"Lost in translation": accounting for between-country differences in the analysis of multinational cost-effectiveness data§

Andrea Manca[1][*] and **Andrew R. Willan**[2],[3]

1 *Centre for Health Economics, University of York, UK*

2 *Population Health Sciences, Sick Kids Research Institute, Toronto, Ontario, Canada*

3 *Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada*

Abstract

Background—Cost-effectiveness analysis has gained status over the last 15 years as an important tool for assisting resource allocation decisions in a budged-limited environment such as healthcare. Randomised (multicentre) multinational controlled trials are often the main vehicle to collect primary patient-level information on resource use, cost and clinical effectiveness associated with alternative treatment strategies. However, trial-wide cost-effectiveness results may not be directly applicable to any one of the countries that participate in a multinational trial, requiring some form of additional modelling to customise the results to the country of interest.

Objective—The aim is to produce recommendations regarding methods that can be (currently) considered 'good practice' when exploring the geographical generalisability of cost-effectiveness data. The manuscript proposes an algorithm to assist with the choice of the appropriate analytical strategy when facing the task of adapting the study results from one country to another. The algorithm considers different scenarios characterised by whether or not (a) the country of interest participated in the trial, and (b) individual patient-level data (IPD) from the trial are available.

Methods—Structured review with description and discussion of case studies.

Conclusions—Methods to reflect between-country variability in cost-effectiveness data are available. It is important to be transparent regarding the assumptions made in the analysis and (where possible) assess their impact on the study results.

1 Introduction

Cost-effectiveness analysis (CEA) has gained status over the last 15 years as an important tool for assisting resource allocation decisions in a budged-limited environment such as healthcare. Many national and provincial/state governments nowadays require cost-effectiveness evidence when assessing new healthcare technologies.[1-6] In the case of emerging technologies, where available clinical and economic evidence is still scarce, multinational and multicentre trials are often the main vehicle for collecting primary (patient-level) information on resource use, cost and clinical effectiveness associated with alternative treatment strategies. The multilocation design of these trials offers the benefit of speedy patient recruitment and large sample size, while facilitating reimbursement submissions in several jurisdictions. In fact, by recruiting participants from different countries (and settings), international clinical studies are believed

* Corresponding author at: Centre for Health Economics, Alcuin A, University of York, Heslington, York YO10 5DD, UK. Tel: +44 1904 321430; Fax: +44 1904 321402. *E-mail address*: am126@york.ac.uk (A. Manca).

[7,8] to offer the advantage to generate evidence more likely to be 'generalisable' across locations than that produced by single-centre trials.

In spite of this, it has been argued that - strictly speaking - multinational trial-wide cost-effectiveness results may not be directly applicable to any one of the countries that participate in the clinical study,[9] requiring some form of additional modelling to customise the results to the country[a] of interest.[10] There are various reasons for this. Decision-makers are inherently country-specific and are more interested in results which are directly relevant to their own jurisdiction. Secondly, it is possible that the country of interest did not participate in the clinical trial. Thirdly, even when the country of interest is part of the original study, the presence of country-specific factors potentially affecting the geographical variability of the study results (effectiveness, cost, and quality of life) means that trial-wide results may not be informative for reimbursement decisions at country-level.

Bernie O'Brien was among the first to raise concerns[11] regarding the generalisability of CEA data collected in one country to inform reimbursement decisions in another. His seminal paper suggested that between-country differences in (i) demography and epidemiology of disease, (ii) clinical practice and conventions, (iii) incentives and regulations for healthcare providers, (iv) relative price levels, (v) consumer preferences, and (vi) opportunity costs of resources, could all be potential threats. Health economists have attempted to address these concerns in many different ways since then and, as methodology become more refined, new and alternative approaches are proposed. Three extensive reviews of conceptual and applied research in the area of generalisability of CEA studies have been recently published by Reed *et al,*[12] Sculpher *et al,*[7] and Goeree *et al* [8] and the interested reader is invited to refer to these reports for further details.

The present manuscript concerns itself with recent methodological developments in the analysis of cost-effectiveness data collected alongside multicentre and multinational RCTs. The aim is to produce recommendations regarding methods that can be (currently) considered 'good practice' when exploring the geographical generalisability of cost-effectiveness data. The manuscript proposes an algorithm to assist with the choice of the appropriate analytical strategy when facing the task of adapting the study results from one country to another. The algorithm considers different scenarios characterised by whether or not (a) the country of interest participated in the trial, and (b) individual patient-level data (IPD) from the trial are available. Given the programme of this conference, the manuscript focuses - specifically - on the methods, and provides only a brief discussion of the policy context. For a more detailed discussion of the latter the reader is invited to refer to the Sculpher and Drummond[13] paper at this meeting.

The manuscript is structured as follow. Section 2 reviews the rationale for assessing the generalisability of cost-effectiveness results from country to country, and summarises the current perception as to which data can be directly 'applied' from one country to another and which one need to be country-specific. Section 3 presents the methodology used to produce country-specific estimates of cost-effectiveness distinguishing four main scenarios, characterised by whether or not (i) the country of interest participated in the study, and (ii) IPD from the trial are available. The implications for the design, data collection, analysis and presentation of the study results are considered next. The final section discusses future lines of applied and policy research in this area.

---

[a]The issue is the mismatch between the source of the data and the location of the decision-maker. Thus, it could be argued that we should be referring to the 'jurisdiction of interest', where the term 'jurisdiction' encompasses both, within-country (e.g. regions, provinces) and country-level decision makers.

2 International variation in cost-effectiveness results

The globalisation of clinical research for pharmaceuticals and medical devices in many disease areas (e.g. cardiovascular disease, oncology, respiratory disease, etc), paired with the need of the industry to seek regulatory approval in different jurisdictions, means that multinational trials are often the preferred vehicle for primary (resource use, clinical and quality of life) data collection. Geographical locations in North America, Western Europe and Asia, traditionally chosen as a base from which to recruit study participants have now been joined by countries in Latin America and Eastern Europe. Possible reasons behind this trend relate to the need to recruit even larger study samples, the expanding market for pharmaceuticals (and devices) in these geographical areas, and the less stringent regulatory regimens operating in some countries.

While increasing the potential for the conduct of large studies, the internationalisation of clinical research poses several challenges in terms of design, management, statistical analysis and interpretation of the study results.[14] In recognition of these challenges the International Committee on Harmonisation (IHC) has developed a series of guidelines[15-19] with the objective to facilitate the conduct of international clinical studies with the expectation that the evidence derived from these studies would be used to meet regulatory requirements in different jurisdictions[14]

One of the consequences of the globalisation of clinical research, though, is that researchers are now more aware of the existence of country and regional variations in multinational trials with respect to the resource use and clinical outcome results.[20,21] Between-country variability in clinical effectiveness,[22] resource use and quality of life results,[23] even after adjustment for patient baseline characteristics, is in fact a well known phenomenon in multinational randomised controlled trials (RCTs) investigating the management of patients following myocardial infarction (MI), for instance. Similar findings have been observed in studies considering patients' management after stroke,[24,25] and the management of acute coronary syndrome (ACS). In the latter clinical area, multinational RCTs found important between[26,27] and within country[28] differences in resource use, therapeutic strategies and short-term mortality. The same trend has been observed in many other clinical areas[22,23, 29-32] specifically with respect to variations in average length of stay when an identical treatment was implemented in similar populations in several countries simultaneously.

There are various factors that could affect the generalisability of the results of CEA studies. Sculpher *et al*[7] reviewed the literature to identify these factors and found 36 papers discussing potential sources of variability between locations.

Patient factors

Patient-level variation feeds through to centre or country variations in cost-effectiveness if patients' characteristics (clinical and socio-demographic) are not evenly distributed between locations. It can be partly explained in terms of differences in demography [33-39] Variation between locations (e.g. centres and countries) in the epidemiology can also translate onto different case-mixes between locations with obvious impact on the cost-effectiveness of a given treatment in a specific location.[36,38,40]

Clinician factors

Clinicians can influence the effectiveness, cost and cost-effectiveness of interventions. This 'clinician effect' is particularly important in non-drug interventions (e.g. surgical), but pharmaceutical trials can also display between-clinician differences – for instance - in background treatments given to patients over and above the study treatment, or in the

management of adverse events. The clinician effect is typically not easy to quantify within clinical[41-43] or economic evaluations.[44,45] In part, variation in how clinical staff perform can be due to the fact that healthcare systems differ also in terms of the incentives that they offer to staff.[11,36,38,46-48]

Healthcare system factors

There may be numerous differences between countries and centres [36,38,40,47,49-56] (other than patients and clinicians characteristics) in terms of the process of healthcare delivery. Between-country differences in relative cost may be influenced by the technology involved in the production of healthcare, the level of substitution between labour and capital, and the types and cost of resource inputs used in production of healthcare.[57] This within-country variation is likely to be particularly pronounced in large and economically heterogeneous countries. Clinical practice and conventions are also known to differ widely between (and within) countries.[33,36,37,47,50,53,55,58,59]

Wider socio-economic factors

Other factors which may have an impact on the generalisability of CEA results from country to country relate to variation between locations in terms of more general socio-economic factors. The willingness (and ability) of a region/country to devote resources to healthcare is one of these factors. Another broader factor discussed by Bernie O'Brien[11] relates to the health-related preferences of the population such as those reflected in health state utilities used to calculate quality-adjusted life years (QALYs).

3 What methods are there to make cost-effectiveness estimates more country-specific?

In view of the arguments developed above and in consideration of the time and effort required to complete a multinational trial-based CEA, it seems reasonable - from the viewpoint of both the industry and national/state governments - to support the use of methods which facilitate the 'translation' of cost-effectiveness data obtained from one country to make them applicable to another. The need to customise the economic study results to a specific jurisdiction is not purely academic, but stems from the decision-makers' need for context-specific information. This raises two overarching methodological questions: 'what methods are there to make cost-effectiveness estimates more country-specific?', and 'how can we account for factors that may affect the between-country generalisability of cost-effectiveness results?'

In broad terms, the analytical options available range from the use of regression-based techniques to the application of decision-analytic models, and are already part of the toolkit of the health economist working in CEA. Decision models are typically used when the evidence base from the trial(s) of interest is available exclusively in summary format.[13] However, there are examples where IPD from a single multinational trial have also been used - in combination with non-trial IPD - to populate a decision model with the objective to generate cost-effectiveness estimates for a country different from the one where the trial had been carried out. Regression-based methods, on the other hand, are used mainly when the country of interest actively recruited patients into the trial and the analyst has access to the study IPD (or at least country-specific summary data).

It emerges that the answers to the questions posed above depend, therefore, upon whether or not (a) the country of interest participated in the trial, and (b) IPD from the trial are available. To structure the discussion in a logical way, this manuscript proposes an algorithm (presented in Figure 1) developed to assist the decision as to which analytical strategy to adopt when faced with the task of generating country-specific cost-effectiveness estimates.

3.1 Individual-patient level data are available but the country of interest did not participate in the trial

Let us start from the situation where, despite the IPD from the multinational (multicentre) trial being available, the country of interest did not participate in the trial. The decision-maker in the country of interest will be interested in the extent to which the results from this trial apply to her own setting. In this case, some form of decision modelling to extrapolate the study results from one country to another will be required. This is a very common situation in health technology assessment (HTA) and there are various examples in the literature (see relevant chapters in the reviews by Sculper *et al*[7] and Goeree *et al*[8]).

As discussed in section 2, it is possible not only for resource use and cost data to vary by location, but the same can apply to clinical data. In the applied work, analysts have addressed this issue by assuming that the baseline risks for particular clinical events are location-specific, whilst the relative treatment effect is more generalisable across locations.[60] In this case it is considered good practice to develop an 'events based model' built around 'generalisable' features of the disease or patient's prognosis, and use the IPD from the trial to estimate the likelihood of occurrence of the clinical events of interest which are expected have an impact on resource use and / or health-related quality of life. The trial-wide relative treatment effect such as, for instance, the relative risk reduction (*RRR*) in the event(s) of interest observed in the trial (e.g. relative reduction in risk of deaths, MI, side effects), is then applied to the reference (baseline) risk (R_0) - i.e. the event rate without the treatment - for the country of interest. The latter information can be ascertained either from long-term follow-up cohort studies, or (more practically) using existing risk equations assuming that the risk factors (e.g. age, tobacco consumption, etc) between the trial population and that in the country of interest are the same, regardless of the country.[61] Different distributions of these risk factors in different countries will translate in differences in country-specific baseline risks. Because cost-effectiveness is essentially concerned with absolute differences (in costs and effects) the absolute number of events averted - for instance, should this be the measure of clinical outcome – in the new country is simply obtained multiplying the trial-wide treatment effect by the baseline event rate in the country of interest.

A well known example of this methodology is the application of the West of Scotland Coronary Prevention Study (WOSCOPS) cost-effectiveness results[62] to Belgium[61], Canada, Sweden and South Africa.[63] The WOSCOP study[64] concluded that treatment with pravastatin reduced the risk of first-time heart attack and death in middle-aged hypercolesterolaemic men. Because of the increased risk of cardiovascular disease (CVD) in the Scottish male population, though, the authors were keen to address possible concerns regarding the generalisability of the study findings to other countries. Using a system of competing risk equations Caro *et al* combined trial (i.e. relative risk reduction of cardiovascular events) and non-trial IPD (i.e. baseline event rates based on risk equations and risk factors distribution for Belgium, local costs, and life expectancy) within a decision-analytic model to estimate the cost-effectiveness of pravastatin in Belgium.[61] A simplified structure of their model is represented in Figure 2.

At any point in time men with hypercholesterolaemia were assumed to be either (i) alive without experiencing cardiovascular events (and be still at risk in the following period), (ii) dead following a non-cardiovascular event, (iii) have a non-fatal cardiovascular event (in which case a given life expectancy was estimated), or (iv) dead following cardiovascular event. Parameter estimates from the risk equations governing the above transitions in a population not receiving the active treatment were obtained applying an exponential regression to the WOSCOP trial IPD, considering a set of risk factors (e.g. age, high diastolic blood pressure, smoking, etc). Parameter estimates from the exponential regression model were then applied to IPD on the same set of risk factors obtained from a Belgian epidemiological study, to predict individual-

patient (and average) probabilities of cardiovascular events for the cohort in the model.[61] For each cardiovascular event considered in the model, direct costs were estimated using a registry which includes 35% of the hospitalisations in Belgium. Similarly, local data was used to estimate drug costs, while mortality tables for Belgium were used to extrapolate long term survival.

It should be pointed out that this approach can also be used in case the country of interest participated in the trial, but IPD are only available for the clinical outcome (i.e. no country-specific resource use data collection for the country of interest had taken place). Furthermore, when the researcher is interested in the cost-effectiveness of the intervention in particular sub-groups of patients the analysis above would need to be run separately for different risk groups (based on the regression results).

3.2 Individual-patient level data are available and the country of interest participated in the trial

The class of models discussed above typically relies on the (often untested) assumption that the relative clinical efficacy is independent of the disease underlying baseline risk, and that while the latter captures a range of country-specific factors (e.g. epidemiology, medical attitude, etc) the relative clinical effectiveness of the intervention does not differ greatly across countries.[36] The use of clinical data collected from different countries (centres) to estimate a single relative treatment effect on clinical outcomes is therefore an accepted practice.[65] However, as recognised by Caro *et al*[61,63] it is possible that between-country variability in the risk factors could lead to different cost-effectiveness results in different locations. Between-country differences in the distribution of individual-patient level risk factors (e.g. age, smoking status, blood pressure) are always likely to exist, which when paired with possible difference in country-specific factors (e.g. type of healthcare system, percentage of national GDP spent on healthcare, etc.) could limit the generalisability of the relative treatment effect observed in the trial.

Despite these considerations, the assumption regarding the generalisability of the relative treatment effect on the clinical outcome from one country (or a set of countries) is only rarely scrutinised,[7] even when access to IPD from the multinational trial is not an issue. When assessed, it is done so using a test of heterogeneity[66] despite the fact that this test is typically underpowered.[14,67] In the 'absence of evidence' (which does not imply 'evidence of absence') about between-country heterogeneity in the data, the results of this test are used as a basis to justify the analysis of the pooled (clinical and resource use) data,[67] regardless of the country of origin. The implication is that non-statistically significant between-country differences in relative treatment effect may not be a concern for a given country decision-maker, even when these are qualitative in nature (i.e. the treatment effect in different countries not only differs in magnitude but also in direction).

It can be argued, though, that the presence of between-country differences in the magnitude and sometimes in the direction of the relative treatment effect is a fundamental consideration which, when paired with international differences in factors affecting resource use and costs, makes the estimate of cost-effectiveness for a particular country based on the trial-wide relative treatment effect unreliable, even when individual risk factors (and their distributions) are similar between countries.

One of the earliest attempts to address the statistical analysis of multinational clinical trials for a cost-effectiveness analysis was presented by Willke *et al.*[68] Using a system of related regression equations, the authors developed a novel approach to explore the between-country variability in the CEA results by looking at the treatment-by-country interactions in both effectiveness and costs. Willke *et al* presented the cost-effectiveness results for five separate

countries under different assumptions regarding the generalisability of the data from the trial. Their results emphasized the differing spread of estimates that could be obtained under different approaches. The authors compared (i) a fully *pooled* analysis with multinational costing, which produced as a single cost-effectiveness estimate for the whole trial, assuming trial-wide effectiveness; (ii) a *pooled* analysis (for each country) with price weights from the individual countries again assuming trial-wide effectiveness, and which produced very little variability in the results; (iii) a strategy relating trial-wide effectiveness and country specific costs (with countries own price weights), which provided a much greater spread, and (iv) the fully *split* analysis, which resulted in the widest between-country variation.

The latter approach is equivalent to *splitting* the data and running a series of regression analyses for each country independently from the others. The potential problem with this approach is that it requires a choice to be made between 'pooling' *vs* 'splitting'. The limitations of a pooled analysis have been discussed already. Splitting the data, on the other hand, is impractical when the country of interest has recruited a limited number of patients compared with the rest of the countries in the trial.[12]

Furthermore, it can be argued that data collected from different countries (and patients) may share some degree of similarity and that there may be advantages in trying to capture such similarities. One way to do this is to reflect the hierarchical structure in the multinational data, inherent in the natural clustering arising from patients being recruited in specific countries and receiving treatment in centres with different characteristics.

Various authors have explored the use of hierarchical regression models[69] for the analysis of multinational (and multicentre) trial-based cost-effectiveness data.[70-75] A simple hierarchical model for either cost (or health outcomes) data can be described as follows. Let Y_{ij} be the observed cost (health outcome) of individual i in country j, and t_{ij} be an indicator variable taking values 0 (control) and 1 (intervention), depending on whether the patient has been treated respectively in the control or the intervention group. The regression model for multinational cost-effectiveness data can be described as follows

$$Y_{ij} = a_j + \beta_j t_{ij} + e_{ij}$$
$$a_j = a + v_j$$
$$\beta_j = \beta + u_j$$

(1)

where, the coefficients α_j and β_j are respectively the country j mean cost in the control arm and the differential mean cost between the two arms of the trial, so that the mean cost in the intervention group in country j is given by ($\alpha_j + \beta_j$). The model can be re-written as

$$Y_{ij} = a + \beta t_{ij} + v_j + u_j t_{ij} + e_{ij}$$

(2)

where the last three terms represent the random components at the country (u_j and v_j) and individual (ε_{ij}) level, usually assumed to follow a normal distribution with mean zero.[69]

The terms α and β in (2) are overall (fixed) effects representing the trial-wide estimates, while v_j and u_j are the 'random effects' representing, respectively, the jth country-specific departure from the overall mean cost in the control group (v_j) and the differential mean cost, u_j.[b] Equation (2) allows to partition the overall variability observed in the data in two components, one associated with variation at patient-level (ε_{ij}), and the other(s) associated with variation at

---

[b]Note that v_j and u_j are assumed to follow a bivariate normal distribution, to reflect the fact that in each country's mean cost in the control arm is correlated to the differential mean cost. In the analysis of the clinical data this assumption would reflect the fact that the baseline events are correlated with the relative treatment effects.

country level (u_j and v_j). Finally, the parameter of interest, the country j specific mean difference in is given by ($\beta + u_j$).

With respect to the *splitting* approach, which requires a treatment-by-country interaction term, (2) assumes that the random effects are 'latent' variables with a specific distribution, representing the potential departures that the country-specific effects could have from the overall mean. By assuming country effects as random, these models allow the country-specific estimates to be obtained using, not only the data from the country of interest, but also the data from the other countries that participate in the trial. In this sense,[76] the estimates are said to be *borrowing strength* from each other. Willan *et al*[70] showed how this class of models can be used to analyse multinational trial-based aggregate country-specific cost-effectiveness data. Others[71-75] illustrated the application of hierarchical regression models for costs and CEA in presence of trial-based IPD.

That the analysis based on IPD from multinational trials offers more flexibility is undisputed. While allowing for the inclusion of a set of country-specific covariates, the aggregate level data analysis does not facilitate inclusion of individual-level covariates, which may lead in some cases to the problem of 'ecological fallacy'.[76,77] The hierarchical model with IPD, on the other hand, offers the potential to analyse the full dataset, hence, the possibility to accommodate both patients- and country-specific covariates. In particular, using both patient and country level covariates the hierarchical modelling can be particularly useful when attempting to explain the observed between-country variability in the results of multinational trial-based CEAs. Given the need to produce robust cost-effectiveness evidence for jurisdiction-specific decision-making, it is only by incorporating both patient and country level covariates that potential between-country heterogeneity in the cost-effectiveness of the intervention can be fully explored and accounted for.

Manca *et al*[78] used hierarchical multivariate models to reanalyse the economic data from a large multinational clinical study, the Assessment of Treatment with Lisinopril and Survival (ATLAS). This multinational trial enrolled 3164 patients in 19 countries, and compared low dose and high dose of the ACE inhibitor lisinopril in patients with chronic heart failure. Details of the main economic and clinical analyses have been reported elsewhere.[79,80] This case study uses a total of 3061 observations (low dose, n=1545; high-dose, n=1516) from 17 countries. The analysis reported here refers to the first three years of follow up. Therefore, due to these assumptions, it must be stressed that the specific results presented here are not to be considered in alternative of the main study report.[79] The authors compared the splitting and pooling approaches in the analysis of international cost-effectiveness data against the recently proposed use of hierarchical models.

Figures 3 and 4 show the results of this analysis respectively for the mean difference in costs and survival gain, and compare the *splitting* and *pooling* approaches (on the left hand graph of each figure) against the hierarchical modelling strategy. It can be seen how the country-specific estimates (empty square markers) obtained using the splitting approach display a large dispersion around the overall mean (black circle marker) obtained using the pooling approach. The hierarchical regression model approach, on the other hand, produces country-specific estimates which are closer to the population mean (the black circle at the bottom of the graph on the right hand side) compared to their counterparts in the analysis obtained by splitting the data. That is, the country-specific estimates in the hierarchical model borrow strength from each other. Some countries are more or less *shrunken* towards the overall mean. The degree of shrinkage is proportional to the between and within country variances, as well as the country-specific sample size.[72]

One of the criticisms often moved to the use of hierarchical models in multinational CEA is that this method may be inappropriate when we expect systematic differences between countries but are not able to 'explain' these adequately using country-level covariates. In other words, the assumption that country-specific random effects are drawn from a common distribution may be erroneous. This criticism can be addressed using the arguments proposed by Gelman *et al* [81], who state 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 θ_j's differ, but it might be perfectly acceptable to consider them as drawn from a common distribution. […] 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." (page 124).

In essence this means that,

"…the usual way to model exchangeability with covariates is through conditional independence $p\left(\theta_1, \theta_2, \theta_3, ..., \theta_{\mathcal{T}}\right) = \int\left[\prod_{j=1}^{\mathcal{T}} p\left(\theta_j \mid \varphi, x_j\right)\right] \cdot p(\varphi \mid x)d\varphi$ *with* $x = (x_1, x_2,...,x_J)$. In this way exchangeable models become almost universally applicable, because any information to distinguish different units should be encoded in the x and y [outcome] *variables"* (page 123).

As discussed in section 2, there are various factors that could explain the between-country variation in costs and effects differences observed in Figures 3 and 4. The re-analysis implemented here developed a Bayesian bivariate hierarchical regression model for cost and survival data in the trial, while controlling for a set of patient and country specific covariates. [78] Figures 5 and 6 plot the country-specific differential cost and survival gain, respectively, against life expectancy at birth and the public expenditure in healthcare as a percentage of the national GDP. Both graphs indicate a positive relationship between the treatment effects and the country-specific covariate, suggesting that these factors may need to be accounted for when assessing the generalisability of the cost-effectiveness results between countries.

3.3 Individual-patient level data are unavailable and the country of interest did not participate in the trial

A more challenging situation is when IPD are unavailable and the country of interest did not participate in the trial. In this case the analyst has to rely on data published in the literature, and to assume that the relative treatment effect estimated from other countries is indeed generalisable to the country of interest. In this case, methods similar to those explained in section 3.1 can be used, again supplementing the evidence base with additional IPD specific to the country of interest.

The example here relates to an analysis undertaken to inform NICE on the cost-effectiveness of using Glycoprotein IIb/IIIa antagonists (GPAs) in the management of non-ST-elevation ACS and illustrates a real life example of a decision model developed for the National Institute for Health and Clinical Excellence.[82] In this section, we focus on the methods used in this example[7,82] and the steps followed in building the model. The paper by Sculpher and Drummond[13] at this conference discusses the policy rationale behind this approach. There were several challenges relating to the generalisability of the data the authors had to deal with in building a model that was relevant for the decision-maker in the UK (i.e. the NHS). These challenges together with the solutions adopted to address them are reviewed in turn.

3.3.1 Relating the model to clinical practice—As mentioned in section 2.3, differences in clinical practice may be an important factor which needs to be accounted for when translating study results from one country to another. In developing the model for NICE, an important consideration was therefore how GPAs would be used in the UK routine clinical practice. The evidence base contained two types of GPAs trial: those comparing the drugs with standard practice (i.e. management without GPAs) in all patients with non-ST elevation ACS regardless of whether a percutaneous coronary intervention (PCI) was subsequently undertaken (medical management); and those which looked at GPAs as an adjunct to PCI. Four treatment strategies were considered to be relevant for the UK, and the model was structured to compare all of them against each other. Lack of trial evidence comparing these strategies head-to-head was overcome by using 'evidence synthesis' methodology such as indirect comparison.[83] Hence, it was necessary to re-structure the effectiveness data to reflect the nature of the indirect clinical comparison which was needed to populate the decision model. This was achieved by separating out the baseline event rates measured in the standard therapy control groups in the trials from the treatment effect observed in the GPA arms relative to the control group. The relative treatment effects for each treatment strategy were pooled across the various groups of trials. [82]

3.3.2 Are baseline event rates in the trials relevant to UK practice?—Given that the trials were undertaken largely outside the UK, the baseline event rates in patients not having GPAs in the UK were considered to be potentially quite different to those patients randomised to the control groups in the trials. As mentioned in Section 2, this could reflect differences in the epidemiology of the disease or, more probably, differences in overall management of patients with ischaemic heart disease (IHD) in the UK. After consulting with UK clinical experts, Palmer *et al*[82] considered the principal difference in the management of IHD in the UK, compared to that in other developed countries, was that fewer patients were considered for PCI at the time of the analysis. It was felt that the lower rates of PCI in the UK could have the effect of generating higher baseline event rates than those observed in the literature. Secondly, the limited availability of 'acute' PCI (i.e. percutaneous procedures undertaken in non-ST elevation patients shortly after presentation) in the NHS could cause clinicians to select ACS patients for acute PCI in a different way than clinicians in the GPA trials. Therefore, baseline event rate data, which were specific to UK practice, were sought and information from the Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK), which is an observational cohort registry of 1046 patients admitted to 56 UK hospitals with ACS in 1999, was used.

3.3.3 Are the relative risk reductions estimated from the trials related to baseline risk?—In the absence of IPD from the trials of interest, one way of adapting the clinical results from international trials to the UK setting is by separating out the baseline event rates associated with standard management (without GPAs), estimating those parameters from UK-specific data and applying the pooled relative treatment effects, for the alternative treatment strategies being considered relative to the control (i.e. no use of GPAs), from the trials. This amounts to assuming that baseline risks are not transferable internationally, but relative risk reductions are. It may be, however, that the relative treatment effect is itself related to baseline risk – for example, the higher the baseline risk, the lower the treatment effect – in which case the assumed independence between the two components of clinical effectiveness is not sustainable. It was important therefore to ascertain whether any relationship existed between the relative treatment effect (i.e. relative risk reduction) and the baseline risk observed in the literature. If that was the case, then this had to be built into the decision model in order to relate the UK baseline risk (taken from local sources) and the relative risk reduction estimated from the literature.

In order to investigate whether the log relative risk in the individual trials varied with log baseline risk (i.e. the log event rate in the control group), a random effects meta-regression

model[84,85] was used.[c] This form of meta-regression works by fitting a regression line between the event rates measured in the control groups of the trials and those of the experimental (i.e. GPA) groups, with the number of points from which the estimate is made being the number of available trials. This function characterises the relationship between the baseline risks and the relative risks. Once the function has been estimated in the meta-regression, the pooled relative risk estimates from the trials could be adjusted according to the point on the regression line which accords with the UK baseline risk. If the results suggest that a relationship exists, the decision model can be used to adjust the relative risk estimates according to the baseline risk employed in the model.

Figure 7 shows the results of this meta-regression, plotting the relationship between baseline risk (in the control group) and relative risk – on the log scale for the group of trials relating to one of the strategies considered in the model (i.e. GPA as part of initial medical management). The results of the analysis showed that there was a negative relationship between log baseline risks and log relative risk. This relationship was found to be not statistically significant, and the authors decided to estimate the relative risk reduction using a meta-analysis. This choice was justified not only by the lack of statistical significance, but probably more realistically by (a) the lack of direct evidence (in the studies included in the meta-regression) directly referring to the baseline risk observed in the UK, and (b) the inherent threats involved in using meta-regression for clinical decision-making.[86-88]

3.3.4 Incorporating UK specific resource use and costs—An important function of many decision models is to extrapolate the effectiveness data available in trials.[d] Palmer *et al* [82] had the problem of taking the short-term effectiveness data in the available trials, which typically had follow-up of no more than six months, and extrapolate them to the lifetime time horizon of the typical patient. Therefore, a long-term (extrapolation) model was developed to estimate the future prognosis for patients who finish the short-term (six month) model in one of two disease states: those having experienced a non-fatal MI and those who have not but remain alive. The structure of the model is illustrated in Figure 8.

The long-term model took the form of a 4-state Markov process. The model assumed that at any point in time, patients could be in one of the following states: ischaemic heart disease (that is, patients who had not experienced a non-fatal MI), non-fatal MI (not showed in figure 8) where patients spent a single cycle of one year, post-MI (where surviving patients entered after one year following an MI) and death. It was necessary to be aware of the variation between countries in long-term survival following cardiac events. Transition probabilities were, therefore, taken from a UK-specific observational study - the Nottingham Heart Attack Register (NHAR). Two cohorts of patients (total n = 1,279) from the NHAR were used, with a diagnoses indicative of ACS, which had follow-up data for up to 5 years. In the context of the GPA model, the relevant 'events' estimated in the short-term model included revascularisation rates and days in hospital. For the longer-term model, these events included MIs, revascularisations and days in hospital. In both cases, these estimates of event rates were based on data from the UK observational studies – PRAIS-UK for the short-term model and NHAR for the long-term model. All estimates of the costs of these events were then taken from UK sources. For further details about this model the reader is invited to refer to the original publications.[82,89]

---

[c]An analysis similar to what would be carried out if one were to explore the generalisability of the absolute treatment effect identified in trials across a range of clinically-defined patient sub-groups where the same separation of baseline risks and relative treatment effect can be employed.
[d]There are various dimensions in the extrapolation problem. This can relate to beyond trial extrapolation (that is from short to long term outcomes), from intermediate endpoints to final outcomes, and from intermediate endpoints or final clinical outcomes to health-related quality of life.

3.4 Individual-patient level data are unavailable and the country of interest participated in the trial

The methods presented in 3.3 section may be a sensible approach even when, although the trial produces evidence directly relevant to the country of interest[90] (e.g. the trial was carried out in, or included, the country of interest and a country-specific estimate is available), neither aggregate (country-level) nor IPD are available. There are several reasons for this, which mainly relate to the perceived limitations of using evidence from a single trial to inform decision-making in a specific jurisdiction (country).

It has been argued that most trials will not randomise individuals to all the relevant management options available in a specific healthcare system. Furthermore, the period of follow-up in a trial will often be shorter than the relevant time horizon for cost-effectiveness decisions; and finally, reliance on a single trial will often ignore other relevant evidence[90-93] possibly more location-specific, which could be used to adapt the trial results to the country of interest. While the use of (clinical and cost-effectiveness) evidence from a single multinational (or multicentre) RCT to inform reimbursement decisions in a given country is still debated,[90,92,93] clinical trials will always play a crucial role in providing unbiased country-specific estimates of (clinical) treatment effects.[92] Key data on the likelihood of particular events (e.g. side effect, complications, etc) and their relationship with quality of life and resource use implications in particular treatment settings (e.g. hospital, country) is also a fundamental piece of information for healthcare decision makers. The evidence base concerning the treatment strategy being examined could include several trials comparing various alternative treatment strategies, giving rise to a 'network of evidence'. In this case it is 'good practice' to synthesise all the available evidence within a comprehensive decision-analytic model,[82,90,92-97] the parameters of which (e.g. treatment effect, transition probabilities, etc) can be estimated through an evidence synthesis model (e.g. indirect[98-100]or mixed treatment comparison[83,101,102]).

The difference with respect to the methodology described in Section 3.3 is subtle. It is argued here that when the country of interest participated in one or more of the trials that form the evidence base used to inform the estimation of the parameters used in the model, the synthesis model should include not only explanatory variables at study-level, but also a country (or at the very least a geographical area) indicator identifying in which the study was carried out. The model could be supplemented with country-specific IPD on baseline risk, resource use, natural history of the disease, and mortality tables as explained in section 3.1.

4 Discussion

Methods which enable policy makers to assess the extent to which the trial (economic and clinical) results are valid in different geographical settings should have a key role in Health Technology Assessment (HTA). There are several reasons for this. First and foremost, reimbursement decisions are made at local (i.e. jurisdictional and country) levels. Furthermore, despite their interest in a set of specific interventions, many healthcare systems (especially in mid-income and developing countries) are unable either to fund (or participate in) relevant trials. Third, the efficient use of R&D funding for HTA at country level requires the avoidance of duplication of funding efforts to address research questions already being addressed in other jurisdictions. Given such a complex scenario, decision-makers need to be able to discern to what extent the observed variability is country-specific and to what extent it is patient-related. There is, therefore, a great value in developing methods that enable decision-makers to use international data to generate jurisdiction-specific cost-effectiveness estimates. Several authors have attempted to address the issue of how should multinational trial-based cost-effectiveness data be analysed, with different degrees of sophistication. Others have developed decision-modelling methods in combination with evidence synthesis techniques to overcome the lack

of IPD to produce cost-effectiveness estimates under specific assumptions (e.g. that the relative treatment effect is generalisable between countries).

This paper reviews the methods that could be used to reflect between-location differences in cost-effectiveness, and provides guidance as to which approaches can be currently considered 'good practice'. An algorithm helping decide which strategy to adopt under different scenarios faced by the analyst is proposed. These scenarios are characterised in terms of trial-based IPD availability and whether the country of interest participated in the study.

Although, these methods have been used mostly to inform decisions at national and state level, there are reasons to believe that they could be useful to inform policy decisions in jurisdictions within a given country/state (eg. HMOs, regions in federalist countries, GP fund holding).

It must be emphasised that the methodology in this research area is currently under rapid development. However, various considerations can be made in terms of recommendations for future research. The paper by Sculpher and Drummond at this conference provides a list of the main issues to consider in trial- and model-based cost-effectiveness analysis in relation to issues of generalisability. These include aspects relevant to the design, data collection, analysis, and reporting of the study results.

This paper has shown that the availability of IPD evidence external to the trial referring to the country of interest is often paramount, especially (but not exclusively) when the country of interest did not participate in the clinical study. It was argued that, in multinational trial-based CEA, patients and country-specific covariates are essential for an accurate assessment of the between-location variability of the cost-effectiveness results. While baseline patient-level data are routinely collected as part of the trial, this is not the case for country (centre) level characteristics considered to have the potential to affect the generalisability of the study results. Questions that remain to be addressed regard, what country (centre) covariates should be collected, can these be routinely collected, if not what alternative sources can be used. This concern affects another element relevant during the design phase of the study: the selection of countries (and centres) participating in the trial. Ideally this selection should be at random. Current methods of selection are unclear, but they probably reflect the level of funding for the cost-effectiveness study.

In terms of modelling methods available it has been shown that hierarchical regression is a useful tool to analyse IPD collected alongside multinational trials. On the other hand, there are valid arguments for accepting that within-trial analysis is often not sufficient to appropriately inform reimbursement decisions at country level and that some form of additional modelling is probably required (e.g. long term extrapolation, synthesis of additional evidence, additional comparators). In this sense, it is possible that Bayesian methods integrating individual- and aggregate-level data could provide additional flexibility in the analysis of cost-effectiveness data for policy decisions. Further methodological developments are expected.

Finally, in terms of reporting of the study results Drummond *et al*[9] indicated that more transparency would help ascertain the generalisability of the study results to a specific context. A recent review by Urdahl *et al*[103] assessed the extent to which applied modelling studies in osteoporosis incorporated data inputs, which were appropriate for the target jurisdiction or decision-maker as stated or inferred from each study. It was found that studies tended to be more assiduous in selecting cost inputs, which were specific to their target decision-maker than they were in identifying appropriate clinical inputs. This is likely to reflect an implicit assumption that parameters relating to clinical effectiveness, whilst needing to be specific to the relevant patient group defined in the decision problem, are inherently more transportable geographically. Whilst this assumption may be justified within healthcare systems and

countries, the factors discussed in section 2 of this paper suggest that this will not necessarily be so between systems and countries.

5 Conclusion

Given the decision-makers' need for jurisdiction-specific cost-effectiveness information, and the presence of country-specific factors that contribute to the international variation in cost-effectiveness results, trial-wide (i.e. pooled) results may not always be useful for jurisdiction-specific resource allocation decisions.[9] Methods to reflect (and address) between-country differences in cost-effectiveness data are available and are likely to develop further.

References

1. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence (NICE); 2004.

2. AMCP. The American Managed Care Pharmacy Format for Formulary Submissions. Alexandria, VA: The Foundation for Managed Care Pharmacy; Apr. 2005 2005

3. CADTH. Guidelines for the economic evaluation of health technologies: Canada. 3rd Edition. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.

4. Commonwealth Department of Health Housing and Community Services. Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: AGPS; 1992.

5. Gricar, JA.; Langley, PC.; Luce, B. AMCP's Format for Formulary Submissions: A Format for Submissions of Clinical and Economic Evaluation Data in Support of Formulary Consideration by Managed Health Care Systems in the United States. Alexandria, VA: Academy of Managed Care Pharmacy (AMCP); 2002. al. e.

6. SMC. New product assessment form. Glasgow: Scottish Medicines Consortium; 2005.

7. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Davies LM, et al. Generalisability in Economic Evaluation Studies in Health Care: a Review and Case-Studies. Health Technology Assessement 2004;8(49):1–206. [PubMed: 15544708]

8. Goeree, R.; Burke, B.; Manca, A.; Sculpher, MJ.; Willan, AR.; Blackhouse, G., et al. Generalizability of economic evaluations: using results from other geographic areas or from multinational trials to help inform health care decision making in Canada. CCOHTA HTA Capacity Building Grants Program. Toronto: Canadian Coordinatin Office for Health Technology Assessment; 2005. Report No.: Grant N. 67

9. Drummond MF, Manca A, Sculpher MJ. Increasing the Generalisability of Economic Evaluations. Recommendations for the Design, Analysis and Reporting of Studies. International Journal of Technology Assessment in Health Care 2005;21(2):165–71. [PubMed: 15921055]

10. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Economics 1997;6(3):217–27. [PubMed: 9226140]

11. O'Brien BJ. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. The American Journal of Managed Care 1997;3:S33–S39. [PubMed: 10180339]

12. Reed SD, Anstrom KJ, Bakhai A, Briggs AH, Califf RM, Cohen DJ, et al. Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. Am Heart J 2005;149 (3):434–43. [PubMed: 15864231]

13. Sculpher, MJ.; Drummond, MF. Paper presented at the conference "Better Analysi for Better Decisions - A conference in honour of Bernie O'Brien". Hamilton (Ontario), Canada: McMaster University; 2006. Issues in the transferability of economic evaluation.

14. O'Shea JC, DeMets DL. Statistical issues relating to international differences in clinical trials. American Heart Journal 2001;142(1):21–28. [PubMed: 11431652]

15. Lewis JA. Statistical principles for clinical trials (ICH E9): An introductory note on an international guideline. Statistics in Medicine 1999;18(15):1903–1904. [PubMed: 10440877]

16. Lewis J, Louv W, Rockhold F, Sato T. The impact of the international guideline entitled Statistical principles for clinical trials (ICH E9). Statistics in Medicine 2001;20(1718):2549–2560. [PubMed: 11523068]

17. International Conference on Harmonisation. International Conference on Harmonisation; guidance on statistical principles for clinical trials; availability--FDA. Notice. Federal register 1998;63(179): 49583–49598. [PubMed: 10185190]

18. International Conference on Harmonisation. International Conference on Harmonisation; choice of control group and related issues in clinical trials; availability. Notice. Federal register 2001;66(93): 24390–24391. [PubMed: 12356096]

19. Chang WC, Midodzi WK, Westerhout CM, Boersma E, Cooper J, Barnathan ES, et al. Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. Journal of Epidemiology and Community Health 2005;59(5):427–433. [PubMed: 15831694]

20. O'Shea JC, Califf RM. International differences in treatment effects in cardiovascular clinical trials. American Heart Journal 2001;141(5):875–880. [PubMed: 11320382]

21. O'Shea JC, Califf RM. International differences in cardiovascular clinical trials. American Heart Journal 2001;141(5):866–874. [PubMed: 11320381]

22. Gupta M, Chang W-C, Van de Werf F, Granger CB, Midodzi W, Barbash G, et al. International differences in in-hospital revascularization and outcomes following acute myocardial infarction: A multilevel analysis of patients in ASSENT-2. European Heart Journal 2003;24(18):1640–1650. [PubMed: 14499226]

23. Mark DB, Naylor CD, Hlatky MA, Califf RM, Topol EJ, Granger CB, et al. Use of medical resources and quality of life after acute myocardial infarction in Canada and the United States. New England Journal of Medicine 1994;331(17):1130–1135. [PubMed: 7935638]

24. Grieve R, Hutton J, Bhalla A, Rastenyte? D, Ryglewicz D, Sarti C, et al. A comparison of the costs and survival of hospital-admitted stroke patients across Europe. Stroke 2001;32(7):1684–1691. [PubMed: 11441220]

25. Weir NU, Sandercock PAG, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke 2001;32(6):1370–1377. [PubMed: 11387501]

26. Van de Werf F, Topol EJ, Lee KL, Woodlief LH, Granger CB, Armstrong PW, et al. Variations in patient management and outcomes for acute myocardial infarction in the United States and other countries: Results from the GUSTO trial. Journal of the American Medical Association 1995;273 (20):1586–1591. [PubMed: 7745771]

27. Barbash GI, Modan M, Goldbourt U, White H, Van de Werf F. Comparative case fatality analysis of the International Tissue Plasminogen Activator/Streptokinase Mortality Trial: Variation by country beyond predictive profile. Journal of the American College of Cardiology 1993;21(2):281–286. [PubMed: 8425987]

28. Pilote L, Califf RM, Sapp S, Miller DP, Mark DB, Weaver WD, et al. Regional variation across the United States in the management of acute myocardial infarction. New England Journal of Medicine 1995;333(9):565–572. [PubMed: 7623907]

29. Postma MJ, Leidl R, Downs AM, Rovira J, Tolley K, Gyldmark M, et al. Economic impact of the AIDS epidemic in the European Community: towards multinational scenarios on hospital care and costs. AIDS 1993;7(4):541–53. [PubMed: 8507419]

30. Rhodes G, Wiley M, et al. Comparing EU hospital efficiency using diagnostic-related groups. European Journal of Public Health 1997;7(Supplement 3):42–50.

31. Cohen MG, Pacchiana CM, Corbala R, Isea Perez JE, Ponte CI, Oropeza ES, et al. Variation in patient management and outcomes for acute coronary syndromes in Latin America and North America: Results from the Platelet IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy (PURSUIT) trial. American Heart Journal 2001;141(3):391–401. [PubMed: 11231436]

32. Lingard EA, Berven S, Katz JN, Gillespie W, Howie C, Annan I, et al. Management and care of patients undergoing total knee arthroplasty: Variations across different health care settings. Arthritis Care and Research 2000;13(3):129–136. [PubMed: 14635286]

33. Stason, WB. Cost-effectiveness analysis in health care: opportunities and challenges to international comparisons. In: Lasser, U.; Roccella, EJ.; Rosenfeld, JB.; Wenzel, A., editors. Costs and benefits in health care and prevention: an international approach to priorities in medicine. Berlin: Springer-Verlag; 1990.

34. Baker AM, Goldberg A, Arnold RJ, Kaniecki DJ. Considerations in measuring resource use in clinical trials. Drug Information Journal 1995;29:1421–1428.

35. Baltussen R, Ament A, Leidl R. Making cost assessments based on RCTs more useful to decision-makers. Health Policy 1996;37(3):163–183. [PubMed: 10160021]

36. Drummond MF. Comparing cost-effectiveness across countries. The model of acid-related disease. Pharmacoeconomics 1994;5(S3):60–67.

37. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world: Effectiveness versus efficacy studies. Pharmacoeconomics 1999;15(5):423–434. [PubMed: 10537960]

38. Drummond MF. The future of pharmacoeconomics: bridging science and practice. Clinical Therapeutics 1996;18(5):969–78. [PubMed: 8930435]

39. Bailey KR. Generalising the results of randomized clinical trials. Controlled Clinical Trials 1994;15:15–23. [PubMed: 8149769]

40. Bryan S, Brown J. Extrapolation of cost-effectiveness information to local settings. Journal of Health Services Research and Policy 1998;3:108–112. [PubMed: 10180659]

41. Spiegelhalter DJ. Surgical audit: Statistical lessons from Nightingale and Codman. Journal of the Royal Statistical Society. Series A: Statistics in Society 1999;162(1):45–58.

42. Normand S-LT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: Issues and applications. Journal of the American Statistical Association 1997;92(439):803–814.

43. Goldstein H, Spiegelhalter DJ. League tables and their limitations: Statistical issues in comparisons of institutional performance. Journal of the Royal Statistical Society. Series A: Statistics in Society 1996;159(3):385–443.

44. Roberts C. The implication of variation in outcome between health care professionals for the design and analysis of randomised controlled trials. Statistics in Medicine 1999;18:2605–2615. [PubMed: 10495459]

45. Hall BL, Hamilton BH. New information technology systems and a Bayesian hierarchical bivariate probit model for profiling surgeon quality at a large hospital. The Quarterly Review of Economics and Finance 2004;44:410–29.

46. Drummond, MF.; O'Brien, BJ.; Stoddart, G.; Torrance, G. Methods for the Economic Evaluation of Health Care Programmes. 2nd ed.. Oxford: Oxford University Press; 1997.

47. Mason J. The generalisability of pharmacoeconomic studies. Pharmacoeconomics 1997;11:503–514. [PubMed: 10168092]

48. Greiner W, Schoffski O, Graf VD, Schulenberg J-M. The transferability of international economic health results to national study questions. Health Economics in Prevention and Care 2000;1:94–102.

49. Bonsel GJ, Rutten FF, Uyl de Groot CA. Economic evaluation alongside cancer trials: methodological and practical aspects. European Journal of Cancer 1993;29A(Suppl 7):S10–4. [PubMed: 8312059]

50. Mason J, Drummond MF, Torrance GW. Some of the guidelines on the use of cost effectiveness league tables. BMJ 1993;306:570–572. [PubMed: 8461776]

51. Bennett CL, Armitage JL, LeSage S, Gulati SC, Armitage JO, Gorin NC. Economic analyses of clinical trials in cancer: are they helpful to policy makers? Stem Cells 1994;12(4):424–9. [PubMed: 7951008]

52. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics 1994;3:95–104. [PubMed: 8044216]

53. Haycox A. Pharmacoeconomics: integrating economic evaluation into clinical trials. Br J Clin Pharmacol 1997;43(6):559–62. [PubMed: 9205814]

54. Rizzo JDPNR. Methodological hurdles in conducting pharmacoeconomic analyses. Pharmacoeconomics 1999;15(4):339–355. [PubMed: 10537953]

55. Drummond M, Brandt A, Luce BC, Rovira J. Standardising methodologies for economic evaluation in health care. International Journal of Health Technology Assessment in Health Care 1993;9(1):26–36. [PubMed: 8423113]

56. Fayers PM, Hand DJ. Generalisation from phase III clinical trials: survival, quality of life, and health economics. Lancet 1997;350:1025–1027. [PubMed: 9329528]

57. Goeree R, Gafni A, Hannah M, Myhr T, Blackhouse G. Hospital selection for unit cost estimates in multicentre economic evaluations: Does the choice of hospitals make a difference? Pharmacoeconomics 1999;15(6):561–572. [PubMed: 10538329]

58. Carr Hill RA. The evaluation of health care. Social Science and Medicine 1985;21(4):367–75. [PubMed: 3931259]

59. Neymark N, Kiebert W, Torfs K, Davies L, Fayers P, Hillner B, et al. Methodological and statistical issues of quality of life (QoL) and economic evaluation in cancer clinical trials: report of a workshop. European Journal of Cancer 1998;34(9):1317–1333. [PubMed: 9849412]

60. O'Connell, D.; Glasziou, P.; Hill, S.; Sarunac, J.; Lowe, J.; Henry, D. Results of clinical trials and systematic trials: To whom do they apply?. In: Stevens, A.; Abrams, K.; Brazier, R.; Fitzpatrick, R.; Lilford, R., editors. The Advanced Handbook of Methods in Evidence Based Healthcare. London: Sage; 2001.

61. Caro JJ, Huybrechts KF, De Backer G, De Bacquer D, Closon MC. Are the WOSCOPS clinical and economic findings generalizable to other populations? A case study for Belgium. Acta Cardiologica 2000;55(4):239–246. [PubMed: 11041122]

62. Caro J, Klittich W, McGuire A, Ford I, Norrie J, Pettitt D, et al. The West of Scotland coronary prevention study: Economic benefit analysis of primary prevention with pravastatin. British Medical Journal 1997;315(7122):1577–1582. [PubMed: 9437275]

63. Caro JJ, Klittich W, McGuire A, Ford I, Pettitt D, Norrie J, et al. International economic analysis of primary prevention of cardiovascular disease with pravastatin in WOSCOP. European Heart Journal 1999;20(4):263–268. [PubMed: 10099920]

64. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, Macfarlane PW, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. New England Journal of Medicine 1995;333(20):1301–1307. [PubMed: 7566020]

65. McAlister FA. Commentary: relative treatment effects are consistent across the spectrum of underlying risks…usually. International Journal of Epidemiology 2002;31:76–77. [PubMed: 11914298]

66. Gail MH, Simon R. Testing for qualitative interaction between treatment effects and patien subsets. Biometrics 1985;41:361–72. [PubMed: 4027319]

67. Cook JR, Drummond MF, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. Statistics in Medicine 2003;22(12):1955–1976. [PubMed: 12802815]

68. Willke RJ, Glick HA, Polsky D, Schulman KA. Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics 1998;7:481–493. [PubMed: 9809707]

69. Snijders, TAB.; Bosker, RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage Publications; 1999.

70. Willan AR, Pinto EM, O'Brien BJ, Kaul P, Goeree R, Lynd L, et al. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. Health Economics 2005;14:327–338. [PubMed: 15685652]

71. Thompson SG, Nixon R, Grieve R. Addressing the issues that arise in multicentre cost data, with application to a multinational study. Journal of Health Economics. 2006forthcoming

72. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing Generalisability by Location in Trial-Based Cost-Effectiveness Analysis: the Use of Multilevel Models. Health Economics 2005;14(5):471–85. [PubMed: 15386662]

73. Manca, A.; Lambert, PC.; Sculpher, MJ.; Hahn, S.; Rice, N. To pool or not to pool? The use of Bayesian hierarchical modelling to analyse cost-effectiveness data collected alongside multinational trials. 5th World Congress of the International Health Economics Association; 2005; Barcelona, (Spain): iHEA; 2005.

74. Grieve R, Nixon R, Thompson SG, Normand C. Using multilevel models for assessing the variability of multinational resource use and cost data. Health Econ 2005;14(2):185–96. [PubMed: 15386660]

75. Pinto EM, Willan AR, O'Brien BJ. Cost-effectiveness analysis for multinational clinical trials. Statistics in Medicine 2005;24:1965–1982. [PubMed: 15803442]

76. Rice N, Leyland A. Multilevel models: applications to health data. Journal of Health Services Research and Policy 1996;1:154–64. [PubMed: 10180862]

77. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Group. A-LAITS. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002;21(3):371–87. [PubMed: 11813224]

78. Manca A, Lambert PC, Sculpher MJ, Rice N. Cost effectiveness analysis using data from multinational trials: The use of bivariate hierarchical modelling. 2006 [PubMed: 17641141]submitted

79. Sculpher MJ, Poole L, Cleland J, Drummond MF, Armstrong PW, Horowitz JD, et al. Low doses vs. high doses of the angiotensin converting-enzyme inhibitor lisinopril in chronic heart failure: a cost-effectiveness analysis based on the Assessment of Treatment with Lisinopril and Survival (ATLAS) study. European Journal of Heart Failure 2000;2:447–454. [PubMed: 11113723]

80. Paker M, Poole-Wilson PA, Armstrong PW, Cleland J, Horowitz JD, Massie BM, et al. Comparative effects of low and high doses of the angiotensin converting-enzyme inhibitor, lisinopril, on morbidity and mortality in chronic heart failure. Circulation 1999;100:2312–2318. [PubMed: 10587334]

81. Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. New York: Chapman & Hall/CRC; 2004.

82. Palmer S, Sculpher M, Philips Z, et al. Management of non-ST-elevation acute coronary syndromes: how cost-effective are glycoprotein IIb/IIIa antagonists in the UK National Health Service? International Journal of Cardiology 2005;100:229–240. [PubMed: 15823630]

83. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23(20):3105–24. [PubMed: 15449338]

84. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. Statistics in Medicine 2000;19:3251–74. [PubMed: 11113958]

85. Sharp SJ, Thompson SG, Altman DG. The relationship between treatment benefot an underlying risk in meta-analysis. British Medical Journal 1996;313:735–738. [PubMed: 8819447]

86. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 2002;21(11):1559–73. [PubMed: 12111920]

87. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005;365(9456):341–6. [PubMed: 15664231]

88. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-analysis. Stat Med 2004;23(11):1663–82. [PubMed: 15160401]

89. Robinson M, Ginnelly L, Sculpher M, Jones L, Riemsma, Palmer S, et al. A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists. Health technology assessment (Winchester, England) 2002;6(25):1–160. [PubMed: 12583818]

90. Sculpher MJ, Claxton K, Drummond MF, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Economics. 2006forthcoming

91. Ades AE, Claxton K, Sculpher MJ. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Economics 2006;15(4):373–81. [PubMed: 16389628]

92. Ades AE, Sculpher MJ, Sutton A, Abrams KR, Cooper NJ, Welton N, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24(1):1–19. [PubMed: 16445299]

93. Sculpher, MJ.; Claxton, K.; Akerhurst, R. It's just evaluation for decision making: recent developments in, and challenges for, cost-effectiveness research. In: Smith, PC.; Ginnelly, L.; Sculpher, MJ., editors. Health Policy and Economics: opportunities and challenges. Maidenhead, Berkshire, England: Oxford University Press; 2005.

94. Abrams KR, Cooper NJ, Sutton AJ, Sculpher MJ, Palmer SJ, Ginnelly L, et al. Populating Economic Decision Models - Bayesian Approaches to Evidence Synthesis. Applied Stochastic Models in Business and Industry. in submission

95. Cooper NJ, Abrams KR, Sutton AJ, Turner D, Lambert PC. Use of Bayesian methods for Markov modelling in cost-effectiveness analysis: An application to taxane use in advanced breast cancer. Journal of the Royal Statistical Society - Series A 2003;166(3):389–405.

96. Cooper NJ, Sutton AJ, Abrams KR. Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section. Stat Methods Med Res 2002;11(6):491–512. [PubMed: 12516986]

97. Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Econ 2004;13(3):203–26. [PubMed: 14981647]

98. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. Health Technology Assesement 2005;9(26):1–148. [PubMed: 16014203]

99. Lim E, Ali Z, Ali A, Routledge T, Edmonds L, Altman DG, et al. Indirect comparison meta-analysis of aspirin therapy after coronary surgery. BMJ 2003;327(7427):1309. [PubMed: 14656836]Erratum in: BMJ. 2004 Jan 17;328(7432):147

100. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. Control Clin Trials 2000;221(5):488–97. [PubMed: 11018565]

101. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direst and indirect evidence. British Medical Journal 2005;331:897–900. [PubMed: 16223826]

102. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. Stat Med 2003;22(19):2995–3016. [PubMed: 12973783]

103. Urdahl H, Manca A, Sculpher MJ. Assessing Generalisability in Model-Based Economic Evaluation Studies: A Structured Review in Osteoporosis. 2006 [PubMed: 17129074]submitted
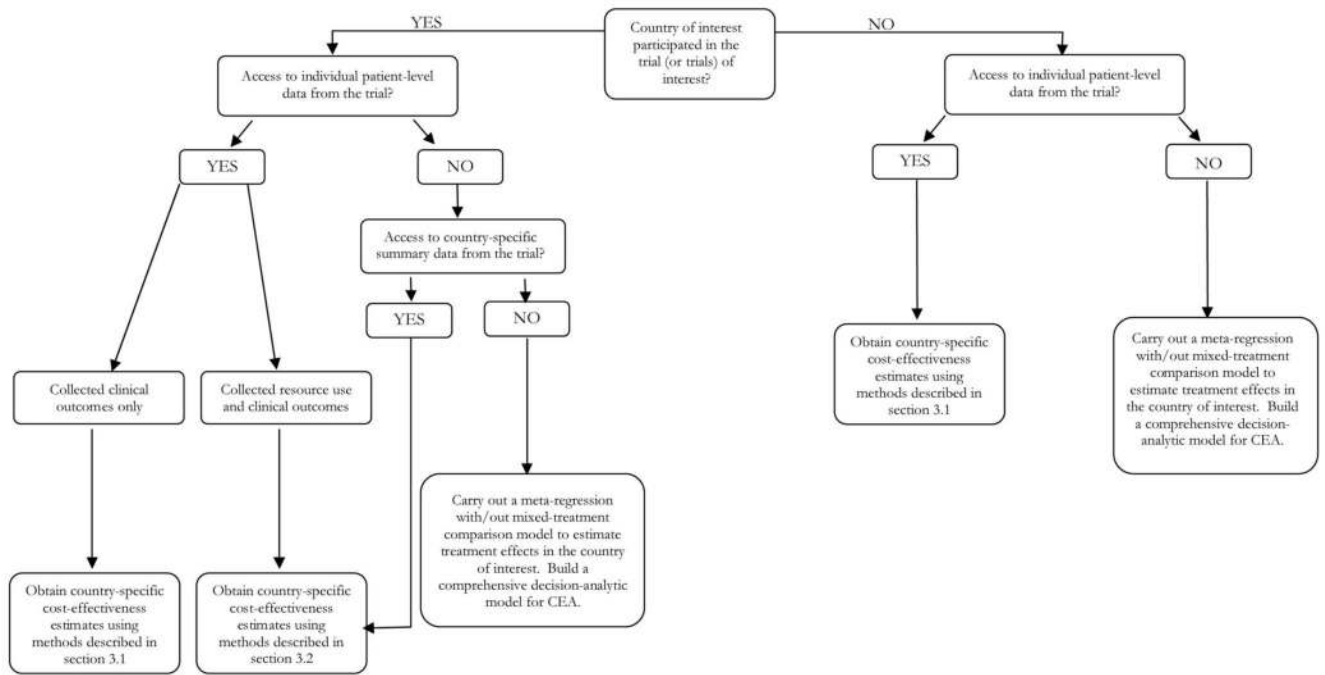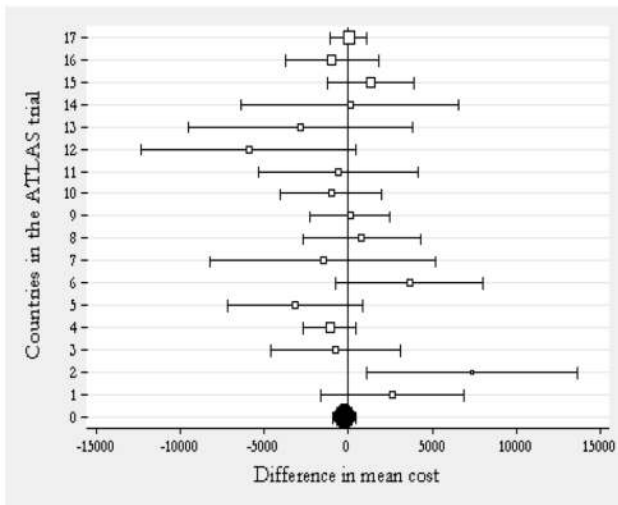
Figure 1. Algorithm to decide which methodology should be used to explore between-country differences in cost-effectiveness
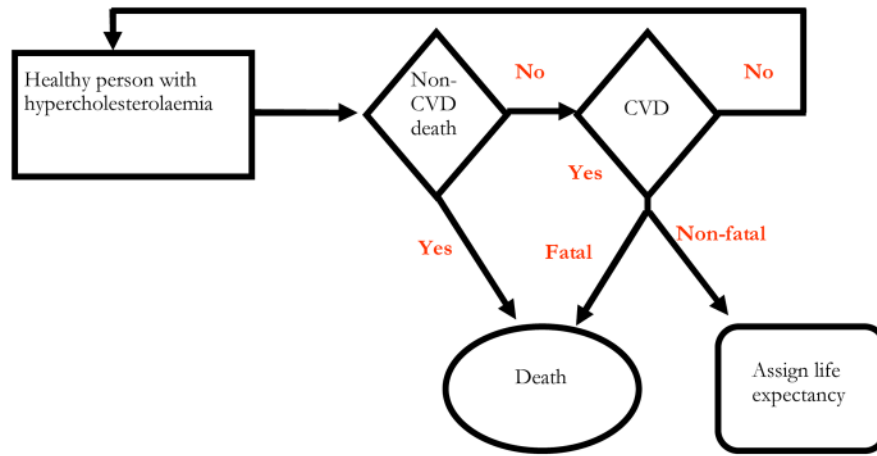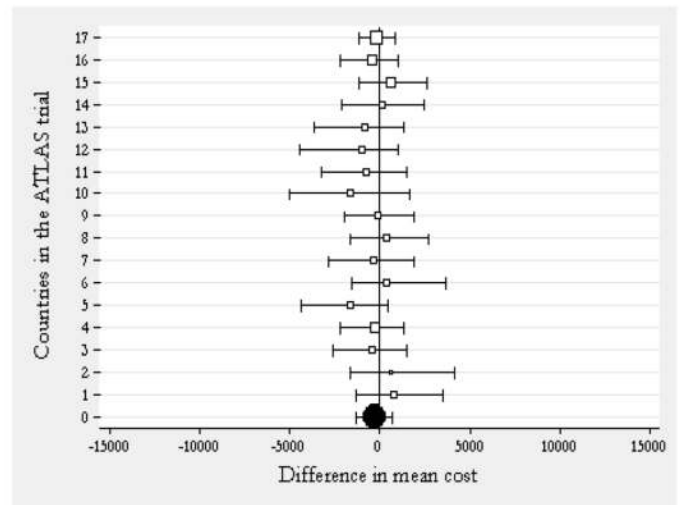
Figure 2. Simplified representation of the WOSCOP model structure for Belgium
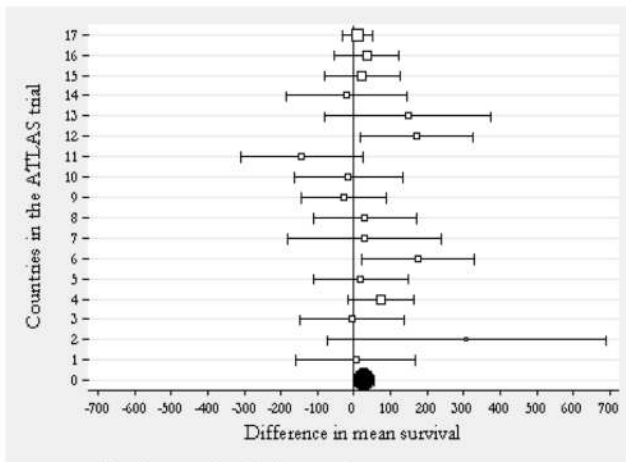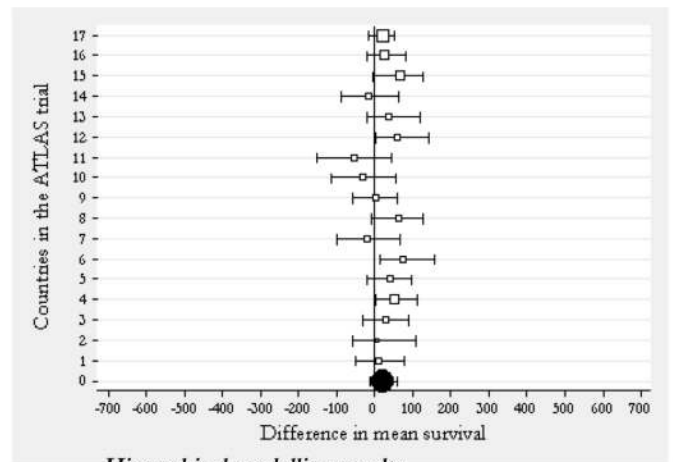**Source**: Caro *et al* Acta Cardiol. 2000; 55(4): 239-46

Figure 3. 'Pooling' and 'splitting' *versus* hierarchical modelling for multinational trials: estimating country-specific mean difference in cost
Note: The black circle indicates the trial-wide estimate. The square markers indicate the country-level mean estimates of differential costs. The size of the markers is proportional to the sample of patients recruited in each country. Finally, the horizontal bars represent the 95% confidence intervals.

*Pooling and splitting analyses*

Hierarchical modelling results

Figure 4. 'Pooling' and 'splitting' *versus* hierarchical modelling for multinational trials: estimating country-specific survival gain
Note: The black circle indicates the trial-wide estimate. The square markers indicate the country-level mean estimates of survival gains. The size of the markers is proportional to the sample of patients recruited in each country. Finally, the horizontal bars represent the 95% confidence intervals.
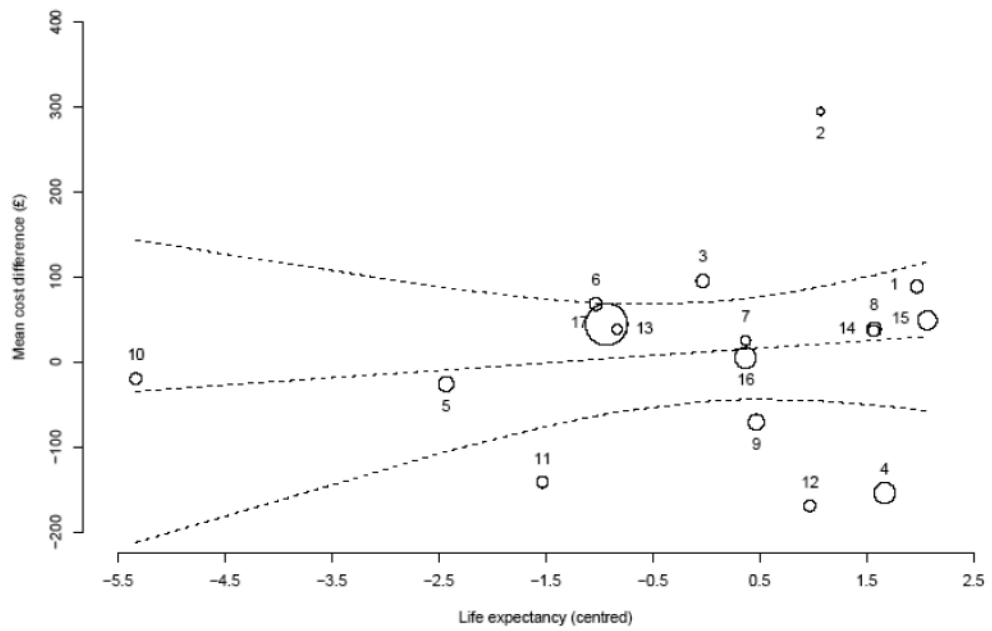
Figure 5. Relationship between country-specific differential costs and life expectancy at birth in the ATLAS trial
Note: The size of the markers is inversely proportional to the standard error of the mean differential cost estimate in each country. The dotted lines represent the mean (95% credibility limits) estimated relationship between the mean difference in costs and the mean life expectancy at birth in each country.
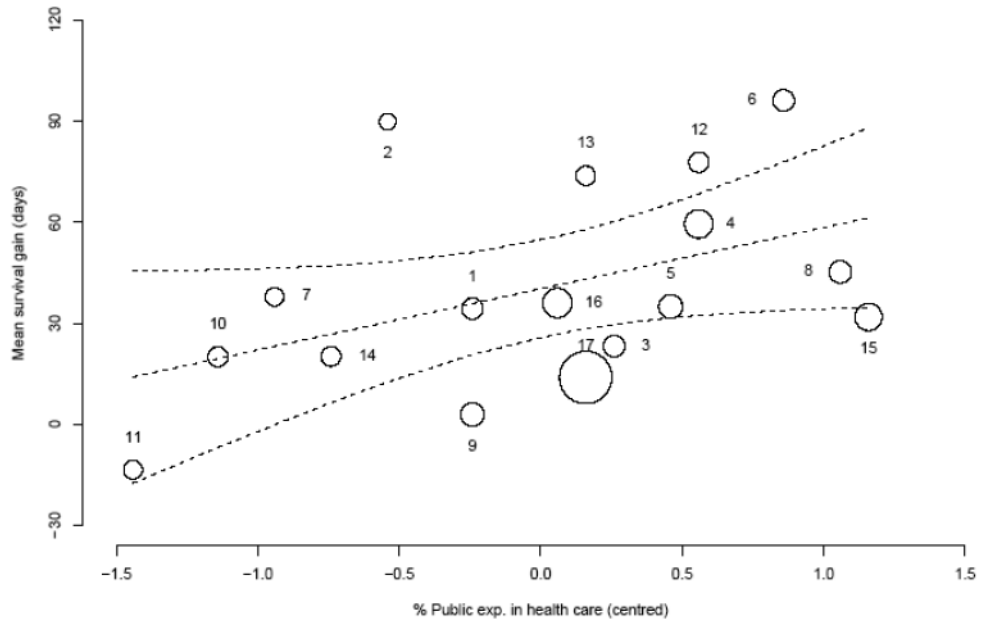
Figure 6. Relationship between country-specific survival gain and public expenditure in healthcare as a percentage of national GDP in the ATLAS trial
Note: The size of the markers is inversely proportional to the standard error of the mean survival gain estimate in each country. The dotted lines represent the mean (95% confidence limits) estimated relationship between the mean survival gain and the mean percentage of national GDP in the ATLAS trial in each country.
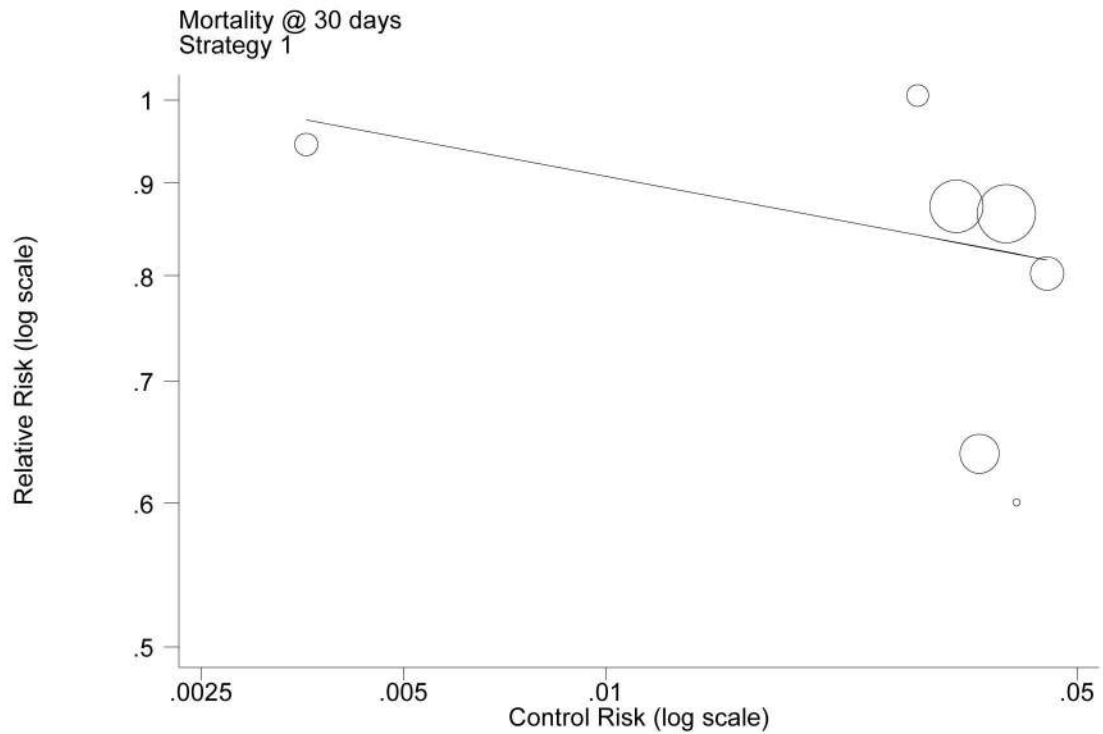
Figure 7. Example of the results of the meta-regression comparing log baseline (control group) risks with log relative risks in the GPAs model.

Note: This example relates to Strategy 1 trials and the event of mortality at 30 days. Each circle represents a trial, and the size of the markers indicates the study relative sample size. In this example, a constant (standard error) of −0.418 (0.524) and a slope (standard error) of −0.066 (0.150) were estimates. This showed that the slope was not statistically different from zero (p=0.44).

Source: Sculpher MJ, Pang F, Manca A, *et al* Health Technology Assessment 2004; 8(49): 1-206
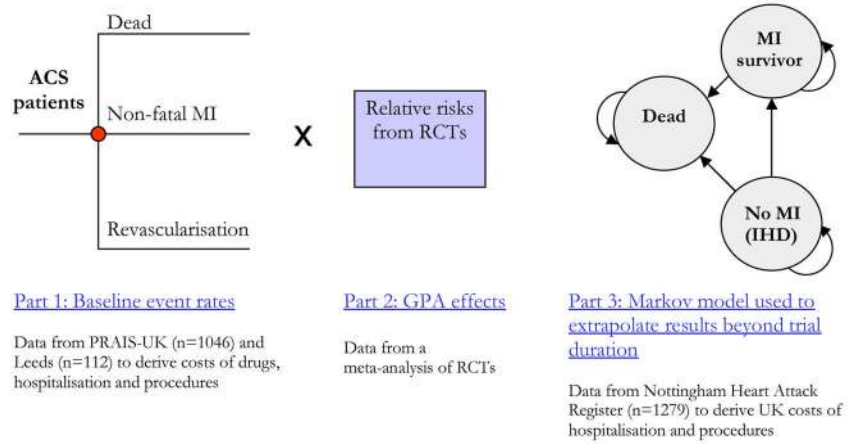
Figure 8. Structure of the model used to evaluate the cost-effectiveness of alternative uses of Glycoprotein IIb/IIIa antagonists in acute coronary syndrome in the UK
Source: Palmer *et al.* International Journal of Cardiology 2005; 100: 229-240.

Table 1

Challenges faced by Palmer *et al.* in developing and analysing a cost-effectiveness models for the use of GPAs in the UK

Limitation with trial data	Methods used to overcome limitations	Additional data source used
Use of GPAs in the trials does not reflect their range of possible uses in the UK. Baseline event rates in trials unlikely to reflect UK practice due to differences in management of ischaemic heart disease	Four strategies relating to how GPAs are being used in the UK were identified with clinical collaborators. Separate out baseline event rates in trial control groups from relative treatment effects. Apply latter to UK-specific baseline event rates taken from another sources	UK-specific baseline event rates taken from the Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK), and a specific survey of patients undergoing acute PCI at Leeds General Infirmary.
The relative treatment effect may also vary by location. In particular, the relative risk may be related to the baseline risk in a location.	Undertake a meta-regression to relate the baseline (control group) risks with those in the experimental (GPA) groups. If a clear relationship is identified, this can be used to adjust the relative risks used for a UK analysis to a level commensurate with UK baseline event rates.	
Recent changes in clinical practice in the UK include the increased use of PCI and use of clopidogrel as part of standard management. However, the UK data sources may not reflect these changes, and their implications for the cost-effectiveness of GPAs is unclear.	Use of sensitivity analysis to assess the effect on the models results of using baseline event rates from the trials instead of the UK observational study. Also modelling a strategy of using clopidogrel as a fifth strategy in the model.	
No data from the trials (UK-specific or otherwise) to extrapolate short-term clinical results to long-term QALYs.	Use of a long-term Markov model populated using UK-specific observational data.	Transition probability and longer-term resource use data taken from Nottingham Heart Attack Register (NHAR).

Source: Sculpher MJ, Pang F, Manca A, *et al* Health Technology Assessment 2004; 8(49):1-20