Cost-Effectiveness Analysis Using Data from Multinational Trials: The Use of Bivariate Hierarchical Modeling
Andrea Manca, Paul C. Lambert, Mark Sculpher and Nigel Rice
Med Decis Making 2007; 27; 471 originally published online Jul 19, 2007;
DOI: 10.1177/0272989X07302132

The online version of this article can be found at:
http://mdm.sagepub.com/cgi/content/abstract/27/4/471
Health care economic evaluation studies benefit considerably from access to individual patient data (IPD) from multinational randomized controlled trials (RCTs). In the context of some health care interventions—in particular, pharmaceuticals—RCTs are often multinational in design. Such studies offer several advantages over single-country counterparts. These include the opportunity to reach rapidly the required target sample size and meeting the needs of regulatory agencies in different jurisdictions. Furthermore, because these studies enroll patients from a wide range of countries, multinational trials are thought to have the potential to increase the geographical generalizability of clinical and cost-effectiveness estimates.

However, 2 sets of arguments suggest that this viewpoint may not be totally defensible. The first relates to the fact that policy makers are essentially jurisdiction specific, and in most instances, this accords with being country specific. An increasing number of health care systems now require economic evidence to make country-specific decisions regarding the reimbursement of specific health technologies, which suggests that trial-wide cost-effectiveness results may not be relevant to any specific jurisdiction. Countries, in fact, will inevitably differ in factors such as availability of health care resources, clinical practice, patients’ case mix, and the costs of delivering health care. These factors can cause the value for money of a health care intervention to vary between different countries in a multinational study, leaving country-specific policy makers uncertain about the extent to which trial-wide cost-effectiveness studies can be used to assess the between-location variability of the study results while controlling for differences in country-specific and patient-specific characteristics. It is demonstrated that results from standard CEA using IPD from multinational trials display a large degree of variability across the 17 countries included in the analysis, producing potentially misleading results. In contrast, “shrinkage estimates” obtained from the modeling approach proposed here facilitate the appropriate quantification of country-specific cost-effectiveness estimates while weighting the results based on the level of information available within each country. The authors suggest that the methods presented here represent a general framework for the analysis of economic data collected from different locations.

Key words: cost-effectiveness analysis; statistics; generalizability; clinical trials.
(i.e., pooled) cost-effectiveness results are relevant to their own jurisdiction.

The second set of arguments relates to the appropriateness of the methodology used to analyze IPD alongside multinational trials. Patients are often recruited and treated (in a particular center) in a given country, and hence cost-effectiveness data take on a natural hierarchical structure. Failure to acknowledge these features of the data may lead to misleading conclusions with regard to the generalizability of the study results. In other words, it can be argued that there is likely to be some degree of statistical heterogeneity in the country-specific cost-effectiveness results between different jurisdictions participating in a multinational RCT, and it is implausible to assume a priori that differential costs and health outcomes are the same in each country.

Three important questions must be addressed when assessing the geographical generalizability of the conclusions of the economic component of any health technology assessment (HTA). First, what are the appropriate methods to analyze IPD relating to cost-effectiveness data from multinational RCTs? Second, how should we assess the extent to which trial-wide results are generalizable between countries participating in the same study? Finally, how do we produce country-specific cost-effectiveness estimates that are directly relevant to local decision makers?

Various analytical strategies, most of which use regression methods, have been proposed to address the above questions. Early approaches ignored the natural clustering in the data and failed to implement methodologies that could facilitate the estimation of the between-country variability in the results. A series of recent studies advocated the use of hierarchical modeling to analyze cost-effectiveness IPD from multilocational trials while simultaneously allowing for (potential) between-location variability in the data. Pinto and others recently explored alternative estimation methods to obtain country-specific estimates of cost-effectiveness, conducting a simulation exercise based on summary estimates of country-level differential costs and effects from large multinational RCTs. Grieve and others compared ordinary least squares (OLS) and hierarchical models in the analysis of cost data from 11 European countries and proposed a generalized linear mixed model to address the skewed nature of the length of stay and cost data. Finally, Manca and others extended the net benefit regression to accommodate the hierarchical structure of economic data in multilocational trials, showing how hierarchical models can be used to obtain trial-wide and location-specific estimates of incremental cost-effectiveness while correctly quantifying the measures of sampling uncertainty around these mean estimates.

This article advocates the use of Bayesian bivariate hierarchical models (BHLMs) for the analysis of multinational cost-effectiveness data. In view of the need to investigate potential between-country differences in cost and health outcomes, it is demonstrated that BHLM can be used to accommodate information at subject and country levels to facilitate robust estimation of overall (i.e., trial-wide) and country-specific differential mean costs and effects. The proposed approach is illustrated using data from the Assessment of Treatment with Lisinopril and Survival (ATLAS) study, a large multinational trial comparing low versus high doses of the angiotensin-converting enzyme (ACE) inhibitor lisinopril in patients with chronic heart failure.

The article begins by briefly summarizing existing approaches for RCT-based cost-effectiveness analysis (CEA). It then describes the bivariate nonhierarchical regression methods for CEA proposed in the literature. Then, the rationale for extending the use of BHLM techniques to the analysis of IPD from multinational cost-effectiveness data is outlined. A case study is presented, and then the results from the application of nonhierarchical and hierarchical bivariate models to the case study are discussed. The final section summarizes the contributions of the article, contextualizing these in light of the current methodological debate in medical economic evaluation.

ANALYTICAL STRATEGIES FOR CEA USING INDIVIDUAL PATIENT DATA FROM MULTINATIONAL TRIALS

Despite the existence of factors that could potentially affect the between-location variability of the study results, most CEAs using IPD collected alongside multinational RCTs have been conducted assuming that the clinical effectiveness of the intervention does not differ greatly (at least in relative terms) across countries and that pooling clinical data across countries (and centers) to assess the effect of treatment on clinical outcomes is an acceptable strategy. Similarly, on the cost side, the standard
method has been to apply a single set of unit costs (typically taken from 1 country) to all patient-specific resource use data collected in the study, regardless of where this resource use had taken place. Pooled costs are related to pooled outcomes to obtain a trial-wide (or pooled) measure of cost-effectiveness. Results for another jurisdiction are then obtained by multiplying the pooled trial wide resource use data by resource-specific unit costs taken from this other jurisdiction. The above practice implicitly assumes that resource use and effectiveness data are perfectly transferable\(^b\) between countries/jurisdictions\(^14\). Despite the popularity of this approach, the proliferation of country-specific reassessments of previously published trial-wide CEAs suggests that local decision makers are often uncertain about the “transferability” of the trial-wide results to their own jurisdiction\(^c\).

An alternative approach, sometimes adopted in multinational studies, assumes that resource use data are not at all transferable between locations, whereas effectiveness data are. Analyses for a particular country/jurisdiction would produce cost estimates by multiplying unit costs for a given country by resource use of patients recruited in the same country while relating these costs to trial-wide effectiveness data (see, e.g., Bjorholt and others\(^15\)). The 2 main limitations with this approach are that, first, it is only feasible for trials that recruited a reasonably large number of patients in the country of interest, and second, it ignores the fact that IPD on costs and effects are naturally correlated. Analyzing 2 different samples for costs and effects will introduce bias in the estimation of the joint distribution of the mean differential costs and effects, influencing the analyst’s ability to correctly quantify the sampling uncertainty surrounding the measure of cost-effectiveness\(^14\).

A third and less commonly used approach to the analysis of multinational cost-effectiveness IPD is to focus on only one of the countries/jurisdictions in the trial, estimating mean costs and effects solely using patient data from such a country (examples can be found in Liljas and others\(^16\) and Mark and others\(^17\)). Although overcoming the limitation of the previous approach, this strategy implicitly assumes that country-specific data are not at all transferable between locations. Most of the time, this is impracticable for countries with a low recruitment rate in the trial. Furthermore, it makes a suboptimal use of the data by discarding potentially relevant information from other countries.

In practical terms, it is likely that IPD from multinational trials will be partially transferable. That is, some components of resource use and health outcomes will be common between locations, whereas others will be more country specific. Analytically, the problem is to disentangle the extent to which CEA data for one country/jurisdiction are generalizable to another. Attempts to address the issue have focused on the analysis of individual patient resource use and cost data only rather than full cost-effectiveness.

Willke and others\(^18\) were the first to use regression methods for the simultaneous analysis of cost and effects data collected alongside multinational RCTs. To explore the between-country variability in the results, their regression model included country-by-treatment and country-by-outcome interaction terms, which facilitated country-specific estimation of mean differential costs and effect. The regression estimates were then used to calculate incremental cost-effectiveness ratios (ICERs) using 1) own-country costs and effectiveness results, 2) own-country costs and trial-wide effectiveness results, or 3) own-country costs but trial-wide resource utilization and effectiveness. The authors found that greatest variation in the country-specific ICER was in the first case, where country-specific resource utilization, unit prices, and outcome levels were all taken into account.

In a recent article, Willke and others\(^18\) used hierarchical modeling to obtain empirical Bayes estimates of country-specific differential costs from summary data derived from a large multinational trial. The authors compared and contrasted 1) overall random effect estimate, 2) the country-level estimates obtained from a stratified analysis (i.e., similar to having a dummy variable for each country), and 3) the empirical Bayes estimates obtained from a random effect model, finding that estimates obtained with (3) to be more informative for the Canadian decision maker. The method in (1) and (3) is essentially similar to a random effect meta-analysis\(^19\).
which assumes that study-specific (country-specific, in the case of the models applied to multinational IPD) estimates are drawn from a distribution of possible realizations of the study-specific treatment effect (country-specific differential cost), which follows a normal distribution. Using the same summary data, Pinto and others\(^5\) conducted a simulation exercise to explore the validity of univariate and bivariate shrinkage models for differential costs and effects derived using an empirical Bayes approach.

Grieve and others\(^7\) and Manca and others\(^8\) recently proposed the use of hierarchical modeling to analyze, respectively, individual-level cost and cost-effectiveness data collected alongside multinational trials. The authors justified their proposed approach on the basis of 2 main considerations: first, that patients treated in the same site are expected to be more similar (in terms of resource use and clinical outcome) than those treated in other locations\(^20,21\) and second, that locations may vary markedly in factors that influence the resource use, unit costs, and outcome data observed in trials.\(^1,2\) Manca and others\(^8\) showed how to use hierarchical regression analysis to obtain trial-wide and location-specific estimates of cost-effectiveness, using center-specific predictions to explore the between-location variability in the results. This approach was developed in the context of a univariate model for net benefit and did not use any patient- or center-specific covariates to explain the observed between-center variability in the results.

This article extends the approach proposed by Manca and others\(^8\) by developing a Bayesian bivariate hierarchical regression model for the analysis of individual patient-level cost-effectiveness data collected alongside multinational trials, showing how to incorporate patient- and country-specific information in the model.

**BIVARIATE (NONHIERARCHICAL) MODELING APPROACHES FOR TRIAL-BASED CEA**

Willan and others\(^22\) recently extended the net benefit regression framework proposed by Hoch and others\(^9\) using of a system of seemingly unrelated regression (SUR) equations for IPD on costs and health outcomes. This (bivariate) approach has 3 main advantages. First, it facilitates explicit modeling of both costs and effects while allowing the inclusion of a set of covariates in the 2 equations. Second, it exploits the existence of correlation, at the patient level, between costs and effects, thereby improving the efficiency of the estimation process when this correlation is different from zero. Third, unlike the standard (univariate) net benefit regression, SUR does not require a new regression to be estimated for every value of the cost-effectiveness threshold (i.e., \(\lambda\)). Using a fully parametric approach, the authors showed how the treatment effect (in terms of differential mean costs and effects) is estimated from a regression model in which the error terms for the cost and effects equations are assumed to follow a bivariate normal (BVN) distribution with (zero) mean and a given covariance matrix.\(^d\)

A more flexible formulation of the nonhierarchical bivariate model for cost-effectiveness IPD has been proposed by O’Hagan and others\(^23\) within a Bayesian statistical framework. Their model can be illustrated as follows. Let the vector of individual-specific costs (\(C_t\)) and effects (\(E_t\)), receiving the treatment \(t = (0, 1)\), be described as follows:

\[
\begin{pmatrix}
    E_t \\
    C_t
\end{pmatrix}
\sim
\text{BVN}
\left(
\begin{pmatrix}
    \mu_c \\
    \mu_e
\end{pmatrix},
\begin{pmatrix}
    \sigma^2_c & \sigma_{ce} \\
    \sigma_{ec} & \sigma^2_e
\end{pmatrix}
\right).
\]

where IPD on cost and effects are assumed to follow a BVN distribution, with the parameters \(\mu_c\), \(\mu_e\), \(\sigma^2_c\), \(\sigma^2_e\), \(\sigma_{ce}\), and \(\sigma_{ec}\) representing, respectively, the sample mean estimates for effects and costs in the 2 arms of the trial (0 = standard intervention; 1 = new intervention). The parameters \(\sigma^2_c\), \(\sigma^2_e\), and \(\sigma_{ce}\) are, respectively, the variance of the mean effects, the variance of the mean cost, and their covariance in treatment arm \(t\). Notice that, unlike the SUR model, this formulation does not require the covariance matrix to be the same in each treatment group.

In each treatment arm, mean cost and effects were assigned a noninformative joint multivariate normal prior distribution, whereas the covariance matrix was characterized by a Wishart distribution. The estimation of the model parameters was carried out using Markov chain Monte Carlo (MCMC) methods, deriving the information of interest from the posterior distribution of the differential mean costs \((\Delta_c = \mu_c - \mu_c^0)\) and effects \((\Delta_e = \mu_e - \mu_e^0)\).

Model (1) can be reparameterized using the so-called product normal formulation, which reinterpret...
the multivariate model into a series of related (univariate) linear models\(^{24-27}\) as follows:

\[
\begin{align*}
E_{ij} & \sim N(\mu_{ij}^e, \tau^e) \\
G_{ij} & \sim N(\mu_{ij}^g, \tau^g) \\
\mu_{ij}^e &= \gamma \\
\mu_{ij}^g &= \alpha + \theta_0 E_{ij} \\
\tau^e &= \sigma^2_e \\
\tau^g &= \sigma^2_g \\
\sigma^2_e &= \sigma^2_q + \tau^e_0 \\
\sigma^2_g &= \sigma^2_q + \tau^g_0 \\
\theta_0 &= 0 \\
\theta_1 &= 1
\end{align*}
\]  

(2)

which assumes (for each treatment arm) a Normal distribution for effects and a Normal distribution for costs, with the latter being conditional on the departure of the individual-specific effectiveness (i.e., \(E_i = E_{ij} - E_t\)) from its treatment-group–specific overall mean (\(E_t\)). The quantities \(\alpha\) and \(\gamma\) represent, respectively, the trial-wide mean cost and effect for the control arm.

When IPD on costs and effects follow a BVN distribution, expressions in model (2) can be backtransformed to obtain the parameters in model (1) as follows:

\[
\begin{align*}
\mu_{ij}^e &= \gamma \\
\mu_{ij}^g &= \alpha + \theta_1 E_{ij} \\
\sigma^2_e &= \sigma^2_q + \tau^e_1 \\
\sigma^2_g &= \sigma^2_q + \tau^g_1 \\
\theta_0 &= 0 \\
\theta_1 &= 1
\end{align*}
\]  

(3)

This formulation captures the correlation between costs and effects at the patient level through the terms \(\theta_0\) and \(\theta_1\), assuming that in each arm, the mean cost is linearly related to the departure of the patient-specific effectiveness from its group mean (i.e., \(E_i\)). The differences in mean cost (\(\Delta_e\)) and mean effects (\(\Delta_e\)) are subsequently combined as follows to estimate the incremental net benefit:

\[
\text{INMB}(\lambda) = \Delta_e \cdot \lambda - \Delta_c.
\]  

(4)

A Bayesian formulation of model (1) requires the specification of prior distributions for both the vector of mean cost and effects and the covariance matrix. The “textbook” definition of a prior distribution for the latter is in the form of a Wishart distribution.\(^{25,26}\) However, several authors have commented on the fact that the definition of a noninformative Wishart prior distribution for variance parameters is not a trivial task.\(^{25,26,28}\) One of the advantages of adopting the product normal formulation in model (2) is that more standard noninformative priors\(^{26}\) for variance parameters can be used.

The models discussed so far assume no differences between countries, and it is reasonable to expect some degree of clustering in multinational economic trials.\(^{7}\)

**BIVARIATE HIERARCHICAL LINEAR MODEL IN TRIAL-BASED CEA**

The model described in the previous section can be extended in 2 ways: first, by formally recognizing the hierarchical structure of the cost-effectiveness data, and second, by incorporating subject- and country-specific covariates to explain observed between-country variability in the economic results. The standard implicit assumption underpinning the validity of this class of models relates to the concept of exchangeability between the units conveying information on the parameter of interest (i.e., country-specific differential costs and effects). The implications of this assumption are considered in the Discussion section of this article. For a full discussion of the concept of exchangeability in Bayesian hierarchical models, the reader is referred to see Gelman and others\(^{25}\) and Spiegelhalter and others.\(^{26}\)

**Accounting for Hierarchical Data Structure**

We start by reformulating (1) to take into account the nested structure of the cost-effectiveness data (i.e., patients clustered within countries) as follows:

\[
\begin{align*}
E_{ij} & \sim \text{BVN} \left( \left[ \mu_{ij}^e, \mu_{ij}^g \right], \left[ \begin{array}{cc} \sigma^2_e & \sigma_{eg} \\ \sigma_{eg}^* & \sigma^2_g \end{array} \right] \right) \\
C_{ij} & \sim \text{BVN} \left( \left[ \mu_{ij}^e, \mu_{ij}^g \right], \left[ \begin{array}{cc} \sigma^2_e & \sigma_{eg} \\ \sigma_{eg}^* & \sigma^2_g \end{array} \right] \right)
\end{align*}
\]  

(5)

where \(E_{ij}\) and \(C_{ij}\) represent, respectively, the observed effect and cost of the \(i\)th patient receiving treatment \(t(=0,1)\) in country \(j\). Note that model (5) is the same as model (1) except that we have now introduced a subscript to indicate that we are interested in both the country \(j\) estimates of the mean costs and effects in each arm, and their differential mean estimates.

It is assumed for simplicity that the covariance matrix is not country specific (i.e., in each arm, the variance of costs and effects between subjects is the same in each country, as is their covariance).\(^{8}\)

---

e. Multivariate normality has the following property.\(^{26}\) Let \(X\) and \(Y\) follow a bivariate normal (BVN) distribution, \(X, Y \sim \text{BVN}[\theta_x, \theta_y, \sigma_x, \sigma_y, \rho_x, \rho_y]\), where the distribution has the following properties: \(E[X] = \theta_x, E[Y] = \theta_y, \text{Var}[X] = \sigma_x^2, \text{Var}[Y] = \sigma_y^2, \text{Cov}[X, Y] = \rho_x \sigma_x \sigma_y\). It can be shown that the conditional distribution of \(Y|x \sim N[E[Y|x], V(Y|x)]\), where \(E[Y|x] = \theta_y + \frac{\sigma_y}{\sigma_x} (x - \theta_x)\) and \(V[Y|x] = \sigma_y^2 (1 - \rho^2)\).

f. This is an important consideration, even in trials where individual patients are the unit of randomization.\(^{56-58}\)

g. This assumption can be relaxed to incorporate between-country differences in the covariance matrix, but its relevance is likely to depend on the data.
This is another important feature of the modeling framework proposed here compared to the approaches described earlier.

As before, model (5) can be reexpressed using the product normal formulation:

\[ E_{ij} \sim N(\mu_{ij}, t_{ij}^2) \quad C_{ij} \sim N(\mu_{ij}, t_{ij}^2) \quad \mu_{ij} = \gamma_j + \Delta_{ij} \]

where \( \gamma_j \) and \( \Delta_{ij} \) are treated as fixed effects, whereas the \( j \)th country-specific differential mean effects (\( \Delta_{ij} \)) and costs (\( \Delta_{ij} \)) are assumed to follow a BVN distribution and to be correlated with each other, as showed in equation (7):

\[ \Lambda_{ij} \sim N(\delta_c, \tau_{ij}^2) \quad \Delta_{ij} \sim N(\delta_e, \tau_{ij}^2) \]

where \( \tau_{ij}^2 \) and \( \tau_{ij}^2 \) represent the between-country variance in the (country-specific) difference in mean costs and effects, \( \delta_c \) and \( \delta_e \) are the overall differences in mean costs and effects, and \( \Lambda_{ij} \) is the departure of the country-specific mean differential effects from the trial overall mean.

This formulation differs from the model proposed by Manca and others,\(^8\) in that the random effects in equation (7) are specified only for the differential mean costs and effects. In this sense, this model resembles a Bayesian (random effects) IPD meta-analysis.\(^26\)

The parameters (\( \Lambda_{ij} \)) and (\( \Delta_{ij} \)) can be combined to calculate the country-specific incremental net benefit as shown in equation (8).

\[ \text{INMB}_{ij} = (\Delta_{ij} \cdot \lambda - \Delta_{ij}). \]

For the \( j \)th country, the expressions in equation (6) can be back-transformed to the parameters in model (5), as before:

\[ \mu_{ij} = \gamma_j + \Delta_{ij} \]

The choice of a truly noninformative prior for variance parameters of the random effects in BHLM is an area of ongoing methodological research.\(^25,26,29\) Following Lambert and others,\(^30\) a range of several noninformative prior distributions for the variances (\( \tau_{ij}^2 \) and \( \tau_{ij}^2 \)) of the “random effects” has been investigated here, and uniform priors on the standard deviation scale have been selected in the final model. This choice was consistent with recent recommendations.\(^25,26\) For the other parameters in equation (9), standard noninformative prior distributions were used.

**Patient and Country Covariates**

An important feature of the framework presented in equations (6) to (9) is its ability to accommodate explanatory variables at patient and country levels. There are several reasons for wanting to do so in RCT-based statistical CEA modeling.

**Subject-Level Covariates**

The selection of patient-level covariates in clinical trials is usually a choice dictated by clinical considerations. There are, however, 3 main justifications for including subject-level covariates in the cost-effectiveness modeling of IPD. The first reason is to adjust for baseline imbalance of important prognostic factors between groups.\(^31\) This improves the efficiency of the estimation process and ensures that “treatment effects” estimates are unbiased. The second reason is to obtain robust estimates of country-specific outcome measures. If there is between-country variation in a given covariate—say, age—considered to be an important predictor of cost (effect) differences, then although age may be balanced between treatment groups within each country, part of the between-country variation in the cost–effectiveness results could be ascribed to by-country variation in mean age. A final argument for including subject-level covariates is to look at effect modification.\(^31\) If the researcher believes that differential costs and effects may differ between subgroups of patients (e.g., by gender), then it is important to look at treatment-covariate interactions. For simplicity, this article does not explore treatment-covariate interaction at the patient level, but the models presented here could be easily extended to this purpose. For an application of these models to the case of patient and center subgroup analysis, see the recent work by Nixon and Thompson.\(^32\)

476 • MEDICAL DECISION MAKING/JUL–AUG 2007
Country/Jurisdiction-Level Covariates

Country-specific variables could also be important predictors of between-country difference in costs and effects, and it is reasonable to expect that their role and selection will vary depending on the context of the study. There are 2 main reasons for including country-level covariates in the type of models proposed here. The first is to investigate if certain “contextual factors” can help explain the observed between-country variability in the results. Between-country difference in clinical treatment protocols for the same condition, for instance, may explain between-country difference in observed resource use and health outcomes. The same can be said for differences in the country-specific level of public expenditure in health care. Once these factors have been factored in, they can improve the model’s predictions for those countries sharing the same mix of contextual factors. A note of caution is due here. The researcher must be careful when selecting these covariates as there are usually few countries in a typical multinational study, and statements about transferability of the results to countries outside the set of those participating in the trial require careful consideration. The second reason for including country-level covariates is to look at covariate–treatment and cross-level–covariates interaction. The former should usually be used for exploratory work, as there are typically very few observations at the country level. The cross-level interaction can be used to explore between-country variation of the treatment effect on costs and health outcome for specific patient subgroups. This type of analysis would also be useful in single-country multicenter trials to explore whether the treatment effect for the more severe patients, for example, is better in teaching hospitals than district general hospitals. Unfortunately, the literature on methods used to identify country-level characteristics that are expected to affect costs and effects in CEA is sparse. Grieve and others explored the role of national gross domestic product (GDP) in explaining between-country differences in differential cost estimates in an observational study of stroke patients. In a recent study, Drummond and others suggested 3 types of variables that could characterize a country in the context of CEA: the percentage of national GDP spent on health care, the reimbursement system for hospitals in a given health care system, and the payment method used for physicians.

The second extension of the analytical framework proposed in this article is aimed at showing how to incorporate subject- and country-specific covariates in the modeling strategy proposed so far. The model described in equations (6) to (9) can therefore be extended as follows:

\[ E_{ij} \sim N(\mu_{ij}, \tau_{e}^2) \]
\[ C_{ij} \sim N(\mu_{ij}, \tau_{c}^2) \]
\[ \mu_{ij} = \eta_1 + \gamma_1 \text{x}_{ij} + \phi_c \tilde{x}_{ij} + \Delta_{ij} \]
\[ \mu_{cij} = \eta_2 + \gamma_2 \text{x}_{cij} + \phi_c \tilde{x}_{cij} + \Delta_{cij} \]

where \( \text{x}_{ij} \) and \( \tilde{x}_{ij} \) are patient-specific variables (e.g., age, gender, smoking status, etc.). These variables are centered with respect to their trial-wide mean to improve the efficiency of the estimation procedure. Country-level covariates can be included in the expression for the difference in costs and effects. The distinction here is important in that this incorporates an effect modifier; that is, the treatment effects vary according to the level of the covariate(s). Therefore, the random effects on the difference in mean costs (\( \delta_c \)) and survival (\( \delta_s \)) between countries in equation (7) can be modified as follows to include country-specific information:

\[ \Delta_{ij} \sim N(\delta_j, \tau_{c}^2) \]
\[ \Delta_{cij} \sim N(\delta_c, \tau_{c}^2) \]

where \( \tilde{z}_{ij} \) and \( \tilde{z}_{cij} \) are country-level covariates, and \( \eta_s \) are the coefficients associated with the country-level regressions. The country-level covariates are centered with respect to the trial overall mean and are not weighted by the number of subjects in each center. Notice that the country-specific variable used in the cost equation (\( \tilde{z}_{ij} \)) can be different from that used in the effects equation (\( \tilde{z}_{cij} \)). Noninformative priors are given to \( \phi, \gamma, \alpha, \eta, \) and \( \theta \), assuming normal distribution with mean zero and large variances, whereas a uniform prior on the standard deviation scale was chosen for parameters \( \tau \) in equation (11) as in equation (7).

Exploring the Variability of the Results between Countries

The flexibility of BHM for CEA using IPD facilitates the estimation of country-specific measures of cost-effectiveness in a way that overcomes the limitations of a stratified analysis, where each country is analyzed independently from the others. This is
achieved through shrinkage estimation, which implements the idea that although different, country-specific data might share some degree of similarity; as such, country-specific estimates of treatment effect (on costs and effects) can borrow strength from each other. In other words, country-specific estimates will be more or less pulled (i.e., shrunken) toward the trial-wide treatment effect, depending on the number of observations in each country and the sampling variability around each country-specific estimate. The amount of shrinkage will also depend on the size of the estimated between-country variance.

The BHLMs proposed in this article are estimated via Bayesian MCMC methods, using the freely available software package WinBUGS, and the code for the models described earlier is reported in the appendix. A Bayesian approach allows inclusion of preexisting evidence in the form of prior information, and the use of MCMC has the benefit of producing the output in a convenient format for calculating cost-effectiveness acceptability curves (CEACs). These curves represent the probability that the intervention is cost effective for a given level of decision maker maximum willingness to pay for an additional unit of outcome, given the available information. In the models presented in this article, the probability that the intervention is cost effective in country \( j \) is simply the probability that the \( \text{INMB}_j \) in equation (8) is greater than zero (at a given \( \lambda \) value), which can be obtained directly from the posterior distribution of the \( \text{INMB}_j \). For an extensive discussion of the reasons for adopting a Bayesian perspective in CEA, see Luce and O’Hagan.

**MOTIVATING EXAMPLE: THE ATLAS TRIAL**

The ATLAS study is a multinational trial that enrolled 3164 patients in 19 countries, comparing a low dose \( (n = 1596) \) and a high dose \( (n = 1568) \) of the ACE inhibitor lisinopril in patients with chronic heart failure. Details of the main economic and clinical analyses have been reported elsewhere. The case study makes use of a total of 3061 observations (low dose, \( n = 1545 \); high dose, \( n = 1516 \)) from 17 countries; 2 countries were excluded because they recruited a very low number of patients and/or had an extremely unequal allocation of patients between the 2 arms. Furthermore, from the end of year 3 on, patients were subject to administrative censoring. Currently, there are no methods in the literature that illustrate how to deal with censored cost-effectiveness data in a bivariate hierarchical modeling framework. Therefore, for simplicity, the analysis uses data observed during the first 3 years of the study. Due to these assumptions, the specific results of the analyses presented here are not the same as in the main study report and should not be interpreted as definitive.

**Trial-Wide Results**

The starting point is to quantify the trial-wide cost-effectiveness results. Table 1 reports the trial-wide results for the 3 models discussed earlier—that is, the nonhierarchical bivariate model (column 1) estimated by implementing the methodology proposed by O’Hagan and others, the BHLM with no covariates (column 2), and the BHLM with the addition of subject- and country-level covariates (column 3).

As expected, the presence of a hierarchical structure in the data means that the differential costs and survival gain estimates obtained using the nonhierarchical model have excessively narrow credibility intervals (CrIs). The 95% CrIs estimated using the BHLM for the differential cost are approximately 5 times wider. Similarly, the 95% CrIs for the trial-wide survival gain are 3.5 times wider than those obtained using the nonhierarchical approach. These results are in line with those found by Manca and others. That is, analytical approaches that ignore the hierarchical structure in multinational trial-based cost-effectiveness data can generate inaccurate estimates of the differential mean costs, effects, and their credibility intervals.

In the BHLM with covariates, the patient-level cost and survival equations (equation (10)) use the patient’s age as a covariate, whereas country-level cost and survival equations (equation (11)) employ, respectively, the country’s mean life expectancy at birth and the public expenditure in health care, where the latter is expressed as a percentage of the national GDP. It should be noted that model (10) assumes that patient-level covariates affect the differential costs and survival estimates indirectly, through their impact on the mean cost and survival in each arm of the trial. In model (10), the effect of mean age on the mean cost \( (\phi_a) \) and survival \( (\phi_e) \) in the trial is, respectively, 4.5 (95% CrI: 2.4 to 6.5) and −4.6 (95% CrI: −5.8 to −3.3), suggesting that the mean cost and survival estimates in each country are clearly affected by the patient’s age and that, on average, a patient who is 1 year older than the overall trial mean would be expected to cost £4.5 more and to live 4.6 days less than the average patient in the study (results not reported in the table).
Similarly, country-specific variables seem to affect the estimated differential mean cost and survival in the trial. Mean life expectancy at birth has a positive impact on the differential cost, indicating that those countries with mean life expectancy greater than the trial average will experience an increased per patient cost (for those patients treated with high-dose lisinopril) of $\frac{C}{131} = 11$ (95% CrI: $-74.8$ to $98.7$). On the survival side of the cost-effectiveness equation, the results reported in the last column of Table 1 suggest that those countries with higher (than the overall mean) public expenditure in health care as a percentage of national GDP can expect, on average, a survival gain of 52.6 days (95% CrI: $-18.9$ to $124.7$) in patients treated with high-dose lisinopril.

Finally, it should be noted that the inclusion of patient- and country-level covariates in the BHLM model allows us to explain the between-country variability in the estimated difference in survival. The estimate of $\frac{C^2}{13}$, in fact, is reduced (by more than half) from 31 in the BHLM with no covariates to 12 when adding explanatory variables. The opposite occurs in the cost difference equation, where the addition of mean life expectancy actually increases the between-country variation of the country-specific estimated differential cost. We will come back to this result in the next section.

### Country-Specific Cost-Effectiveness Estimates

As mentioned earlier, an additional advantage of using BHLM is that this methodology facilitates the quantification of country-specific cost-effectiveness estimates, through shrinkage estimation. To illustrate this point, Figures 1 and 2 provide a graphical representation of the country-specific differential cost and survival gain estimates (together with their credibility intervals) obtained under the 3 models discussed above. The marker and the horizontal bars represent, respectively, the posterior mean estimate and 95% CrI of differential mean costs (Figure 1) and survival (Figure 2) obtained from MCMC estimation, running 3 chains for 40,000 iterations following a burn-in period of 20,000 iterations. Convergence was checked using the Gelman-Rubin convergence criteria, as implemented in WinBUGS.

#### Nonhierarchical Bivariate Model

A number of interesting considerations can be made when comparing country-specific versus trial-wide CEA results obtained under a nonhierarchical bivariate model, the most obvious relating to the degree of within-country variability. Larger sampling uncertainty in country-specific estimated differences in cost and survival gain is observed in those countries that recruited a smaller number of patients into the trial (e.g., countries 2, 7, 11, and 14). Notice, though, that the sample size itself is not always a predictor of the degree of sampling uncertainty around the estimated mean differences. The country-specific estimated difference in costs for country 13, for instance (where $n = 49$), has a narrower CrI than the estimate for country 12 (where $n = 70$). The same consideration can be made with respect to the sampling uncertainty surrounding the differential survival estimates between countries 13 and 10.

### METHODOLOGY

#### Table 1  Trial-Wide CEA Results in the ATLAS Trial

<table>
<thead>
<tr>
<th>Model</th>
<th>Differential cost (UK pounds): $\Delta c$</th>
<th>Mean life expectancy: $\Delta \hat{c}$</th>
<th>Survival gain (days): $\Delta e$</th>
<th>Public expenditure in health: $\Delta \hat{e}$</th>
<th>Between-country variance in differential cost: $\tau_c^2$</th>
<th>Between-country variance in survival gain: $\tau_e^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian Bivariate Nonhierarchical Model</td>
<td>8.2 ($-32.9$ to $50.1$)</td>
<td>—</td>
<td>24.9 ($0.2$ to $51.2$)</td>
<td>11.7 ($-74.8$ to $98.7$)</td>
<td>—</td>
<td>61 (42 to 117)</td>
</tr>
<tr>
<td>BHLM with No Covariates</td>
<td>11.9 ($-192.8$ to $217.1$)</td>
<td>—</td>
<td>40.2 ($-50.9$ to $132.8$)</td>
<td>35.2 ($-64.8$ to $136.6$)</td>
<td>61 (42 to 117)</td>
<td>31 (18 to 73)</td>
</tr>
<tr>
<td>BHLM with Covariates</td>
<td>10.3 ($-200.4$ to $222.1$)</td>
<td>—</td>
<td>52.6 ($-18.9$ to $124.7$)</td>
<td>52.6 ($-18.9$ to $124.7$)</td>
<td>71 (42 to 155)</td>
<td>12 (7 to 31)</td>
</tr>
</tbody>
</table>

Results are reported for 3 different models: coefficients and 95% credibility intervals (in parentheses) $N = 3061$. Markov chain Monte Carlo estimates are obtained after 40,000 iterations and a 20,000-iteration burn-in period. CEA, cost-effectiveness analysis; ATLAS, Assessment of Treatment with Lisinopril and Survival; BHLM, Bayesian bivariate hierarchical models.

a. Expressed as departure from the overall mean.
b. As percentage of the gross domestic product (GDP).
Figure 1  Pooled and country-specific cost differences obtained using alternative modeling strategies. Markers indicate country-specific mean differential cost estimates, and horizontal bars across the markers represent 95% credibility intervals. The covariates used in the cost difference equation are patient’s age and country-specific mean life expectancy, both expressed as departure from the overall mean.
Figure 2  Pooled and country-specific survival gains obtained using alternative modeling strategies. Markers indicate country-specific mean differential cost estimates, and horizontal bars across the markers represent 95% credibility intervals. The covariates used in the survival equation were patient's age, expressed as departure from the overall mean, and public expenditure in health, expressed as percentage of the gross domestic product.

METHODOLOGY

© 2007 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.
This phenomenon may be due to several factors. One is the degree of correlation between observations within the same country, where the higher the correlation, the narrower the CrI and the variability of the observed cost (and survival) seen in the different countries. Another explanation may be that countries with large resource use have higher mean costs and larger spread of costs.

More interestingly for the purpose of this article is the between-country variability of the CEA results, which leads to potentially different country-specific recommendations regarding the value for money of the “intervention” (i.e., high dose). For example, high dose (strictly) dominates low dose in countries 4, 5, and 12 but is dominated by low dose in country 14. The question for the policy makers in the latter country is, therefore, whether to consider the trial-wide results as the most valid results in their own context or, instead, to rely only on the results obtained analyzing data collected in their own jurisdictions.

Bivariate Hierarchical Model
Results without Covariates

A comparison of the above results with those obtained from a BHLM with no covariates suggests that, in most cases, thanks to the effect of the shrinkage estimation, jurisdiction-specific differential cost and survival are much closer to their trial-wide counterparts. Countries 2, 4, and 12, for example, display particularly large point estimates with correspondingly wide credibility intervals using the nonhierarchical model. By explicitly modeling the hierarchical structure in the data, the BHLM exploits similarities in the subject-level data between countries (i.e., they borrow strength), hence obtaining efficiency gains. A direct result is that the mean country-specific estimates of differential costs and survival gains are shrunken toward the trial-wide mean. Furthermore, country-specific estimates are more precise than those obtained using a nonhierarchical modeling, as indicated by their 95% CrIs.

Bivariate Hierarchical Model CEA Results,
with Subject- and Country-Level Covariates

There might be important factors that, if considered, could contribute to partially explain the observed between-country variability in cost-effectiveness. The argument in favor of their inclusion in the CEA of multinational IPD is that, if left unaccounted for, these factors could inflate the between-country unobserved variability in the results, leading to potentially erroneous conclusions concerning the transferability of the study findings. In this sense, it is reasonable to expect that a more realistic country-specific estimate will lie between the 2 extreme results obtained with the (stratified) nonhierarchical bivariate model and the BHLM with no covariates.

The application of the latter assumes perfect exchangeability between the information conveyed by the 17 countries included in the analysis. However, it could be argued that assuming perfect exchangeability may on some occasions be unreasonable. Additional information available at country and patient levels could be used to reassess the between-country variability in the cost-effectiveness results. Table 2 reports the mean values of a selection of patient- and country-specific variables that might be important to consider in the ATLAS trial. These are variables that could be expected, a priori, to affect the country-specific mean differential mean estimates of cost and survival gain, their sampling uncertainty, and their between-country variability. Unfortunately, the country-specific variables reported in this table were not collected as part of the ATLAS trial but have been obtained ex post from published Organization for Economic Cooperation and Development (OECD) health data.

An examination of the country-level summary statistics for the patient in the trial and of the country-specific variables from OECD indicates some degree of between-country variability in these covariates. Mean age at randomization, for instance, ranges from a minimum of 56.76 in country 10 to a maximum of 67.92 in country 15. Similarly, the mean left ventricular ejection function (LVEF) varies from 0.3654 in country 13 to 0.6081 in country 5.

Intercountry variability can also be detected when looking at variables such as life expectancy (LE) in the total population at birth. This ranges from a minimum of 72.4 years in country 10 to a maximum of 79.7 in country 1. Similarly, the total private (public) expenditure on health care as a percentage of the national GDP varies respectively between 0.6% in country 5 (4.5% in country 11) and 7.6% in country 17 (7.1% in country 15). Finally, the per capita level of alcohol and tobacco consumption in the population older than 15 years of age displays a remarkable between-country variation, with alcohol consumption ranging from 5.5 units in country 12 to a maximum of 14.5 in country 11 and
tobacco consumption varying from 992 grams per capita in country 7 to 2670 in country 14.

In the attempt to account for some of these between-country differences in factors that could help explain the between-country variability in cost-effectiveness results, we ran the BHLM with the addition of patient- and country-specific covariates. Results have been already reported in Table 1 (with regard to the trial-wide estimates of differential costs and survival) and Figures 1 and 2 (with regard to the country-specific shrunken estimates). For simplicity of exposition, this model used patient-specific age at randomization (for both costs and survival) as the explanatory variable in the patient-level cost and survival equations, public expenditure on health care as the percentage of the national GDP in the country-specific differential survival equation, and mean LE of the total population at birth in the country-specific differential cost equation. The trial-wide results have been discussed earlier and indicate that age has a positive impact on mean costs and survival in both arms. Furthermore, the greater the life expectancy in a specific country compared to the overall trial mean, the greater the incremental cost. On the other hand, the higher the public investment in health care with respect to the overall mean in the countries participating in the ATLAS trial, the higher the expected survival benefit that one can expect. It is important to notice that the inclusion of subject- and country-specific variables helps to reduce the between-country variability in the differential survival. This does not occur for the cost side. A closer look at the between-country distribution of the variables used for the cost equation at the patient (i.e., age) and country levels (i.e., LE) may help to shed light on this apparently contradicting result. Patients recruited in the trial from country 10 are in fact the oldest in the trial sample as well as having (on average) the lowest life expectancy at birth, and these factors will set country 10 somewhat apart from the others when estimating country-specific differential costs conditional on these variables, therefore increasing the between-country variation in the estimated mean difference in cost.

---

**Table 2** Patient- and Country-Specific Baseline Mean Values for a Sample of Covariates in the Assessment of Treatment with Lisinopril and Survival (ATLAS) Trial

<table>
<thead>
<tr>
<th>Center</th>
<th>( n_i )</th>
<th>Age</th>
<th>LVEF(^b)</th>
<th>% Male</th>
<th>Life Expectancy(^c)</th>
<th>Private Expenditure in Health Care as % of the GDP(^d)</th>
<th>Public Expenditure in Health Care as % of the GDP(^e)</th>
<th>Alcohol Consumption(^f)</th>
<th>Tobacco Consumption(^g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88</td>
<td>65.53</td>
<td>0.495</td>
<td>81.32</td>
<td>79.7</td>
<td>2.4</td>
<td>5.7</td>
<td>9.8</td>
<td>1265</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>57.24</td>
<td>0.52</td>
<td>84</td>
<td>78.8</td>
<td>2.5</td>
<td>5.4</td>
<td>11.3</td>
<td>2210</td>
</tr>
<tr>
<td>3</td>
<td>102</td>
<td>64.34</td>
<td>0.579</td>
<td>71.03</td>
<td>77.7</td>
<td>0.5</td>
<td>6.2</td>
<td>10.2</td>
<td>2310</td>
</tr>
<tr>
<td>4</td>
<td>281</td>
<td>63.84</td>
<td>0.454</td>
<td>79.58</td>
<td>79.4</td>
<td>2.8</td>
<td>6.5</td>
<td>7.7</td>
<td>1432</td>
</tr>
<tr>
<td>5</td>
<td>142</td>
<td>61.15</td>
<td>0.608</td>
<td>81.76</td>
<td>75.3</td>
<td>0.6</td>
<td>6.4</td>
<td>11.8</td>
<td>2500</td>
</tr>
<tr>
<td>6</td>
<td>97</td>
<td>62.32</td>
<td>0.56</td>
<td>80</td>
<td>76.7</td>
<td>1.5</td>
<td>6.8</td>
<td>11.4</td>
<td>1527</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>57.44</td>
<td>0.519</td>
<td>78.85</td>
<td>78.1</td>
<td>1.7</td>
<td>5</td>
<td>9.9</td>
<td>992</td>
</tr>
<tr>
<td>8</td>
<td>119</td>
<td>63.88</td>
<td>0.492</td>
<td>81.45</td>
<td>79.3</td>
<td>2.3</td>
<td>7</td>
<td>10.5</td>
<td>2037</td>
</tr>
<tr>
<td>9</td>
<td>149</td>
<td>67.11</td>
<td>0.503</td>
<td>76.82</td>
<td>78.2</td>
<td>2.9</td>
<td>5.7</td>
<td>10</td>
<td>2319</td>
</tr>
<tr>
<td>10</td>
<td>73</td>
<td>56.76</td>
<td>0.568</td>
<td>82.43</td>
<td>72.4</td>
<td>1.7</td>
<td>4.8</td>
<td>12.3</td>
<td>1809</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
<td>63.37</td>
<td>0.472</td>
<td>88.89</td>
<td>76.2</td>
<td>1.3</td>
<td>4.5</td>
<td>14.5</td>
<td>1834</td>
</tr>
<tr>
<td>12</td>
<td>72</td>
<td>64.39</td>
<td>0.553</td>
<td>82.89</td>
<td>78.7</td>
<td>1.2</td>
<td>6.5</td>
<td>5.5</td>
<td>1433</td>
</tr>
<tr>
<td>13</td>
<td>49</td>
<td>60.44</td>
<td>0.365</td>
<td>90.38</td>
<td>76.9</td>
<td>2.8</td>
<td>6.1</td>
<td>13</td>
<td>2010</td>
</tr>
<tr>
<td>14</td>
<td>70</td>
<td>59.84</td>
<td>0.486</td>
<td>76.39</td>
<td>79.3</td>
<td>2.1</td>
<td>5.2</td>
<td>11.7</td>
<td>2670</td>
</tr>
<tr>
<td>15</td>
<td>227</td>
<td>67.92</td>
<td>0.526</td>
<td>83.19</td>
<td>79.8</td>
<td>1.2</td>
<td>7.1</td>
<td>6.5</td>
<td>1710</td>
</tr>
<tr>
<td>16</td>
<td>284</td>
<td>63.77</td>
<td>0.441</td>
<td>84.83</td>
<td>78.1</td>
<td>1.2</td>
<td>6</td>
<td>10.5</td>
<td>1225</td>
</tr>
<tr>
<td>17</td>
<td>1164</td>
<td>63.6</td>
<td>0.516</td>
<td>76.58</td>
<td>76.8</td>
<td>7.6</td>
<td>6.1</td>
<td>8.4</td>
<td>1542</td>
</tr>
</tbody>
</table>

---

a. Source: Organization for Economic Cooperation and Development (OECD).\(^{39}\)
b. Left ventricular ejection function.
c. Life expectancy total population at birth (years).
d. Total private expenditure on health (% gross domestic product [GDP]).
e. Total public expenditure on health (% GDP).
f. Alcohol consumption in liters per capita (15 +).
g. Tobacco consumption in grams per capita (15 +).
Figures 1 and 2 show the impact of including patient- and country-level covariates in the model by plotting the country-specific treatment effects on costs and survival. The level of shrinkage, now conditional on the covariates included in the model, is more attenuated for some jurisdictions, thereby capturing potentially important differences between countries. Results for countries 7 and 14, for example, are now different from those obtained using the BHLM without covariates, suggesting that low-dose lisinopril dominates high-dose lisinopril. Similarly, results suggest that in countries 9 and 10, high dose no longer dominates low dose, whereas in some other countries (e.g., 4 and 5), the conclusions remain the same regardless of the model used. As in the case of the BHLM with no covariates, the impact of shrinkage is both qualitative and quantitative and varies across countries in the trial.

Figure 3 plots the country-specific cost-effectiveness planes for a selection of countries in the trial. These graphs show the country-specific confidence ellipses obtained from 1) the nonhierarchical bivariate model, 2) the BHLM with covariates, and contrasting these against the 3) trial-wide confidence ellipse derived from the BHLM with covariate results. For some countries, the application of a BHLM in combination with patient and country covariates affects both the joint sampling uncertainty and the point estimates of differential mean costs and survival, leading in some cases to opposite recommendations regarding the cost-effectiveness of high-dose lisinopril. Country 2, for instance, is shrunken toward the overall mean obtained from the BHLM with covariates. Similar behavior is observed in countries 3 and 9. On the other hand, countries 10 and 14 show that the addition of covariates to their specific cost-effectiveness estimates slightly increases the joint variability of ($\Delta_c$, $\Delta_e$), for the reasons discussed earlier in this section. Figure 4 plots the CEACs for the same set of countries inspected in Figure 3.

A brief inspection of the CEACs confirms that the use of the trial-wide result does hide an important piece of information for location-specific policy makers. In fact, not only is there clearly variation in the cost-effectiveness by country, but there is also considerable variation in the decision uncertainty at a given $\lambda$ level, represented by the probability that the intervention (i.e., high dose) is cost effective. These results suggest that it is impossible to predict the impact on the results deriving from the explicit consideration of patient- and country-specific covariates. In country 2, for instance, where the sample size was very small ($n = 24$), the probability of high dose being cost effective is still considerably low despite the country-specific results being heavily shrunken toward the trial-wide BHLM estimate. In other words, although the treatment effects are shrunken toward the trial overall mean in country 2, the probability that the intervention is cost effective is still penalized.

**DISCUSSION**

This article proposes a novel approach to the analysis of individual- patient-level cost-effectiveness data collected in multinational RCTs by extending the use of Bayesian bivariate hierarchical modeling for health care economic evaluation purposes.

Two recent contributions\(^5,32\) have proposed methods somewhat similar to those advocated in this article. Nixon and Thompson\(^32\) have used Bayesian modeling for cost-effectiveness IPD to illustrate ways in which alternative research questions could be investigated (i.e., subgroup CEA at the patient and center levels) while relaxing basic assumptions often made in regression analysis (i.e., common variance between treatment arms for costs and effects, alternative parametric assumptions for costs and effects). There are 3 main differences between the methods proposed in this article and those advocated by Nixon and Thompson. First, we consider the inclusion of explanatory variables at both patient and country levels. Second, our example includes a much larger number of higher level units (i.e., 17) compared to the Nixon and Thompson example, which was based on 5 units only. We feel that this is an important distinction as a greater confidence can be placed on our country-specific random effects being drawn from an appropriate distribution; this also allows us to correlate the country-specific effects across the 2 equations. This leads us to the third difference between the 2 studies—that is, unlike Nixon and Thompson, who only consider a bivariate model by linking individual-level random terms across the equations (due to the small number of higher level units in their case study), we consider a bivariate model where both the country- and the individual-level terms are each correlated across equations.

Pinto and others\(^5\) recently explored alternative shrinkage estimation methods, presenting a simulation exercise based on summary country-level cost–effectiveness estimates from large multinational RCTs. The authors explore the impact of univariate versus bivariate shrinkage, for costs and...
effects, as well as exploring the impact of different distributions for the random effects at the country level. We take a different approach by using individual-patient-level cost-effectiveness data. In addition, although Pinto and others obtained shrinkage estimates through an empirical Bayes estimation procedure, we chose to implement a fully Bayesian analysis. The advantages of our approach

**Legend**

- □ = Country-specific mean values;  ———— 95% confidence ellipse (bivariate non-hierarchical model);
- • = Trial-wide mean values;  ———— 95% confidence ellipse (BHLM with covariates);
- △ = Country-specific mean values;  ———— 95% confidence ellipse (BHLM with covariates).

**Figure 3** Shrinkage on the cost-effectiveness plane. The y-axis and x-axis represent, respectively, the estimated difference in mean survival and difference in mean costs. The markers represent the country-specific mean estimate of the point with coordinates \((\Delta_c, \Delta_s)\), together with the relative 95% confidence ellipse.
Figure 4 Country-specific cost-effectiveness acceptability curves obtained using alternative modeling strategies for selected countries.

Legend:
- Country-specific estimate (BHLM with covariates)
- Trial-wide – BHLM with covariates
- Country-specific estimate bivariate non-hierarchical model
THE USE OF BIVARIATE HIERARCHICAL MODELING

compared to empirical Bayes are outlined above. A final difference relates to the fact that, unlike the methods proposed here, the methods proposed by Pinto and others did not include covariates at patient and country levels. Although the inclusion of these variables is still possible in their model, such an analysis based on summary data only would be subject to a number of issues (e.g., aggregation bias, patient subgroup analysis), which have been well rehearsed in the meta regression literature.

It could be argued that knowing “which country is which” will convey some information about the costs in a given country, hence violating the exchangeability assumption. Although this may be true for absolute costs in each treatment arm, knowing that one country has higher costs than another may not tell us much about the differential costs (and indeed the differential health outcomes). The approach proposed here is standard methodology in many Bayesian random effects meta-regression models, where the knowledge of the characteristics of the trials that enter the model does not prevent the analyst from making a prior assumption of exchangeability. Gelman and others argue that “in virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ, . . . The fact that the experiments differ implies that [the parameter of interest in the unit] differ, but it might be perfectly acceptable to consider them as drawn from a common distribution.” The authors continue by saying, “Objecting to exchangeability for modelling ignorance is no more reasonable than objecting to an iid model for samples from a common population, objecting to regression models in general, or, for that matter, objecting to displaying points in a scatter plot without individual labels. As with regression, the valid concern is not about exchangeability, but encoding relevant knowledge as explanatory variables where possible.” In essence, “the usual way to model exchangeability with covariates is through conditional independence. . . . In this way exchangeable models become almost universally applicable, because any information to distinguish different units should be encoded.”

The models presented and discussed in this article assume joint bivariate normality of costs and effects at patient and country levels. Although the “Normal” case is a very general framework in multivariate statistical modeling, it can be argued that especially for CEA data, one may need to use alternative parametric distributions that describe the data more appropriately. In this sense, the product normal formulation facilitates a more flexible modeling process that allows the use of different distributions for costs (i.e., gamma) and effects (i.e., binomial, Poisson). However, Bayesian modeling outside the bivariate normality framework is highly complex. Future work will explore the impact of alternative distributional assumptions.

A further issue concerns the selection of country-specific variables for the economic analysis. Unfortunately, when country-level data are not collected as part of the study, as is the case in the vast majority of multinational and multicenter RCTs, the selection of country (center)-specific variables must rely on some judgment regarding which readily available statistics can be used as proxy to explain between-country variability in costs (and effects). In the present study, the country-specific variables reported in Table 2 were chosen among those suggested to be factors that might affect the cost-effectiveness of a health technology between countries, and it is recognized that their explanatory power may be limited. The issue of which variable to use to adjust for factors that may have an impact on the difference in cost and effects is both a theoretical and an empirical one, and it deserves further research.

Should we be concerned about whether there is statistically significant difference between country-specific estimates? Cook and others have advocated a prior approach, based on a test for homogeneity, in which multinational economic RCT data are analyzed to assess the appropriateness of a “pooled” CEA. The general idea is to follow the methods used to analyze clinical data from such trials, where tests of homogeneity are typically performed before pooling the data. The authors suggest that if there is no evidence of a treatment-by-center interaction, the data can be pooled for analysis across centers/countries, thus offering improved precision of cost-effectiveness ratios for each country. The authors take a strong statistical inferential view, stressing that their interest is “not in whether the countries differ with respect to their average clinical or economic outcome but rather whether there are important differences in their effect of treatment among the countries.” Although this is one possible way to explore the variability of the results between countries, it could be argued that decisions based on the test of interaction could be misled by the fact that this test is often underpowered and that non-statistically significant between-country differences in the treatment effect may still lead to different decisions in each jurisdiction. More important, policy makers are clearly interested in jurisdiction-specific
results, and as argued above, results from pooled analyses might not be as informative as they could be. Finally, as decision makers’ information needs increase requiring cost-effectiveness evidence comparing all available treatment strategies, in many circumstances, cost-effectiveness studies will need to synthesize both IPD and summary data from trials (and other sources) to reflect all available evidence relating to all relevant management options.\textsuperscript{46} In this sense, the framework outlined in this article is highly suitable for evidence synthesis for cost-effectiveness modeling as it can be used to incorporate not only prior information regarding the treatment strategies compared in the trial but also information on the clinical- and cost-effectiveness of alternative treatment strategies in any form this may be available (i.e., summary or IPD).

APPENDIX

model {

for (i in 1:Nsubject0) {
  C0[i] ~ dnorm(mu.C0[i], tau.C0)
  S0[i] ~ dnorm(mu.S0[i], tau.S0)
  mu.C0[i] <- beta.C(country0[i])
    + theta0*(S0[i]-mean(S0[]))
    + gamma[1]*(age0[i]-mean(age0[]))
  mu.S0[i] <- beta.S(country0[i])
    + gamma[2]*(age0[i]-mean(age0[]))
}

for (i in 1:Nsubject1) {
  C1[i] ~ dnorm(mu.C1[i], tau.C1)
  S1[i] ~ dnorm(mu.S1[i], tau.S1)
  mu.C1[i] <- beta.C(country1[i])
    + delta.C[country1[i]]
    + theta1*(S1[i]-mean(S1[]))
    + gamma[1]*(age1[i]-mean(age1[]))
  mu.S1[i] <- beta.S(country1[i])
    + delta.S[country1[i]]
    + gamma[2]*(age1[i]-mean(age1[]))
}

for (j in 1:Ncountry) {
  delta.C[j] ~ dnorm(mu.delta.C[j], tau.dC)
  delta.S[j] ~ dnorm(mu.delta.S[j], tau.dS)
  mu.delta.C[j] <- alpha.C +
    phi*(delta.C[j]-mean(delta.C[])) +
    gamma[3]*(le[j]-mean(le[]))
  Dcost[j] <- -delta.C[j]
  Dsurv[j] <- -delta.S[j]
  beta.C[j] ~ dnorm(0, 0.0001)
  beta.S[j] ~ dnorm(0, 0.0001)
}

for (j in 1:Ncountry) {
  for (i in 1:NK) {
    nb[j,i] <- step(K[i]*(Dsurv[j])-(Dcost[j]))
  }
}

for (k in 1:4) {gamma[k] ~ dnorm(0.0, 1.0E-6)}

theta0 ~ dnorm(0, 0.00001)
theta1 ~ dnorm(0, 0.00001)
alpha.C ~ dnorm(0, 0.0001)
alpha.S ~ dnorm(0, 0.0001)
phi ~ dnorm(0, 0.0001)

ss.C0 <- -s.C0*s.C0
s.C0 ~ dnorm(0, 0.0001)
s.C1 ~ dnorm(0, 0.0001)
ss.C1 <- -s.C0*s.C1
s.C1 ~ dnorm(0, 0.0001)

ss.S0 <- -s.S0*s.S0
s.S0 ~ dnorm(0, 0.0001)
ss.S1 <- -s.S0*s.S1
s.S1 ~ dnorm(0, 0.0001)

ls.C0 ~ dunif(-1.6,2)
ls.C1 ~ dunif(-1.6,2)
ls.S0 ~ dunif(-2,1)
ls.S1 ~ dunif(-2,1)

Dcost[j] ~ dnorm(0, 0.0001)
Dsurv[j] ~ dnorm(0, 0.0001)
Dcost[j] ~ dnorm(0, 0.0001)
Dsurv[j] ~ dnorm(0, 0.0001)
REFERENCES


METHODOLOGY

s.dS < -exp(ls.dS)
ls.dS~dunif(-2,4)

{ }