Bayesian meta-analysis on medical devices

Youn, Ji-Hee; Lord, Joanne; Hemming, Karla; Girling, Alan; Buxton, Martin

Link to publication on Research at Birmingham portal

BAYESIAN META-ANALYSIS ON MEDICAL DEVICES: APPLICATION TO IMPLANTABLE CARDIOVERTER DEFIBRILLATORS

Ji-Hee Youn
Brunel University
email: cmp11jy@sheffield.ac.uk

Joanne Lord
Brunel University

Karla Hemming and Alan Girling
University of Birmingham

Martin Buxton
Brunel University

Objectives: The aim of this study is to describe and illustrate a method to obtain early estimates of the effectiveness of a new version of a medical device.
Methods: In the absence of empirical data, expert opinion may be elicited on the expected difference between the conventional and modified devices. Bayesian Mixed Treatment Comparison (MTC) meta-analysis can then be used to combine this expert opinion with existing trial data on earlier versions of the device. We illustrate this approach for a new four-pole implantable cardioverter defibrillator (ICD) compared with conventional ICDs, Class III anti-arrhythmic drugs, and conventional drug therapy for the prevention of sudden cardiac death in high risk patients. Existing RCTs were identified from a published systematic review, and we elicited opinion on the difference between four-pole and conventional ICDs from experts recruited at a cardiology conference.
Results: Twelve randomized controlled trials were identified. Seven experts provided valid probability distributions for the new ICDs compared with current devices. The MTC model resulted in estimated relative risks of mortality of 0.74 (0.60–0.89) (predictive relative risk [RR] = 0.77 [0.41–1.26]) and 0.83 (0.70–0.97) (predictive RR = 0.84 [0.55–1.22]) with the new ICD therapy compared to Class III anti-arrhythmic drug therapy and conventional drug therapy, respectively. These results showed negligible differences from the preliminary results for the existing ICDs.
Conclusions: The proposed method incorporating expert opinion to adjust for a modification made to an existing device may play a useful role in assisting decision makers to make early informed judgments on the effectiveness of frequently modified healthcare technologies.

Keywords: Bayesian analysis, Meta-analysis, Medical devices, Expert opinions, Defibrillator, Implantable

Medical devices differ from pharmaceuticals in terms of the nature and quantity of evidence about their effectiveness and cost-effectiveness. The medical device industry is a dynamic sector where new technologies are continuously developed, and the life cycle of a particular version of a device is often less than 2 years. Frequent technological improvements are made without changing major features or functionality. With such rapid incremental change, it can be difficult to keep the evidence up-to-date.

The regulatory environment for devices also differs from that for pharmaceuticals. Evidence required for marketing approval may only describe technical performance rather than clinical safety or efficacy, let alone effectiveness or cost-effectiveness. In the United States, new products can obtain access to the market without being tested on patients if they are considered "substantially equivalent" to previously approved devices. Among numerous examples is the artificial hip known as the articular surface replacement from DePuy, which has now been recalled due to the unusually high rate of implant failure.

The U.S. Food and Drug Administration has recently approved the use of an implantable cardioverter defibrillator (ICD) with four-pole (IS-4) lead connectors without premarket clinical testing, and instead required a post-approval study. Although the new IS-4 device is said to confer benefits such as a smaller, lower-profile pulse generator, there have been some concerns that such a modification should be clinically tested before approval (10).

This leaves potential early adopters—clinicians, patients, and payers—with a difficult decision. Unless a new version of a medical device is a genuine breakthrough, its effects will be largely predictable from information on previous generations. But this does require a judgment on whether demonstrated technical improvements are likely to translate into clinical gains, and if so, whether these are sufficient to justify the (usually higher) cost. Furthermore, potential users have to judge the likelihood of unforeseen adverse effects.

The aim of this study is to develop a method to help decision makers make an informed, early assessment of a new version of a medical device. We suggest that Bayesian bias-adjustment techniques may be adapted to combine existing information on earlier generations of the device with expert judgments about the extent to which the new version is likely to differ. In the

following section, we explain the method in more detail. We then illustrate the method using the example of four-pole ICDs. The advantages and disadvantages of this approach are then discussed.

METHODS FOR EARLY EVALUATION OF A MODIFIED TECHNOLOGY

The Bayesian approach resembles the human cognitive procedure of acquiring knowledge by updating a prior understanding as new information becomes available. Usually, the prior view is supported by scant or poor-quality data, or it may even be based on subjective opinion alone. As more or better empirical evidence emerges, priors are updated to derive posterior estimates. In such cases, there is usually less uncertainty around the posterior, than around the prior state of knowledge. We propose to adapt this method for the situation where there is a strong body of evidence on an existing technology, and only weak evidence or beliefs about how a new modified version of the technology differs. Here, posterior estimates of the effectiveness of the modified technology will be less certain than the prior knowledge about earlier generations of the device.

Bayesian methods of meta-analysis pool evidence from clinical trials to estimate the effectiveness of health technologies. The meta-analysis requires assumptions about the heterogeneity of treatment effects: a fixed effect model assumes a common true effect underlying all study results, whereas a random effects model assumes that trials draw a random sample from the distribution of true effects. The standard pair-wise meta-analysis can be extended to a Mixed Treatment Comparison (MTC) meta-analysis when exploring the relative efficacy of a range of interventions (2). This integrates information from a network of direct and indirect comparative trials within a single model (12). This does, of course, require assumptions about the consistency of the underlying evidence. Bayesian computation is possible with Monte Carlo Markov chain (MCMC) methods using WinBUGS, a freely available statistical software package (16).

These Bayesian meta-analysis techniques may be extended to adjust for suspected "biases" in the evidence base. For example, Welton and colleagues (25) included additive bias parameters to adjust the results of studies that lacked appropriate allocation concealment, estimating the extent of this bias from a large meta-epidemiological study. And Turner et al. (23) incorporated a wide range of possible biases, from selection bias to health outcome bias, using expert judgment to identify and quantify them.

We borrow this idea, introducing an adjustment parameter to reflect beliefs about the difference in effectiveness between a new and an existing version of a device. This can be added to an estimate of treatment effect from a random-effects Bayesian MTC meta-analysis of trial evidence for the existing device, to obtain an estimate of effectiveness for the new device (see Appendix 1 for a more formal mathematical description of this idea). This relies on the assumption that the new device is sufficiently similar to the existing one, and that its effect is predictable from data on the earlier version. And, of course, the same assumptions about the consistency between direct and indirect evidence for the existing device must still hold, as for any MTC analysis.

The probability distribution for the difference parameter, denoted as β, can be estimated from expert opinion. There are many ways in which such distributions can be elicited to reflect individuals' degrees of belief and extent of uncertainty over the location of an unknown parameter. In our example below, we elicited this information in the form of histograms, with experts asked to state the probability of the parameter lying within each of a series of finite intervals.

There are also many ways in which individuals' subjective probability estimates can be pooled. In our example, we combined histograms from several experts by dividing the intervals on which opinion was elicited to obtain the smallest subintervals, and taking the arithmetic means of the probabilities for each of the subintervals across the experts. To illustrate, suppose that Expert 1 said there was a 50 percent probability that the parameter lay in each of two intervals: (0,1) and (1,2). Expert 2 said there was a 50 percent probability that it lay in each of the intervals (0,1) and (1,3). Then, one might say that the elicited probabilities for three subintervals, (0,1), (1,2) and (2,3), are 50 percent, 37.5 percent, and 12.5 percent, respectively.

Appropriate scale conversion (Appendix 1) provides a pooled discrete probability distribution for β, which can be used to directly sample values for this parameter in the Bayesian MCMC model. Alternatively, one might fit a suitable parametric distribution to reflect the pooled opinion. In our example reported below, we found that the elicited distribution was approximately normal, so we calculated the mean of β from the discrete distribution, and used a variation on a univariate least squares fitting method (19) to find its standard deviation. Samples from the fitted Normal distribution for β were then incorporated into the main MTC model.

In other circumstances, other parametric distributions might provide a better fit for the elicited data: for a more complete discussion of distribution selection, see O'Hagan et al. (19).

ILLUSTRATIVE EXAMPLE: ICDS

The Review Question

Our example concerns the use of ICDs in the prevention of sudden cardiac death in high risk patients with a low left ventricular ejection fraction (LVEF < 35–40 percent). These patients often have a history of myocardial infarction and/or congestive heart failure. We included interventions targeted at individuals who have not yet experienced a major arrhythmic event (primary prevention), and those targeted at people who have survived such an event (secondary prevention) (9).
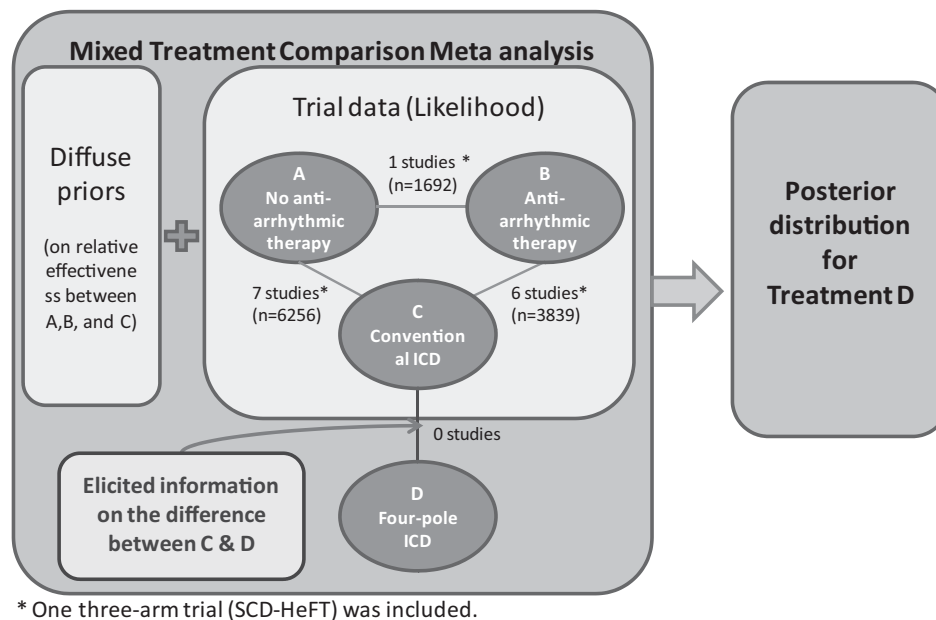
Figure 1. Illustration of analytical approach.

We distinguish four broad types of Treatment: (A) Conventional drug therapy (which may include beta-blockers or ACE inhibitors); (B) Conventional drug therapy and Class III Anti-arrhythmic drug therapy (such as amiodarone or sotalol); (C) Conventional drug therapy and conventional ICD therapy; and (D) Conventional drug therapy and four-pole connector ICD therapy.

"Conventional" drug therapy was defined to include any drugs other than class III anti-arrhythmics, provided they are also prescribed for patients in the other arms of the trial. Interventions might also include other surgical procedures such as coronary artery bypass graft, provided these were also common "underlying" treatments across all arms within the same RCT.

All four interventions can be used in primary or secondary prevention, although conventional drug therapy alone is relatively unusual in secondary prevention.

Review of Evidence on Current ICDs

Clinical trial data on existing ICDs was obtained from a recently-published systematic review by Ezekowitz and colleagues (9). They report details about the search strategy, the characteristics of the included studies and the extent of heterogeneity. Although this review included both randomized and observational studies, we focused only on RCT data for illustrative purposes.

Two authors (J.Y. and K.H.) extracted data from the original papers cited in the Ezekowitz review, and also checked other papers related to each trial if information given in the cited ones was insufficient.

Meta-analysis Model

We used a Bayesian MTC meta-analysis to estimate the relative risk of death for patients under the existing ICD therapy (treatment C) compared with non-ICD therapies (treatments A and B). A hierarchical random effects model was used, assuming study-level treatment effects to be a random sample drawn from a probability distribution for the overall treatment effect governed by a set of common parameters. The model assumes a lognormal distribution for the relative risks, and applies diffuse priors (this is a common MTC model) (12). Conventional drug therapy (treatment A) was taken as the reference treatment for the meta-analysis.

The distribution for the relative effectiveness of the new ICD (treatment D), for which there is not yet any RCT evidence, is assumed to be the same as that for C, except its mean is shifted by an adjustment parameter β, which is also assumed to be normally distributed. The mean and variance of β was estimated from elicited expert opinion.

The meta-analysis model and adjustment for the ICD example is illustrated in Figure 1, and the BUGS code is provided in Appendix 2.

Elicitation Method

Preliminary discussions with two practicing cardiologists led us to focus on the practicality of our elicitation methods, so we adopted a rather simpler approach than that advocated in the literature (15). Our paper-based questionnaire was refined after these discussions, to focus and standardize the elicitation process (see Appendix 3).

Data were collected through face-to-face interviews with individuals recruited at a cardiology conference (after a session devoted to discussion on advanced device treatments for arrhythmias at the Heart Rhythm Congress 2009). This provided an opportunity to clarify the questions if required. Participants were first asked about their clinical expertise and
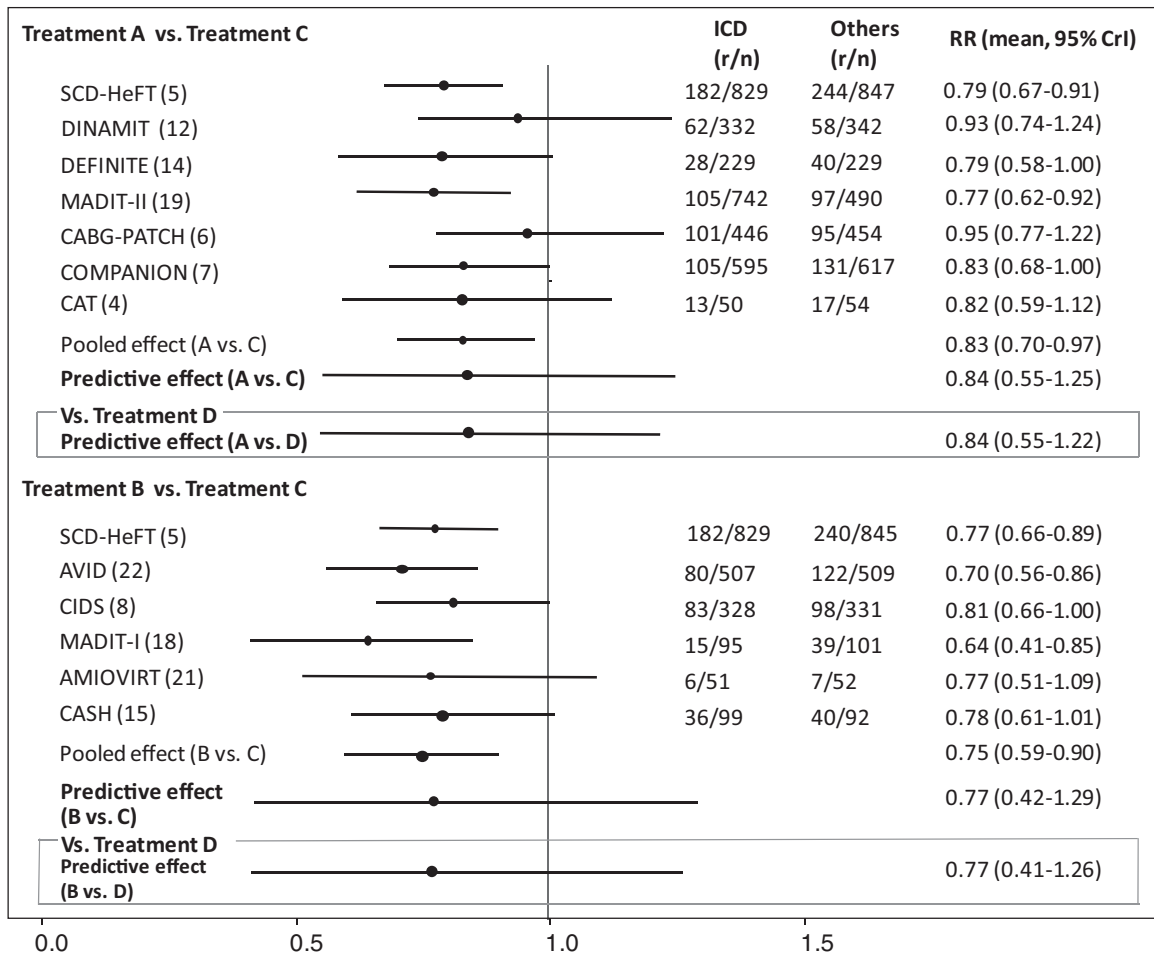
Figure 2. Mixed treatment comparison of the conventional ICD (Treatment C) and other therapies (Treatments A & B), and the predicted new ICD effect. Treatment A: Conventional drug therapy; Treatment B: Anti-arrhythmic drug therapy; Treatment C: Conventional ICD therapy; Treatment D: Four-pole connector ICD therapy. Dashed lines: the predictive posteriors estimated from the MTC model with and without adjustment for β.

familiarity with ICD implantation. They were then given some simple summary information about the 5-year mortality rate with "conventional" ICDs, based on the results of two, large trials. The experts were then asked to estimate the most likely value, and lower and upper limits, for the relative mortality rate with the four-pole IS-4 ICD. Finally, they were asked to distribute 100 points along a line showing different numbers of lives gained or lost with the four-pole device in comparison with a "conventional" ICD to reflect their subjective beliefs. The comparison against "conventional" ICD was intended to simplify elicitation and analysis, although it does neglect possible differences between current ICD devices. We used statistical tests of heterogeneity on the trial data to look for evidence of such differences.

As the model estimates the logarithm of relative mortality, elicited absolute mortality values were converted to an appropriate scale (see Appendix 1). Microsoft Excel and R software tools were then used to calculate a combined discrete distribution for the difference-adjustment parameter, and to test and fit a normal distribution, as described above.

RESULTS

Literature Review

The systematic review by Ezekowitz and colleagues (9) identified twelve RCTs that reported all-cause deaths during the follow-up period and the number of study participants for each treatment arm (3–7;11;13;14;17;18;20;21).

Ezekowitz et al. made pair-wise comparisons between ICD and non-ICD treatments. However, for our MTC analysis, we further classified the non-ICD treatments as conventional drug therapy alone (A) and conventional drug therapy combined with a Class III Anti-arrhythmic drug therapy (B). One study, the CASH trial (14), reported data on three randomized groups: (i) metoprolol (a beta blocker), (ii) amiodarone (a Class III anti-arrhythmic), and (iii) conventional ICD. We excluded the metoprolol arm, because this did not fit our definition of "conventional drug therapy" (treatment A), because it was not also co-prescribed in the other treatment arms. The final list of data included in our analysis is shown in Figure 2.

Table 1. The Posterior Mean (95% Credible Interval) of Selected Relative Effectiveness Measures Regarding Mortality (τ^2: Between-Study Variance, i.e., Heterogeneity Parameter)

	ICDs vs. Treatment B		ICDs vs. Treatment A	
a) Existing ICDs (Treatment C)	**Primary and secondary prevention trials combined (12 trials)** ($n = 9266$)			
	Log RR	−0.30 (−0.52, −0.11)	Log RR	−0.20 (−0.36, −0.03)
	RR	0.75 (0.59, 0.90)	RR	0.83 (0.70, 0.97)
	Predictive RR	0.77 (0.42, 1.29)	Predictive RR	0.84 (0.55, 1.25)
	$\tau^2 = 0.031$,	Posterior mean of residual deviance = 26.08		
b) Four-pole connector ICDs (Treatment D)	**Primary and secondary prevention trials combined (12 trials) with β adjustment** ($n = 9266$)			
	Log RR	−0.30 (−0.52, −0.12)	Log RR	−0.19 (−0.36, −0.03)
	RR	0.74 (0.60, 0.89)	RR	0.83 (0.70, 0.97)
	Predictive RR	0.77 (0.41, 1.26)	Predictive RR	0.84 (0.55, 1.22)
	$\tau^2 = 0.031$,	Posterior mean of residual deviance = 26.55		

Note. Treatment A: Conventional drug therapy; Treatment B: Anti-arrhythmic drug therapy; Treatment C: Conventional ICD therapy; Treatment D: Four-pole connector ICD therapy.

Primary versus Secondary Prevention

Ezekowitz et al. (9) reported similar estimates for the relative risk of mortality in the primary prevention trials as in the secondary prevention trials. They noted that the extent of heterogeneity across all primary and secondary prevention trials was moderate ($I^2 = 44.4$ percent), although the primary prevention trials showed an intrinsically wider variance.

We initially conducted separate Bayesian meta-analyses for primary and secondary prevention trials to test whether they are sufficiently homogeneous to be combined. For this analysis, we excluded trials for which the control arm was Treatment A because no secondary prevention trials included this treatment. The results of this C-B comparison showed little difference between the estimated effects for primary prevention (relative risk [RR] 0.72 (0.22–1.81), based on 3 trials (4;17;20) $N = 1973$) and those for secondary prevention (RR 0.81 (0.41–1.46), based on 3 trials (7;14;21), $N = 1866$). (Note that the differences between these results and those reported in (9) are due to differences in the meta-analysis models (24), and in the data included – we excluded three primary prevention trials with a Treatment A control arm, and the metoprolol arm from the CASH study, as mentioned above.)

Although this analysis suggested a relatively higher level of heterogeneity between the primary prevention studies ($\tau^2 = 0.69$) than between the secondary prevention studies ($\tau^2 = 0.27$), pooling all these trials resulted in a much lower level of heterogeneity ($\tau^2 = 0.03$). This led us to conclude that the use of ICD technology results in similar mortality benefits in primary and secondary prevention, and, therefore, we combined these results.

Meta-analysis Results for Existing ICDs

Our main MTC model included 12 trials: nine primary prevention trials (3–6;11;13;17;18;20) and three secondary prevention trials (7;14;21).

The forest plot of treatment effects for all comparisons involving the existing ICD therapy (Treatment C) and the pooled estimates of mortality benefits from the use of existing ICDs are shown in Figure 2 and Table 1, respectively. Note that, in Figure 2, the results for some studies (particularly DINAMIT, CABG-PATCH, and MADIT-I) appear rather different to what one might expect from the raw data, because of a "shrinkage effect" (1). The relative risk estimate for the C-B comparison (mean = 0.75 [95 percent CrI 0.59–0.90]) is lower than that for the C-A comparison (mean = 0.83 [95 percent CrI 0.70–0.97]), indicating a higher mortality risk associated with the Class III anti-arrhythmic drug therapy (B) than that for conventional drug therapy alone (A) (mean = 1.12 [95 percent CrI 0.89–1.42]).

(a) Combined subjective distribution

(b) Normal CDF vs. subjective CDF

Dots: cumulative probabilities for mid-interval values from the combined histogram

Line: Fitted Normal distribution CDF values

β= log(mortality on old ICD / mortality on new ICD); *Mortality on new ICD* = (Lives saved or lost/1000) + mortality on old ICD
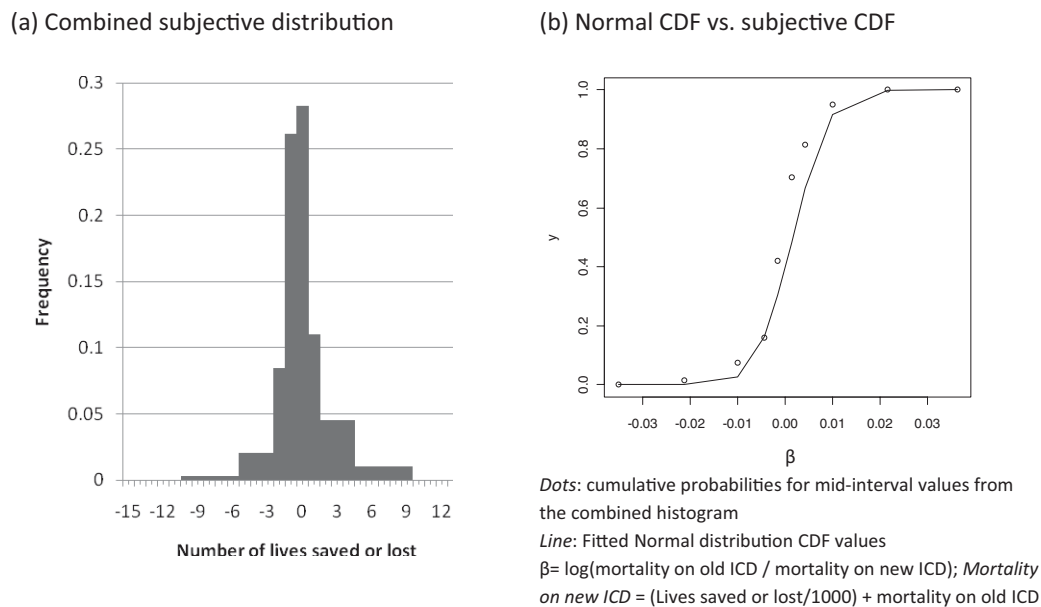
Figure 3. Combined experts' subjective distributions for the difference in mortality with the new four-pole ICD compared with conventional ICD and the fitted Normal distribution.

(Similar findings were also reported in Tung et al. (22).) The predictive posteriors for a future trial on conventional ICDs show wider intervals: 0.77 (0.42–1.29) for the C-B comparison and 0.84 (0.55–1.25) for the C-A comparison. Between-study variability was moderate ($\tau^2 = 0.031$).

The model predictions were considered to offer adequate goodness of fit when the posterior mean residual deviance of 26.08 was compared to the number of data points, 25. For all meta-analyses, standard checks for model convergence led to the use of 10,000 iterations as a burn-in, and the results presented were obtained based on 20,000 iterations.

Elicitation of Expert Opinion

Seven experts provided valid probability distributions for the mortality difference with the four-pole ICD. We cannot calculate a response rate, as the invitation to participate was made in an open conference session and two of the authors (K.H. and J.Y.) approached individuals after the session. However, most of those approached did agree to offer their opinion and complete the elicitation form.

The mean and standard deviation of the fitted distribution for β were both close to zero, i.e., Normal ($\mu_\beta = 0.0017$, $\sigma_\beta = 0.0060$), reflecting expert opinion that the mortality impact of the new ICDs would be similar to that of current devices. However, the positive mean value signifies the belief that the new device might slightly reduce mortality.

The pooled cumulative subjective probabilities from the experts were compared against the cumulative density function of the Normal distribution (Figure 3). Although a weak degree of skewness is observed in the empirical probability distribution, the assumption that β follows a normal distribution appears to be reasonable.

Adjusted Meta-analysis: Prediction for New ICD

The fitted Normal distribution for β was incorporated into the meta-analysis to estimate the effectiveness of the four-pole connector ICDs. The adjusted MTC results are provided in Table 1.

The estimates in Table 1 show extremely small differences between the estimated mortality effects of the old and new ICDs: RR of 0.75 (0.59–0.90) for C versus B and 0.74 (0.60–0.89) for D versus B. This was expected due to the near-zero mean and standard deviation estimates of β. Also, the predictive distributions for future trials of the new ICD were very similar to those for existing ICDs (Figure 2). The between-trial variance, τ^2, was estimated as having a mean of 0.031(on the log RR scale). The Monte Carlo standard errors of the mean estimates, were as small as 0.003 for the D-B comparison.

DISCUSSION

We have presented a Bayesian, random-effects, MTC meta-analysis incorporating pertinent external information. This can be used for the evaluation of new versions of medical devices, for which trial data is not yet available. It adapts an established technique of plugging a functional relationship between a target parameter and an adjustment parameter into an evidence synthesis model (1). Based on the current literature on bias modeling (23;25), an intuitively simple additive parameter was used to capture expected departures of the effectiveness of the new device from the old one.

We illustrated this approach, obtaining mortality estimates for a new ICD device based on a mixture of expert opinion and empirical evidence on existing ICDs. This example was chosen as there is relatively large existing evidence base for ICDs, while further data collection on a new variant may be expensive and

difficult due to the invasive nature of the technology. In this example, the incorporation of expert opinion on the new technology made very little difference. However, the proposed method could potentially change a technology adoption decision, depending on how positive and strong the experts' beliefs are.

Whether the suggested method should help to inform manufacturers' decisions about development, or payers' decisions about reimbursement, depends on the intrinsic reliability of expert opinion—are they right in their expectations about the worth of a new version of a device more often than they are wrong? It also depends on the robustness of methods used to elicit expert opinion and to combine it with prior evidence.

Even well-designed elicitation processes can be subject to cognitive or behavioral bias, and there are further uncertainties associated with the selection of experts and methods of synthesizing elicited data. There is not currently a prescription for how best to elicit and pool expert opinion. In our example, we recruited experts from a relevant conference. These individuals might not reflect the spectrum of beliefs of cardiac specialists, which could explain the relatively small within- and between-expert variation. The use of individual interviews, rather than an actual or nominal group approach might also have influenced the results; with feedback and/or a chance for discussion, opinions might have evolved (19). The specific design and wording of our questionnaire might also have had an impact. For example, we informed participants that there is a fixed 5-year mortality rate of 35 percent with existing version of ICDs, which may underestimate uncertainty over β.

The chosen effectiveness parameter in our analysis was all-cause mortality. However, the dominant opinion of the interviewed experts was that the change in the lead connector is more likely to be associated with other complication rates, as the modified lead may be related to higher rates of lead dislodgement and insulation defects. If other effectiveness measures had been chosen, it is possible the experts might have anticipated a "substantial departure from current ICD connector systems" (10). This suggests that preliminary, qualitative discussion with experts is very important to identify the key parameters of interest. It would also be relatively straightforward to elicit a range of outcomes, including resource use or cost estimates required for an early assessment of cost-effectiveness. However, elicitation of probability distributions is time-consuming, so care is needed not to overburden respondents.

Validity of the meta-analysis model must also be considered. Various assumptions on the model structure could have been made. An overarching assumption is that heterogeneity in the data associated with the earlier versions of ICDs is less than the difference between the new IS-4 device and the earlier versions. Although around 44 percent of the total variation in trial estimates that is due to heterogeneity is considered reasonable in the current study, there is much variation in the levels of doses of anti-arrhythmic drugs and other adjuvant therapies such as the use of beta blockers across studies. There are several ways to address such heterogeneity problems (8). As in any systematic review and meta-analysis, it is also important to consider the potential for other forms of bias; such as publication bias.

We also note that, whereas Bayesian incorporation of expert opinion can help consolidate existing opinion and data, these methods cannot substitute for long term evaluation of rare adverse events. Post-marketing follow-up studies are essential in ensuring that new variants of a device do not impose additional harms. Expert opinions on rare adverse events have limited use. Because of the rarity of these events, not all experts will have encountered them. Furthermore, the issue of when a difference in a new variant of a device is sufficiently large to warrant primary randomized controlled trials is fundamental. In relation to this, the proposed method has another potential use: if future research generates evidence on the new variants, the estimates obtained by this method can also serve as prior information in the sequential use of Bayesian evidence synthesis.

CONCLUSIONS

Despite these caveats, the suggested method may play a useful role in assisting decision makers to make early informed judgments about evolving healthcare technologies. Before any formal clinical evidence is gathered, quantified opinion can provide some indication of expected benefits from early adoption and further evidence collection. This method provides a means for formally including experts' opinion into an evidence synthesis model, which can be updated as more evidence emerges.

CONTACT INFORMATION
Ji-Hee Youn, MSc, Research Fellow, Joanne Lord, PhD, Reader, Health Economics Research Group, Brunel University, Uxbridge, United Kingdom
Karla Hemming, PhD, Senior Lecturer, Alan Girling, MA, Senior Research Fellow, Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham, United Kingdom
Martin Buxton, BA, Professor, Health Economics Research Group, Brunel University, Uxbridge, United Kingdom

APPENDIX 1: MATHEMATICAL REPRESENTATION OF THE MTC MODEL WITH ADJUSTMENT FOR THE NEW TECHNOLOGY
For each trial i, the included treatment earliest in the alphabet (labeled with the index b below) was taken as the baseline treatment (e.g., in trials comparing

A with B, b = A). The Bayesian MTC model, with adjustment for a new, modified device, is shown in Eq. 1.

$$
\begin{aligned}
& r_{k,i} \sim Binomial(p_{k,i}, n_{k,i}) \\
& \log(p_{k,i}) = \begin{cases} \mu_{ib} & b = A, B, C. \text{ if } k = b \\ \mu_{ib} + \delta^{U}_{ikb} & \text{if } k \text{ alphabetically after } b \end{cases} \\
& \delta^{U}_{ikb} = \min\{\delta_{ikb}, -\mu_{ib} = -\log(p_{b,i})\} \\
& \delta_{ikb} \sim Normal(d_{kb} + \beta_i, \tau^2) \sim Normal((d_{Ak} - d_{Ab}) + \beta_i, \tau^2) \quad \text{(Eq. 1)} \\
& \beta_i \sim Normal\left(\mu_\beta, \sigma^2_\beta\right)
\end{aligned}
$$

$$
\begin{aligned}
& Priors: -\mu_{ib} \sim Gamma(.001, .001), \\
& d_{Ak} \sim Normal(0, 10000), \text{ where } k = B, D, \\
& \tau \sim Uniform(0, 2)
\end{aligned}
$$

where $r_{k,i}$ are the number of deaths out of the total number of patients at risk $n_{k,i}$ for each treatment k in study i; $p_{k,i}$ is the probability of death in study i under treatment k; μ_{ib} is the logarithm of baseline mortality risk for trial i. The parameter δ_{ikb} represents the log relative risk (LRR) of treatment k relative to baseline treatment b for study i. The nuisance parameter δ^{U}_{ikb} is included to constrain the range of probabilities of an event below 1 (see (40)). These trial-specific LRR are drawn from a common distribution, and the pooled LRR, d_{bk}, is defined with respect to the reference treatment A (e.g. $d_{BC} = d_{AC} - d_{AB}$). The inclusion of β signifies the mean mortality effect of the conventional device estimated from observed data is deviated from the effect of the new device, d_{kb} by the magnitude of β.

Finally, a flat Gamma prior was specified for the negative of μ_{ib} to constrain the range of μ_{ib} below 0, and Normal priors for the 'basic parameters' whose comparators are the reference treatment A. The between-study variance (heterogeneity parameter), τ^2, was assigned a Uniform prior. The potential correlation between relative effectiveness parameters for trials involving more than two arms were not considered in this study for simplicity.

Expert opinions form the basis of estimating μ_β and σ_β. Although β_is represent the difference between two log relative risks, actual questions asked experts to directly compare the proportion of patients who would die under the modified device therapy with that under the conventional device treatment. This was justified by the relationship $E(\delta_{iCb}) = d_{Db} + E(\beta_i) = d_{Db} + \mu_\beta$, where both δ and d denote the log relative risks, but δ compares the baseline treatment with treatment C while d does with treatment D. Taking away the logarithm from the above relationship gives:

$$
e^{\beta_i} = \frac{E\left(RR^C_i\right)}{RR^D} \approx \frac{P_C/P_{Control}}{P_D/P_{Control}} = \frac{P_C}{P_D}
$$

P is a risk of death and RR a relative risk. The mortality risk under control treatment, $P_{Control}$, cancels out, making possible the direct comparison between treatments C and D, i.e., P_C and P_D.

APPENDIX 2: FULL BUGS CODE FOR MTC MODEL WITH ADJUSTMENT PARAMETER INCORPORATED

```
model{
# Model defined at treatment arm level (Narm = 25)
for(i in 1:Narm)
{
log(p[i]) <- -mu[s[i]] + delta.u[i] * (1-equals(t[i],b[i]))
r[i] ~ dbin(p[i],n[i])
rhat[i] <- p[i]*n[i]
dev[i] <- 2*(r[i]*(log(r[i])-log(rhat[i])) + (n[i]-r[i])*(log(n[i]-r[i])
    - log(n[i]-rhat[i])))
```

```
delta.u[i] <- min(delta[i], -mu[s[i]])
delta[i] ~ dnorm(md[i], prec)
md[i] <- d.study[i] + beta[s[i]]*equals(t[i], 3)
# beta to be included only when the treatment is ICD (trt = 3)
d.study[i] <- d[t[i]] - d[b[i]]
}
#Information on distribution of beta
mu.beta <- 0.001680164
prec.beta <- 1/ pow(0.006045638, 2)
# Priors and adjustment at the study level (Nstud = 12)
for(j in 1:NStud)
{
mu[j] <- 0 - mu.minus[j]
mu.minus[j] ~ dgamma(0.001, 0.001)
beta[j] ~ dnorm(mu.beta, prec.beta)
}
#Treatment level prior (NTrt = 3)
d[1] <- 0
for (k in 2:NTrt)
{
d[k] ~ dnorm(0,0.0001)
}
#Predictive distribution
d.pred[1] <- 0
for (k in 2:NTrt)
{d.pred[k] ~ dnorm(d[k],prec) }
#Pairwise LRR and RR: new ICD
for (c in 1:2) {
for (k in (c+1):3) {
lrr[c,k] <- d[k]-d[c]
log(rr[c,k]) <- lrr[c,k]
lrr.pred[c,k] <- d.pred[k]-d.pred[c]
log(rr.pred[c,k]) <- lrr.pred[c,k]
}
}
prec <- 1/ tau.sq
tau.sq <- tau*tau
tau ~ dunif(0,2)
resdev <- sum(dev[]) #Residual deviance
}
```

```
#Data
list(Narm = 25, NStud = 12, NTrt = 3)
```

s[]	t[]	r[]	n[]	b[]
1	1	244	847	1
1	2	240	845	1
1	3	182	829	1
2	1	58	342	1
2	3	62	332	1
3	1	40	229	1
3	3	28	229	1
4	1	97	490	1

4	3	105	742	1
5	1	95	454	1
5	3	101	446	1
6	1	131	617	1
6	3	105	595	1
7	1	17	54	1
7	3	13	50	1
8	2	122	509	2
8	3	80	507	2
9	2	98	331	2
9	3	83	328	2
10	2	39	101	2
10	3	15	95	2
11	2	7	52	2
11	3	6	51	2
12	2	40	92	2
12	3	36	99	2
END				

APPENDIX 3: SAMPLE QUESTIONNAIRE USED TO ASSIST VERBAL INTERVIEWS

What is your opinion of the likely impact of the four-pole connector on mortality?

According to the SCD-HeFT* and AVID** studies, on average around 35% of ICD patients (both primary and secondary prevention) currently die within 5 years due to cardiac and non-cardiac causes.
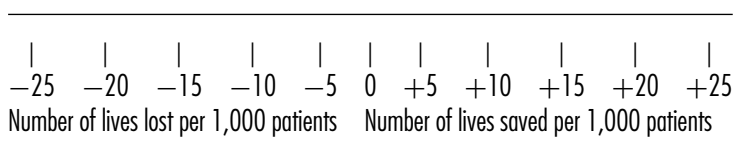
Do you think it is likely that 5-year mortality would improve as a result of using a four-pole connector, compared to the current connector system?

Yes ☐ No ☐ Don't know ☐

Suppose that 1,000 patients were to have an ICD implanted. How many deaths, on average, do you think might be averted/caused over a 5 year period if a four-pole connector was used instead of a conventional ICD?

Minimum [number] more/ less (tick) deaths	Your Best Guess [number] more/ less (tick) deaths	Maximum [number] more/ less (tick) deaths

Please try to represent your opinion on the scale below by distributing a total of 100 points between the lines to reflect your subjective beliefs about each range occurring.

−25	−20	−15	−10	−5	0	+5	+10	+15	+20	+25

Number of lives lost per 1,000 patients   Number of lives saved per 1,000 patients

REFERENCES

1. Abrams RK, Lambert CP, Shaw C, Sanso B. Meta-analysis of heterogeneously reported study results: A Bayesian approach. In: Stangl DK, Berry DA, eds. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker; 2000:29-63.
2. Ades AE, Welton NJ, Caldwell D, et al. Multiparameter evidence synthesis in epidemiology and medical decision-making. *J Health Serv Res Policy.* 2008;13(Suppl 3):12-22.
3. Bansch D, Antz M, Boczor S, et al. Primary prevention of sudden cardiac death in idiopathic dilated cardiomyopathy: The cardiomyopathy trial (CAT). *Circulation.* 2002;105:1453-1458.
4. Bardy GH, Lee KL, Mark DB, et al. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med.* 2005;352:225-237.
5. Bigger JT Jr. Prophylactic use of implanted cardiac defibrillators in patients at high risk for ventricular arrhythmias after coronary-artery bypass graft surgery. coronary artery bypass graft (CABG) patch trial investigators. *N Engl J Med.* 1997;337:1569-1575.
6. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med.* 2004;350:2140-2150.
7. Connolly SJ, Gent M, Roberts RS, et al. Canadian implantable defibrillator study (CIDS): A randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation.* 2000;101:1297-1302.
8. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. 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;28:1861-1881.
9. Ezekowitz JA, Rowe BH, Dryden DM, et al. Systematic review: Implantable cardioverter defibrillators for adults with left ventricular systolic dysfunction. *Ann Intern Med.* 2007;147:251-262.
10. Hauser RG, Almquist AK. Learning from our mistakes? Testing new ICD technology. *N Engl J Med.* 2008;359:2517-2519.
11. Hohnloser SH, Kuck KH, Dorian P, et al. Prophylactic use of an implantable cardioverter-defibrillator after acute myocardial infarction. *N Engl J Med.* 2004;351:2481-2488.
12. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: An introduction to mixed treatment comparisons. *Value Health.* 2008;11:956-964.
13. Kadish A, Dyer A, Daubert JP, et al. Prophylactic defibrillator implantation in patients with nonischemic dilated cardiomyopathy. *N Engl J Med.* 2004;350:2151-2158.
14. Kuck KH, Cappato R, Siebels J, Ruppel R. Randomized comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from cardiac arrest : The cardiac arrest study hamburg (CASH). *Circulation.* 2000;102:748-754.
15. Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: An applied example. *Value Health.* 2007;10:195-203.
16. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput.* 2000;10:325-337.
17. Moss AJ, Hall WJ, Cannom DS, et al. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. multicenter automatic defibrillator implantation trial investigators. *N Engl J Med.* 1996;335:1933-1940.
18. Moss AJ, Zareba W, Hall WJ, et al. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *N Engl J Med.* 2002;346:877-883.
19. O'Hagan A, Buck EC, Daneshkhah A, et al. *Uncertain Judgements: Eliciting Experts' Probabilities.* New York: John Wiley & Sons; 2006.
20. Strickberger SA, Hummel JD, Bartlett TG, et al. Amiodarone versus implantable cardioverter-defibrillator: Randomized trial in patients with nonischemic dilated cardiomyopathy and asymptomatic nonsustained ventricular tachycardia–AMIOVIRT. *J Am Coll Cardiol.* 2003;41:1707-1712.

21. The Investigators of the Antiarrhythmic versus Implantable Defibrillators (AVID) Trial. A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. the antiarrhythmics versus implantable defibrillators (AVID) investigators. *N Engl J Med.* 1997;337:1576-1583.

22. Tung R, Zimetbaum P, Josephson ME. A critical appraisal of implantable cardioverter-defibrillator therapy for the prevention of sudden cardiac death. *J Am Coll Cardiol.* 2008;52:1111-1121.

23. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc.* 2009;172:21-47.

24. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Stat Med.* 2002;21:1601-1623.

25. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc*. 2009;172:119-136.