SSB	0.88	(0.76, 1.03)	0.4	0.96	(0.77, 1.22)	0.6	0.89	(0.77, 1.04)	0.6	0.89	(0.77, 1.04)	0.6	0.84	(0.70, 0.99)	0.2
	HH	1.05	(0.85, 1.29)	16.1	1.63	(0.76, 3.53)	59.2	1.03	(0.83, 1.27)	14.9	1.03	(0.84, 1.27)	15.0	1.03	(0.84, 1.28)	11.1
	ZINC	0.77	(0.41, 1.42)	6.2	0.78	(0.37, 1.62)	2.8	0.78	(0.41, 1.44)	6.5	0.78	(0.41, 1.43)	6.7	0.75	(0.03, 29.49)	17.5
	2LB	1.40	(0.65, 3.05)	71.9	1.39	(0.62, 3.30)	36.0	1.38	(0.66, 3.05)	71.8	1.38	(0.63, 3.04)	72.0	1.59	(0.61, 5.34)	67.0
Baseline characteristics	Log area	0.71	(0.66, 0.76)	—	0.71	(0.66, 0.76)	—	0.70	(0.65, 0.75)	—	0.70	(0.65, 0.75)	—	0.71	(0.65, 0.76)	—
	Log duration	0.92	(0.90, 0.94)	—	0.92	(0.90, 0.94)	—	0.92	(0.91, 0.94)	—	0.93	(0.91, 0.94)	—	0.92	(0.90, 0.94)	—
	Difficulty in walking	0.71	(0.60, 0.85)	—	0.73	(0.60, 0.86)	—	0.72	(0.60, 0.85)	—	0.72	(0.60, 0.85)	—	0.71	(0.60, 0.80)	—
	Immobile	0.67	(0.23, 1.52)	—	0.66	(0.23, 1.51)	—	0.72	(0.24, 1.65)	—	0.72	(0.25, 1.67)	—	0.68	(0.24, 1.59)	—
Interactions	Log area	—	—	—	—	—	—	—	—	—	—	—	—	1.00	(0.97, 1.10)	—
	Log duration	—	—	—	—	—	—	—	—	—	—	—	—	1.00	(0.99, 1.00)	—
	Btw-centre SD	0.04	(0.01, 0.13)	—	0.05	(0.01, 0.13)	—	0.05	(0.01, 0.15)	—	0.05	(0.01, 0.15)	—			—
	Btw-study SD	—	—	—	0.13	(0.01, 0.51)	—	—	—	—	—	—	—	—	—	—
	λ	1.07	(1.01, 1.13)	—	1.07	(1.01, 1.14)	λ_1**	0.93	(0.86, 1.01)	λ_1	0.93	(0.86, 1.01)	—	1.07	(1.01, 1.14)	—
							λ_2	1.27	(1.17, 1.38)	λ_2**	1.27	(1.17, 1.38)				
	DIC		5396.2			5396.2			5371.2			5371.5			5377.4	

Model A – Fixed-effects NMA, Weibull model of IPD+AD; model B – Random-effects NMA, Weibull model of IPD+AD; model C1 - Fixed-effects NMA, Weibull model of IPD+AD, shape parameter derived from IPD study 1 only; model C2 - Fixed-effects NMA, Weibull model of IPD+AD, shape parameter derived from IPD study 2 only; model D - Fixed-effects NMA, Weibull model of IPD+AD, considering 2 treatment-effect modifiers.
**Shape parameter used in the synthesis model section for summary data.
4LB = four layer bandage; SSB = Short stretch bandage; HH = two layer hosiery; 2LB = two layer bandage.
HR = hazard ratios; CrI = credibility interval; SD = standard deviation; P = probability of being the best treatment choice in terms of healing (%); DIC – deviance information criteria.

Figure 2 Graphical representation of model A results reflecting uncertainty over relative treatment effects in the probability of healing over time for the five main high compression ulcer treatments. The main figure **(a)** shows the expected probabilities of healing (point estimates) across time (25 months); figures **(b)**, **(c)**, **(d)** and **(e)** compare the expected values for four layer bandage with the healing probability (point estimates and uncertainty) of each of the other four high compression treatments. Estimates reflect the average participant in the trial data from VenUS IV (IPD study 2) (mean ulcer area at baseline of 9.4cm2 and ulcer duration at baseline of 11.5 months). (4LB, SSB, HH, Zinc Paste and 2LB as described in Additional file 1).

Treatment-covariate associations

Model D tested the inclusion of interaction terms. Two covariates on ulceration area (in cm^2) and duration (in months) were considered and their association with treatment evaluated. Results (column 5 of Table 2) show that the covariates included did not appear to be treatment effect modifiers in this case study. However, estimating these two additional regression terms increased uncertainty in relative treatment effects estimates, specifically for ZINC and 2LB.

Table 3 Goodness of fit (AIC statistics) of alternative time to ulcer healing models for IPD studies 1 and 2

Time to event model	Akaike Information Criteria (AIC)	
	IPD study 1 (30)	IPD study 2 (14)
Weibull PH	1102.1	1021.0
Gompertz PH	1072.5	1065.5
Exponential PH	1102.7	1068.4
Log-Logistic AFT	1026.1	971.8
Log-Normal AFT	1032.2	961.5

PH = proportional hazards model; AFT = accelerated failure time model

Discussion

This paper introduces a novel NMA modelling approach that allows IPD on time to event (with censoring) and AD on event count (for a given follow-up time) to be synthesised jointly, by assuming an underlying, common, distribution of time to healing. Available IPD is used directly to inform this distribution (likelihood). Studies reporting the number of participants healed (AD) are used to inform a probability parameter, and a Binomial likelihood was defined for this subset of the evidence-set. The probability of healing was then related (algebraically) to the common distribution of time to healing, by taking the duration of follow-up in each AD study into account. This modelling framework extends approaches in the literature [5,14,15] and is also a natural extension of previously published methodologies of synthesising IPD and AD jointly [6-8,16]. This work was motivated by a real data example looking at the effectiveness of high compression treatments on the healing of venous leg ulcers.

We found that the key strength of the use of IPD in this context (additional to the known advantages described in the introduction) was the flexibility in modelling these data allowed. For example, had all evidence been available

as AD, the modelling process would have been limited to the specification of uniparametric distributions for time to healing i.e. the Exponential, with constant healing hazard over time [5,14,15]. In our motivating example, the Exponential distribution was shown to be less adequate than other distributions in describing the time to event data in the studies for which IPD were available. The availability of IPD allowed a more complex distribution for time to event outcomes to be implemented, in this case the Weibull. This may be of particular importance when absolute effectiveness estimates, and not just relative effect estimates, are of interest – and especially where results may need to be extrapolated beyond the follow-up time horizon.

We note that even with the flexibility offered by the use of IPD we were, in practice, limited to using the Weibull distribution. Semi-parametric methods are also unmanageable as the evidence in aggregate form consists of a single data point and thus does not allow defining a non-parametric distribution of hazard of healing. Given this limitation, the synthesis of time to event data will still often require the use of potentially suboptimal distributional assumptions, in which case estimates obtained may be biased. We suggest further research could focuses on using numerical analysis techniques within NMA, to try and resolve this issue.

This work was also relevant in again highlighting the importance of including IPD when wanting to consider patient characteristics in the model. In the presence of heterogeneity, incorporating information on either *a)* patients' baseline characteristics or *b)* control for treatment-effect modifiers within the synthesis model has been shown to improve estimates and, by doing so, possibly resolve evidence inconsistencies [6,34]. The first relates to potential heterogeneity in the baseline hazard, which cannot be explicitly explored with AD alone (this is important when analyses aim to explore determinants of baseline hazard, for example). In this study we assumed a common effect of baseline covariates on the hazard of healing across IPD studies (as commonly assumed in IPD meta-analysis). The second relates to treatment-covariate interactions, which are generally acknowledged to be best estimated using IPD, as ecological bias can be avoided [9]. For the proportion of evidence only available as AD, the model here implemented considered study level mean covariate values. Nonetheless, not all studies provided information for these, and imputation was undertaken (imputation is naturally done through the MCMC, and assumes values are 'missing at random').

In our work we assumed the shape parameter of the Weibull time to event distribution to be common across studies. However, testing proved this assumption was not valid, thus highlighting the importance of evaluating any assumptions of similarity imposed across studies.

Despite relative effectiveness estimates being mainly unaffected, such potential heterogeneity between studies should be explored and accounted for in analyses. Such assumptions of commonality also mean that information may be shared throughout the network, in which case evidence on treatments other than those on our decision set (the five treatments of interest for which results were reported). This is the case of *model D* that makes use of all evidence (including *ad hoc* treatments) to estimate treatment-covariate interaction, which may indirectly affect the relative effectiveness estimates of interest.

In the network of evidence there was one closed loop where both direct and indirect data informed relative treatment effect estimates. The existence of inconsistency was explored elsewhere [17], showing no evidence of statistically significant discrepancies between the direct and the indirect data. We note, however, that given the fairly high uncertainty in the evidence base, only large differences in direct and indirect data within the loop would have returned a statistically significant result.

The evidence included as AD in our case study was limited to the primary outcome data on proportion of patients healed at the end of follow-up. However, we could have extended the modelling framework to use any other summary data reported, such as number of patients healed at different time points or Kaplan-Meier (KM) curves. Information conveyed in Kaplan Meier plots can be reconstructed [41], and further research could focus on including this information in our NMA modelling framework to strengthen inferences.

Conclusions

The use of IPD for a time to event outcome is particularly useful in guiding HTA decision making by allowing flexibility in the specification of more appropriate survival distributions and in dealing with potential existing study heterogeneity [42]. Increasingly, data sharing is promoted in research [43,44] and this example highlights how use of IPD allows the development of more informative and flexible models that are better able to summarise existing evidence. However, it is important to acknowledge that accessing and analysing IPD can be time consuming and may cause delay [4]. The process needs to be well planned and implemented. In our case both sources of IPD were easily accessed and that directly facilitated the conduct of these analyses and the associated methodological work presented here.

Endnotes

[a.]The shape parameter of the Weibull distribution, s, can be interpreted directly as follows: i) if $0 < s < 1$, hazard rate decreases over time; ii) if $s = 1$, hazard rate is constant over time (hazard exponentially distributed); and if $s > 1$, it indicates that the hazard increases with time.

[b.]When imputing missing information, MCMC generates independent draws of the missing data from its predictive distribution. Multiple imputation through MCMC techniques is attractive for exploratory or multi-purpose analyses involving a large number of estimands.

[c.]The missing-at-random assumption (sometimes called the ignorability assumption) considers that the probability that an observation is missing may depend on the observed values but not the missing values, as sufficient data has already been collected.

Additional files

Additional file 1: Description of main compression systems evaluated.

Additional file 2: WinBUGS code.

Additional file 3: Observed and fitted survival distributions for IPD studies 1 and 2.

Abbreviations
2LB: Two layer bandage; 4LB: Four layer bandage; AIC: Akaike information criteria; AD: Aggregate data; DIC: Deviance information criteria; HH: Two layer hosiery; HTA: Health technology assessment; IPD: Individual patient-level data; MCMC: Markov chain Monte Carlo; NMA: Network meta-analysis; RCT: Randomised controlled trial; SSB: Short stretch bandage; ZINC: Zinc paste bandage.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
PS developed and implemented the novel synthesis models, conducted the analysis and prepared the final manuscript. L-HC structured the evidence base and conducted preliminary data analysis using existing published synthesis methodology. MS oversaw and advised on all elements of the work undertaken and prepared the final manuscript. All authors read and approved the final manuscript.

Author details
[1]Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK. [2]Pharmerit International, Rotterdam, The Netherlands.

References
1. Bucher H, Guyatt G, Griffith L, Walter S: The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997, **50**(6):683–691.
2. Sutton A, Ades A, Cooper N, Abrams K: Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008, **26**(9):753–767.
3. Dias S, Welton NJ, Sutton AJ, Ades AE: Evidence synthesis for decision making 1: introduction. *Med Decis Making* 2013, **33**(5):597–606.
4. Saramago P, Manca A, Sutton AJ: Deriving input parameters for cost-effectiveness modelling: taxonomy of data types and approaches to their statistical synthesis. *Value Health* 2012, **15**(5):639–649.
5. 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):607–617.
6. Saramago P, Sutton AJ, Cooper NJ, Manca A: Mixed treatment comparisons using aggregate- and individual-participant level data. *Stat Med* 2012, **10**; 31(28):3516–3536.
7. Jansen J, Cope S: Network meta-analysis of individual and aggregate level data. *Value Health* 2012, **15**(4):A159.
8. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith C: 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):914–930.
9. Berlin J, Santanna J, Schmid C, Szczech L, Feldman H: Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002, **21**(3):371–387.
10. Riley R, Lambert P, Staessen J, 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):1870–1893.
11. Sutton AJ: *Methods for Meta-Analysis in Medical Research.* Chichester ; New York: J. Wiley; 2000.
12. Welton NJ, Willis SR, Ades AE: Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Res Synth Method* 2010, **1**:239–257. doi: 10.1002/jrsm.21.
13. Smith C, Williamson P, Marson A: An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *J Eval Clin Pract* 2005, **11**(5):468–478.
14. Woods B, Hawkins N, Scott D: Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. *BMC Med Res Methodol* 2010, **10**:54.
15. Soares MO, Dumville JC, Ades AE, Welton NJ: Treatment comparisons for decision making: facing the problems of sparse and few data. *J R Stat Soc A Stat Soc* 2014, **177**:259–279. doi: 10.1111/rssa.12010.
16. Sutton A, Kendrick D, Coupland C: Meta-analysis of individual- and aggregate-level data. *Stat Med* 2008, **27**(5):651–669.
17. Ashby RL, Gabe R, Ali S, Saramago P, Chuang LH, Adderley U, Bland JM, Cullum NA, Dumville JC, Iglesias CP, Soares MO, Stubbs NC, Torgerson DJ: VenUS IV (venous leg ulcer study IV): a randomised controlled trial of compression hosiery versus compression bandaging in the treatment of venous leg ulcers. *Health Technology Assessment (forthcoming)* 2014.
18. Duby T, Hoffman D, Cameron J, Doblhoffbrown D, Cherry G, Ryan T: A randomized trial in the treatment of venous Leg ulcers comparing short stretch bandages, 4 layer bandage system, and a long stretch paste bandage system. *Wounds-a Compendium Clin Res Pract* 1993, **5**(6):276–279.
19. Scriven J, Taylor L, Wood Rgn A, Bell P, Naylor A, London N: A prospective randomised trial of four-layer versus short stretch compression bandages for the treatment of venous leg ulcers. *Ann R Coll Surg Engl* 1998, **80**(3):215–220.
20. Partsch H, Damstra R, Tazelaar D, Schuller-Petrovic S, Velders A, De Rooij M, Tjon Lim Sang R, Quinlan D: Multicentre, randomised controlled trial of four-layer bandaging versus short-stretch bandaging in the treatment of venous leg ulcers. *Vasa* 2001, **30**(2):108–113.
21. Ukat A, Konig M, Vanscheidt W, Munter K: Short-stretch versus multilayer compression for venous leg ulcers: a comparison of healing rates. *J Wound Care* 2003, **12**(4):139–143.
22. Franks P, Moody M, Moffatt C, Martin R, Blewett R, Seymour E, Hildreth A, Hourican C, Collins J, Heron A: Randomized trial of cohesive short-stretch versus four-layer bandaging in the management of venous ulceration. *Wound Repair Regen* 2004, **12**(2):157–162.
23. Junger M, Wollina U, Kohnen R, Rabe E: Efficacy and tolerability of an ulcer compression stocking for therapy of chronic venous ulcer compared with a below-knee compression bandage: results from a prospective, randomized, multicentre trial. *Curr Med Res Opin* 2004, **20**(10):1613–1623.
24. Kralj B, Kosicek M: Randomised Comparative Trial of Single-Layer and Multi-Layer Bandages in the Treatment of Venous leg Ulcer. Harrogate, UK: 6th European Conference on Advances in Wound Management: 1995; 1995:158–160.
25. Polignano R, Bonadeo P, Gasbarro S, Allegra C: A randomised controlled study of four-layer compression versus Unna's boot for venous ulcers. *J Wound Care* 2004, **13**(1):21–24.
26. Wilkinson E, Buttfield S, Cooper S, Young E: Trial of two bandaging systems for chronic venous leg ulcers. *J Wound Care* 1997, **6**(7):339–340.
27. Colgan MP, Teevan M, McBride C, O'Sullivan L, Moore D, Shanik G: Cost Comparisons in the Management of Venous Ulceration. In Harrogate, UK: 5th European Conference on Advances in Wound Management: 21–24 November 1995; 1995.
28. Blecken S, Villavicencio J, Kao T: Comparison of elastic versus nonelastic compression in bilateral venous ulcers: a randomized trial. *J Vasc Surg* 2005, **42**(6):1150–1155.

29. Moffatt C, Edwards L, Collier M, Treadwell T, Miller M, Shafer L, Sibbald G, Brassard A, McIntosh A, Reyzelman A, Price P, Kraus SM, Walters SA, Harding K: **A randomised controlled 8-week crossover clinical evaluation of the 3M coban 2 layer compression system versus profore to evaluate the product performance in patients with venous leg ulcers.** *Int Wound J* 2008, **5**(2):267–279.

30. Szewczyk M, Jawie A, Cierzniakowska K, Cwajda-Bialasik J, Moscicka P: **Comparison of the effectiveness of compression stockings and layer compression systems in venous ulceration treatment.** *Archives of Medical Science* 2010, **6**(5):793–799.

31. Wong I, Andriessen A, Charles H, Thompson D, Lee D, So W, Abel M: **Randomized controlled trial comparing treatment outcome of two compression bandaging systems and standard care without compression in patients with venous leg ulcers.** *J Eur Acad Dermatol Venereol* 2012, **26**(1):102–110.

32. Iglesias C, Nelson E, Cullum N, Torgerson D: **VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.** *Health Technol Assess* 2004, **8**(29):iii, 1–iii, 105.

33. Collett D: *Modelling Survival Data in Medical Research.* 2nd edition. Boca Raton, Fla: Chapman & Hall/CRC; 2003.

34. 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):1861–1881.

35. Lunn DJ, Thomas A, Best N, Spiegelhalter D: **WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility.** *Statistics and Computing* 2000, **10**(4):325–337.

36. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A: **Bayesian measures of model complexity and fit.** *J R Stat Soc B Stat Soc* 2002, **64**:583–616.

37. 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.

38. StataCorp A: **Stata Statistical Software: Release 12.** In College Station, TX: StataCorp LP; 2011.

39. Akaike H: *Information Theory and the Maximum Likelihood Principle.* 2nd International Symposium in Information Theory: 1973; 1973:267–281.

40. O'Meara S, Tierney J, Cullum N, Bland J, Franks P, Mole T, Scriven M: **Four layer bandage compared with short stretch bandage for venous leg ulcers: systematic review and meta-analysis of randomised controlled trials with data from individual patients.** *BMJ (Clinical research ed)* 2009, **338**(7702):1054–1057.

41. Guyot P, Ades AE, Ouwens MJ, Welton NJ: **Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves.** *BMC Med Res Methodol* 2012, **12**:9.

42. Espinoza M: *Heterogeneity in cost-effectiveness analysis: methods to explore the value of subgroups and individualised care in a collectively funded health system.* University of York; 2012.

43. **Medical Research Council Data Sharing.** http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/ (accessed in September 2014).

44. **National Institute for Health Research - Health Technology Assessment Data Sharing.** http://www.journalslibrary.nihr.ac.uk/information-for-authors/data-sharing (accessed in September 2014).