ABSTRACT

A growing number of jurisdictions now request economic data in support of their decision-making procedures for the pricing and/or reimbursement of health technologies. As more jurisdictions request economic data, the burden on study sponsors and researchers increases. There are many reasons why the cost-effectiveness of health technologies might vary from place to place. Therefore, this report of an ISPOR Good Practices Task Force reviews what national guidelines for economic evaluation say about transferability, discusses which elements of data could potentially vary from place to place and recommends good research practices for dealing with aspects of transferability, including strategies based on the analysis of individual patient data and based on decision-analytic modeling.
In December 2004, the ISPOR Health Science Policy Council recommended the issue “Transferability of Economic Data: When Does a Difference Make a Difference?” be considered by the Board of Directors. The Council’s recommendations concerning transferability of economic data were as follows: define key variable economic data, define guidelines for acceptance data from outside a country taking into consideration existing national guidelines. After further development of the issue by the Health Science Policy Council, the Board approved the creation of a Task Force on Good Research Practices on Transferability of Economic Data in Health Technology Assessment in May 2005, and advised that the Task Force be under the leadership of the Health Science Policy Council. Task Force leadership and reviewer groups were finalized by December 2005.

Task Force members were experienced in health economics and technology assessment and were drawn from both industry and academia. In addition, several members had experience in working with reimbursement agencies in their respective countries. The members came from Italy, the Netherlands, Poland, Switzerland, the United Kingdom and the US.

Beginning in May 2006, the task force conducted monthly and bi-monthly teleconferences to develop core assumptions and an outline before preparing a draft report. A face to face meeting was held in January 2005 to develop consensus for the draft final report. The draft final report was posted on ISPOR’s website in April 2008, and the task force’s reviewer group and ISPOR general membership was invited to submit their comments for a one-month review period. In total, 50 individuals or groups responded. All comments received were posted on the ISPOR website and presented for discussion at the Task Force Forum at ISPOR’s 13th Annual International Meeting in May 2008. Comments and feedback from the forum was considered, and when appropriate, incorporated and acknowledged in the final report.
INTRODUCTION

A growing number of jurisdictions now request economic data in support of their decision-making procedures for the pricing and/or reimbursement of health technologies. In most cases the requests for data are supported by national guidelines on the conduct of economic evaluation [1, 2] As more jurisdictions request economic data, the burden on study sponsors and researchers increases, especially as the various national guidelines may insist on the presentation of local data, or the use of specific methods.

There are many reasons why the cost-effectiveness of health technologies might vary from place to place, including the incidence and severity of the disease in question, the availability of health care resources, clinical practice patterns and relative prices [3]. The extent of variation in estimates has been shown in a review of economic evaluations of medicines undertaken in Western Europe [4]. They found that in 17 out of 27 cases the variation in the estimates of the incremental cost-effectiveness ratios could be considered to be substantial (a two-fold difference likely to change the decision to reimburse the drug). Therefore, it is reasonable for national guidelines to request that analyses be relevant to the local context.

However, the requirement that economic evaluations should use local data, or that particular methods should be used, means that analyses increasingly need to be customized for each setting. Therefore, an ISPOR Good Research Practices Task Force on Economic Data Transferability was established with the following objectives: 1) to review what national guidelines for economic evaluation say about transferability; 2) to discuss which elements of data could potentially vary from setting to setting; and 3) to recommend good research practices for dealing with aspects of transferability (including analytic strategies and guidance for considering the appropriateness of evidence from other countries).

The Task Force’s working definitions were that economic evaluations were generalizable if they applied, without adjustment, to other settings. On the other hand, data were transferable if they could be adapted to apply to other settings. Also, the generic term ‘jurisdiction’ was used to mean any setting where there is a need for local estimates of cost-effectiveness. Often this would be a country, but could also be a region
within a country, or a particular payer, such as a health plan. However, when referring to a particular study, more specific terms like ‘country’ or ‘clinical center’ are used if they help in the explanation of the study’s methods.

Finally, the Task Force’s work focused on the transferability of the data and analyses used to produce local estimates of cost-effectiveness. It was not within the remit to discuss the transferability of decision-making criteria, such as the threshold value of the incremental cost-effectiveness (ICER) deemed to be ‘acceptable’ in different jurisdictions.

**VARIATIONS IN CURRENT GUIDANCE ON THE TRANSFERABILITY OF ECONOMIC AND CLINICAL DATA FOR ECONOMIC EVALUATION**

The full results of our review of existing national methods guidelines are published elsewhere [5], but are summarized here. All the guidelines recognize that there are several issues relating to the transferability of economic data but differ in the extent to which they discuss these in detail. There is also considerable variation in the methods for addressing transferability, not all of which can be justified.

Most guidelines (16 out of 21 reviewed) recognize the potential for differences in the clinical parameters from one country to another, and that these can lead to differences in cost-effectiveness. The majority of guidelines suggest that jurisdiction-specific estimates of baseline risk should be used in cost-effectiveness studies and that treatment effect (i.e., relative risk reduction) from clinical trials might be more generalizable across settings. However, there is not unanimous agreement on this and a small number of guidelines do not make a distinction between relative and absolute risk reduction.

Less than 40% of the guidelines (8 out of 21) make any recommendations about the transferability of health state valuations or utility estimates between jurisdictions. The remaining guidelines recommend the use of utility values from the jurisdiction of interest, or at least values that are applicable to the jurisdiction of interest. One set of guidelines [4] even specifies the method by which the utility estimates should be obtained, suggesting that a generic instrument is preferred.
Most guidelines recommend that resource quantities should be reported separately from unit costs (prices) in order to increase the transparency of the analysis. All the guidelines agree that unit cost prices should be jurisdiction-specific, due to differences in relative, or absolute, price levels among jurisdictions. Some guidelines also provide sources for unit costs (e.g. an official list). Almost all guidelines require that data on resource use should be from, or adapted to, the jurisdiction of interest. However, there is no consistent guidance on how such data should be obtained (e.g. from trials or from local databases), or on how any adjustments should be made.

Therefore, it does appear that the various methods guidelines differ sufficiently in their requirements and those issues of economic data transferability need to be addressed.

**DEVELOPING GOOD RESEARCH PRACTICES FOR DEALING WITH ASPECTS OF TRANSFERABILITY**

**Overview of Possible Strategies**

The approach for dealing with aspects of transferability is likely to depend on two key factors: the decision-maker’s requirements and data availability. Many economic evaluations published in the literature have been performed with no specific decision-maker in mind. In such cases the approach for dealing with aspects of transferability is based entirely on data availability and the attributes of the various analytic methods.

However, the main focus of this report is on those situations where a particular decision-maker can be specified. Here the decision-maker’s requirements may also help determine the analytic approach. For example, in choosing comparators for a given evaluation in England and Wales, the National Institute for Health and Clinical Excellence (NICE) [6] states that ‘all relevant alternatives should be compared’. This means that an analysis based on a single trial would be unlikely to be sufficient, unless the trial compared all the relevant treatment options. Therefore, an approach based on modeling is likely to be preferred. On
the other hand, in the Netherlands the guidelines for pharmaco-economics [7] stipulate that the comparator should be the standard treatment as is mentioned in practice guidelines and for which effectiveness is demonstrated, and in Germany the Institute for Quality and Efficiency in Health Care (IQWiG) favours estimates of effectiveness obtained directly from clinical trials. Therefore, in Germany in particular, decision-makers are more likely to favor an approach based on the analysis of individual patient data from one or more randomized controlled trials (RCTs), rather than a modeling approach involving a synthesis of estimates from a range of sources [8].

The availability of data is another important factor determining transferability of cost-effectiveness results. Manca and Willan [9] have proposed an algorithm based on the availability of data. For example, if the jurisdiction of interest has participated in a multinational clinical trial in which data on resource use and/or cost have been collected, the preferred strategy would be to analyze the individual patient data from that trial. If this were not the case, a modeling approach would usually be required, using as much clinical, resource use and cost data as possible from the jurisdiction of interest. This report takes the argument further and considers analytic strategies based on the analysis of individual patient data gathered in clinical trials, and strategies based on decision-analytic modeling. The various strategies are initially discussed separately but, later in the paper the synergies between them are also explored.

Defining the Decision Problem

For any evaluation it is imperative to have a clear and appropriately specified decision problem, defined in advance. In the context of decisions about a new technology, this would require a statement of the patient population and sub-populations of interest and the comparators to the new technology. In this context, 'comparator' will be a therapeutic option or strategy which could be used instead of the new technology. It is quite feasible for an appropriately specified decision problem to include a large number of alternative options.

It is possible that an appropriately specified decision problem relating to a new technology would vary between jurisdictions. One reason, in the case of pharmaceuticals, is that licenses can differ between
jurisdictions and this can affect the clinical applications of the new product, or those of the comparators. A second reason relates to differences in routine clinical practice, where health technologies used in one jurisdiction are simply not used in another, and this can affect the choice of comparators. Differences in clinical practice may also influence the specification of treatment strategies. For example, in defining appropriate comparators for aspirin plus clopidogrel in the management of acute coronary syndrome, it may be appropriate to specify aspirin as the only alternative in the UK, but in Italy it may be considered appropriate also to include ticlopidine. It should be noted, however, that just because particular health technologies are not currently used in a particular jurisdiction does not mean they should never be used, and a completely specified decision problem may include such options to fully inform decision making.

To guide the reader through this report we propose four steps that consider data availability and determine whether simple or more elaborate methods for adjusting cost-effectiveness information to a particular jurisdiction are needed. These four steps are outlined in Box 1. First, in some situations a cost-effectiveness study may already exist for another jurisdiction, but generally will not be directly applicable to the jurisdiction of interest. Therefore, prior to considering developing a new economic evaluation for the jurisdiction of interest, it will be useful for the analyst to assess whether the results of the existing study can be adapted in order to make them transferable. Therefore, the question arises as to whether cost-effectiveness estimates can be transferred to another specified jurisdiction and if not, what adjustments of these estimates, or even the cost-effectiveness study, are possible and/or necessary?

It is possible to identify two situations where cost-effectiveness estimates for another jurisdiction are available and where the level of transferability could be assessed:

*The cost-effectiveness results are not transferable because the starting points of the study are irrelevant to the decision maker*

If either the experimental technology or the comparator(s) is (are) not relevant in the jurisdiction of interest, the cost-effectiveness results are irrelevant. Additionally, if the methodological quality of the cost-effectiveness study does not meet the local standards, transfer of the cost-effectiveness estimate is not
valid. These aspects are so-called ‘general knockout criteria’ (i.e. factors that preclude transferability) [10]. Furthermore, the patient population that is relevant to the use of the health care technology may be different between jurisdictions and therefore the cost-effectiveness information is not transferable.

The cost-effectiveness results are only transferable after adjustment for differences in treatment patterns, in unit costs or other aspects

For instance, the setting where a patient is treated (e.g. in a hospital setting or by family physicians) can differ between jurisdictions and this can make a difference to the results. Therefore, practice variations between jurisdictions may make the transferability of cost-effectiveness estimates impossible without adjustment [10]. Furthermore, differences in unit costs between jurisdictions make recalculation necessary and for this reason either jurisdiction-specific cost data might be required, or a more simple adjustment may be made, for instance based on purchasing power parities [10]. Also, the definition of the time horizon of the analysis may be important, as the prerequisites for this might differ between jurisdictions. Other characteristics of existing cost-effectiveness information, for which adjustment may be required, concern the perspective used and the impact this has on the approach for estimating health care costs (e.g. charges, fees, real costs), the discount rates applied, and whether productivity or time costs are included and how they are valued (e.g. human capital or friction cost method). These are called ‘specific knockout criteria’ (i.e. aspects of the analysis that would have to be addressed before transfer of estimates is possible) [10].

If, from the answers to the questions in the first three steps in Box 1, it becomes clear that the available cost-effectiveness information is not directly transferable through a simple adjustment procedure, Step 4 indicates that there are two options. If patient data from multi-location studies are available including the jurisdiction of interest, there are analytic techniques to calculate adjusted cost-effectiveness estimates for the jurisdiction. These methods differ in level of sophistication. However, if the trial was undertaken wholly outside the jurisdiction of interest, the transferability issues will usually be addressed through parametrization of a decision analytic model, although it may be possible to make some inferences from
the adjusted cost-effectiveness estimates for jurisdictions that did participate in the trial, particularly if one or more of them was thought to be similar to the jurisdiction of interest.

Analyzing Individual Patient Data from Multi-location Studies

Most large clinical trials enroll patients from several jurisdictions. When data on resource use (and/or cost and/or utilities) are gathered alongside the trial, the analysis of individual patient data can be the basis for exploring issues of economic data transferability.

Analytic methods to address the many complexities inherent in multi-location studies have rapidly evolved in recent years. In this section, the current literature is summarized, identifying the advantages and limitations of various approaches to addressing issues of transferability, and making recommendations about which approaches should be avoided and which should be considered for application. Several areas where there are likely to be further methodological advances are also identified.

With regard to transferability, analytic approaches address two sets of objectives. The first is to evaluate whether there is evidence of heterogeneity in patterns of resource use, costs, survival and/or utilities exist among jurisdictions included in the trial and to explore potential sources of heterogeneity. The second objective is to obtain estimates of incremental resource use, cost and/or cost-effectiveness that are appropriate for decision making within particular jurisdictions that may or may not have been included in the trial. Analytic methods for both objectives will vary depending on the types of data available to the analyst. An algorithm developed by Manca and Willan\(^9\) differentiates strategies according to whether the trial enrolled patients from the jurisdiction of interest and whether individual patient-level data (IPD) from the trial are available to the analyst. In this section, we assume the analyst has access to patient-level data from the multi-location trial.

Three general statistical methods comprise the literature: 1) detection of heterogeneity; 2) fixed effects models; and 3) multilevel, or hierarchical, models.
Detection of Heterogeneity

A prerequisite step in identifying presence of heterogeneity prior to statistical modeling is the examination of descriptive statistics to identify key differences between jurisdictions in incremental costs and effects. In many cases, it may be necessary to group patients across jurisdictions with small sample sizes to avoid jeopardizing patient confidentiality. Exploration of potential heterogeneity using point estimates and confidence intervals is encouraged, but should not represent the sole basis for conclusions about similarities or differences between jurisdictions.

The identification of heterogeneity in economic measures (e.g., incremental resource use, cost, and cost-effectiveness) across jurisdictions participating in a multi-location trial can be handled in the same statistical manner as the identification of heterogeneity in the clinical measures from the trial [11]. One approach involves tests of qualitative and quantitative interactions developed by Gail and Simon [12]. A quantitative interaction represents a situation where the direction of the treatment effect is consistent but not the magnitude (i.e., treatment decreases resource use in all jurisdictions, but to a greater extent in some and to a lesser extent in others); a qualitative interaction represents a situation where the direction of the treatment effect is different (i.e., treatment increases resource use in some jurisdictions and decreases resource use in others) (Box 2) [12]. Although these tests can be applied separately to a health technology’s impact on resource utilization, costs and effects, results from separate tests do not necessarily provide insight as to whether jurisdictions differ (in either direction or magnitude) with respect to cost-effectiveness. Instead, the appropriate application for cost-effectiveness analysis is to employ the statistical tests directly on the incremental net benefits (health or monetary) or cost-effectiveness ratio. If there is no evidence of heterogeneity, and if the test is powerful enough to rule out economically meaningful differences, it is recommended that the pooled or fixed treatment effect across all jurisdictions can be used to summarize the study’s results. However, if the test reveals evidence of heterogeneity, a pooled estimate of cost-effectiveness should not be used to represent the trial-wide results and thus is not transferable across jurisdictions.
There are two main disadvantages of this approach. One is that tests of homogeneity are typically underpowered, especially when relatively small numbers of patients are enrolled in individual jurisdictions. This limitation can be partially addressed by combining patients from jurisdictions with similar practice characteristics (e.g., types of facilities and providers, or characteristics of reimbursement systems). Unless there is a theoretical basis for pooling jurisdictions, there is no empirical basis about how such pooling should be accomplished. *Ex post*, one might pool jurisdictions with the most similar results in an attempt to maximize the opportunity to identify any evidence of heterogeneity that may exist. Limited power can also be addressed by using larger $\alpha$ levels (e.g., 0.1 or 0.2). An additional disadvantage to this approach is that it can only be used to evaluate whether there is evidence of heterogeneity; it does not offer a natural extension to generate jurisdiction-specific estimates of cost-effectiveness.

**Fixed Effect Models**

The simplest way analysts have accounted for potential differences among jurisdictions has been to include a set of fixed effects for jurisdictions that participated in the trial. However, the inclusion of these fixed effects was typically to ‘control for’ differences across jurisdictions, rather than to explore evidence of heterogeneity or to produce jurisdiction-specific results. Thus, this approach is not adequate for addressing the transferability issue that is the focus of this paper.

Willke and colleagues [13] extended the use of fixed effects models to separate the direct effect of a study treatment on costs, to include changes in resource use independent of the patient’s clinical outcome and the indirect effect of a study treatment on costs through changes in the patient’s clinical status. Through the use of multiple treatment-by-country and country-by-outcome interaction terms, the approach estimates country-specific direct and indirect effects of study treatment on costs. The advantage of this approach is that it uses patient-level data and standard statistical procedures to estimate country-specific differences of both costs and effects. It also enables sensitivity analyses, particularly with respect to country-specific treatment outcomes.
Koopmanschap et al. [14] also applied a regression approach to adjust estimates of resource use for differences observed across countries and evaluated costs as though all patients were enrolled from a single country. Although these regression-based approaches may be more familiar to most analysts, the application of fixed effect regression models do not account for the inherently hierarchical data structure in multi-location clinical trials. Failure to incorporate the clustering that can exist within each jurisdiction will result in an overestimation of precision, leading to confidence intervals that are too narrow.

**Multilevel Models**

Most of the recent literature has focused on the application of multilevel (or hierarchical or random effects) models to economic information (i.e. costs, effectiveness, incremental net benefit (INB) and the incremental cost-effectiveness ratio) derived from multi-location trials [15-22]. These models may not be necessary when just a few jurisdictions (i.e., less than 4 or 5) are represented in the trial. However, one of their advantages over fixed effects models and tests for detecting heterogeneity is that they can appropriately handle the hierarchical nature of the data which manifests itself as a lack of independence of the errors between the observations. This hierarchical structure occurs because of within jurisdiction similarities in clinical practice patterns, price weights for health care resources, lifestyle and health behaviors [23], and other factors. When used to analyse data from multi-location studies involving study sites in several jurisdictions, these models also allow lower level clustering (e.g. study sites or clinical centers, physicians), as well as heterogeneity in treatment effect across countries, to be incorporated in the analysis. By extension, they provide a formal means of estimating jurisdiction-specific measures of cost-effectiveness through the calculation of random intercepts and random slopes, also described as empirical Bayesian shrinkage estimation [19].

By borrowing information from the pooled estimates, shrinkage estimation yields smaller standard errors for jurisdiction-specific measures of costs and cost-effectiveness relative to the standard errors derived from each jurisdiction’s data alone [19-21]. Each jurisdiction’s estimate of costs or cost-effectiveness is also shrunken toward the summary estimate by ‘borrowing strength from across jurisdictions’ [20]. The
extent to which this shrinkage affects each jurisdiction’s estimate is dependent on the variability among the jurisdiction-specific estimates in the trial (more between-jurisdiction variability, less shrinkage), the variability within a jurisdiction (more within-jurisdiction variability, more shrinkage) and the number of trial participants from the jurisdiction (more participants, less shrinkage) [15].

The impact of within versus between variability and the number of patients on the degree of shrinkage has a direct parallel to the statistical assessment of heterogeneity and the use of the pooled or fixed effect estimate. The greater the within-jurisdiction variability relative to the between-jurisdiction variability, the greater the shrinkage and the lower the likelihood that one will detect heterogeneity. Without evidence of heterogeneity, the pooled effect may be used to represent the study’s results. Conversely, the greater the number of trial participants, the less the shrinkage and the greater the power to detect heterogeneity. With evidence of heterogeneity, fixed effects or shrunken estimates may be preferred for reporting the study’s results. In essence, the pooled estimates can be viewed as fully shrunk, and the fixed effect estimates can be viewed as not shrunk toward the summary measure. Shrinkage estimation allows for a partial shift towards the pooled estimate rather than assuming it is all or nothing.

Multilevel models can also be used to derive a pooled, random effects estimate across all jurisdictions, equivalent to a random effects summary estimate from meta-analysis [19, 21]. When applying multilevel models, analysts must make two sets of assumptions. One pertains to the estimates at the level of an individual jurisdiction and one pertains to the collection of jurisdiction-specific estimates. At the country level, most analysts have assumed a normal distribution with a mean and variance specific to each country [19]. Others have extended the models by employing gamma distributions and multiplicative effects, and have shown that these specifications are preferable for cost data [22]. In regard to the collection of country-specific, random effects estimates, most applications of multilevel models have been based on the assumptions that the parameters are drawn from a common distribution, or are ‘exchangeable’, and are represented by a common variance across all countries. The term ‘exchangeable’ means that there are no other a priori reasons why one jurisdiction may have more or less favorable measures of costs or cost-effectiveness than another [24]. Making an a priori assumption that
one does not expect differences in jurisdiction-specific measures of cost-effectiveness may be unreasonable when the question we are trying to answer is whether or not such differences exist.

Some authors have tried to overcome this problem by including center-level and/or country-level covariates, thought to correlate with measures of costs or cost-effectiveness, which allows one to make an assumption about exchangeability through conditional independence [25]. For example, the inclusion of these higher-level covariates that are thought to distinguish between centers or countries with worse or better measures of cost or cost-effectiveness allows for shrinkage towards two (or more) separate pooled (i.e., local) means. Currently, evidence is lacking as to what types of higher-level covariates may be useful in this regard, but candidate variables may include those that are indicative of macroeconomic characteristics (e.g. gross national product), capacity constraints (e.g. limited availability of intensive care beds), economic incentives (e.g. pressure to minimize length of stay), or financing characteristics (e.g. global budgets vs. fee-for-service reimbursement) associated with various practice sites or health systems. There is also little agreement about tests that should be used to evaluate the appropriateness of shrinkage towards one, two, or more such local means. Finally, the inclusion of higher-level variables also allows estimation of cost-effectiveness for centers or countries that did not participate in the trial. This approach assumes that the analyst has identified the appropriate set of higher-level covariates for the exchangeability assumption to hold, and that the characteristics of the country of interest are represented appropriately by countries participating in the trial.

**Displaying the Results**

Findings from fixed effect, pooled effect and random effects can be displayed as forest plots. To illustrate this, we reproduce two forest plots based on an analysis of the ATLAS trial [21]. The forest plot on the left-hand side of Box 3 represents the fixed effects (means and confidence intervals) for individual countries and the pooled effect across the 17 countries participating in the trial. The plot on the right-hand side represents the shrunken estimates for each of the countries along with a pooled, random effects estimate across all countries, equivalent to a random effects summary estimate from a meta-analysis. As expected, countries shrink more towards the pooled estimate if they have 1) greater variability in the fixed
effects analysis (as identified by the width of the confidence interval), or 2) smaller sample sizes (as identified by the relative size of the box for the point estimate).

The application of multilevel modeling to data from multi-location studies is continuing to evolve [21, 26]. Therefore, it would be premature to comment on best practices or on situations when multilevel modeling may, or may not, be advantageous over simpler analytic approaches.

**Summary**

Where IPD are available, all three of these methods can be used to identify heterogeneous treatment effects on measures of cost and cost-effectiveness across jurisdictions. However, they vary in regard to their utility in exploring jurisdiction-level and patient-level variables that may be associated with heterogeneity. With regard to developing jurisdiction-specific estimates, the fixed effect model and the multilevel models have the advantage that such estimates are natural extensions of these analytic frameworks. However, each of these approaches assumes that patient-level estimates of cost and effect are measured quantities. In the majority of cost-effectiveness analyses, particularly those that take a lifetime perspective, these quantities are estimated through some type of modeling process. Thus, these approaches may not be directly applicable. In the future, we may expect to see further methodological advances aimed at adapting the multilevel modeling framework to other analytic strategies to extrapolate estimates beyond the trial period [9].

**Addressing Transferability Issues in Decision-Analytic Models**

**Introduction**

Decision analytic models are increasingly being used as vehicles for economic evaluation. One of the key features of decision models is that they provide a framework within which evidence from a range of sources can be assembled. This could, for example, relate to treatment effectiveness data from a meta-
When considering undertaking a cost-effectiveness analysis to inform decision making in a particular jurisdiction, there are a number of situations where decision analytic modeling might be considered to be the preferred basis for cost-effectiveness analysis [27-28] including:

- When the available trials were undertaken wholly outside the jurisdiction of interest and one or more components of evidence cannot be considered generalizable to that jurisdiction.
- When there is more than one relevant source of evidence relating to any aspect of the analysis – e.g. treatment effect, baseline risks, resource use or quality of life.
- Where a trial exists, but some other aspect of the economic evaluation is not consistent with the design of that trial. Obvious examples would be that the period of follow-up in the trial is not the same as the appropriate time horizon for the economic analysis, or the options being compared in the trial represent only a sub-set of those considered relevant in the country of interest.
- A decision model has been developed (using a range of evidence sources) for another jurisdiction and might be adapted to support decision making in another.

Determining Model Structure

A model's structure relates to how it specifies the natural history of a disease and the effects of a set of interventions. Depending on the type of model selected (e.g. a state transition model or a decision tree), the structure embodies the choice of states/pathways and how they interact. In general terms, a model's structure will be determined by the clinical course of a disease (its 'natural history') and the mode of action of the alternative interventions.
Again, there may be legitimate differences between jurisdictions in model structure. This is most likely to relate to differences in the specified decision problem. If a specific comparator is relevant to one jurisdiction but not to another, this may affect model structure if it needs to reflect the particular mode of action or side effects of the comparator. Similarly, if the relevant patient population differs between jurisdictions, then the natural history of the disease may also differ in a way that requires different structural assumptions.

**Parameter Estimation**

Perhaps the greatest challenge in tailoring a cost-effectiveness model to the requirements of a particular jurisdiction relates to parameter estimation; that is, using evidence to estimate the range of inputs necessary for the model. In many cases evidence for a model will be sought from a range of sources – for example, a meta-analysis of clinical trials relating to the alternative options being evaluated. However, evidence is generated globally and, at least for some parameter estimates, the relevance of evidence generated outside of the jurisdiction of interest is open to doubt. There are numerous studies showing that ‘economic’ measures vary considerably between jurisdictions – e.g. resource use and unit costs (see Sculpher et al. [3] for a review). In addition, there are studies showing that a treatment’s effectiveness can also vary [29].

In addition, it is possible that population health state valuations might vary among jurisdictions, although the evidence on this is mixed. Jurisdiction-specific value sets are now available for some of the generic instruments, such as the EQ-5D, although these have been estimated using different methods (e.g. visual analogue scale and time trade-off) [30]. Some analysts argue that differences between the value sets cannot be neglected and that transferring utility scores across jurisdictions might be questionable [31-33]. Others argue that when measurement methods to derive utilities are truly replicated, there do not seem to be substantial differences across jurisdictions [34]. Certainly, in the context of most economic evaluations, the method for estimating health state valuations is as important as their jurisdiction of origin.
The challenge facing analysts is to be able to estimate parameters in such a way as to use as much of
the evidence as possible, but to reflect variation between jurisdictions. Where IPD are available, but the
jurisdiction of interest did not participate in the trial, Manca and Willan [9], argue that the best approach is
to develop an ‘events-based model’, built around the generalizable features of the disease or patients’
prognosis, and use the IPD from the trial to estimate the likelihood of occurrence of the clinical events of
interest which are expected to have an impact on resource use and/or health-related QoL. An increasingly
common approach is to apply the trial-wide relative risk reduction in the events of interest observed in the
trial (e.g. relative reduction in risk of deaths, myocardial infarction, adverse effects) to the baseline risk in
the jurisdiction of interest, unless there was good reason to believe that the relative risk reduction from
any particular single jurisdiction was more relevant. Of course, the resource use and costs of the events,
and perhaps the values for health states, would come from the jurisdiction of interest. An important
challenge in taking this approach is to locate jurisdiction-specific baseline risk data. Furthermore, in some
clinical areas (e.g. cardiovascular disease and osteoporosis), the use of statistical models to predict event
risks as a function of patients’ characteristics are widely used within cost-effectiveness analyses, but
these ‘risk algorithms’ themselves may be jurisdiction-specific (e.g. the Framingham Heart Study
Cardiovascular risk equations which were developed in the USA). In principle, it is possible these
algorithms may be more generalizable to other jurisdictions than the data that underlie them because they
can be applied to health care systems with different patient case-mix. However, the relevance of the
clinical practice reflected in the data used to estimate the algorithms would always need to be carefully
considered, and this will depend on both the health systems in which the data were collected and the time
at which they were gathered.

A range of approaches have been used in modeling studies as outlined in Box 4. These include using
baseline event risks from the jurisdiction of interest and a meta-analysis of international randomized trials,
on the assumption that the latter can be assumed generalizable; the estimation of clinical and economic
parameters for a model using the only trial from the jurisdiction of interest and a secondary analysis
where relative treatment effects are based on a meta-analysis of all international trials; and the use of
data from a multi-national trial with regression modeling to estimate model parameters for the jurisdiction
of interest.
The need for pharmacoeconomics studies is now emerging in several countries in Central and Eastern Europe, Asia, Latin America and Southern Africa. While it is beginning to change, countries in these regions have not historically been included in those large (Phase IIIa) multinational trials that collect individual patient data on resource use or cost. Furthermore, in countries where the market for pharmaceuticals is small, there is typically less interest in enrolling patients in Phase IIIb/IV clinical trials. In such cases jurisdiction-specific data on clinical effectiveness of the new therapy would not be available. To add to the problems, routine data sources on resource use and costs may not be very well developed in the jurisdictions concerned.

In this situation one or other of the approaches outlined above can be applied. If one or more cost-effectiveness studies already exist for other jurisdictions, minor adaptations to the jurisdiction of interest, such as adjusting for local unit costs, may be possible, based on the approach outlined by Welte et al. [10]. However, owing to the differences between the jurisdictions in these regions and those for which most pharmacoeconomic studies are conducted (ie North America and Western Europe), it is likely that the general or specific knockout criteria will apply, such as the lack of availability of a given comparator therapy or substantial differences in treatment patterns.

Therefore, it is much more likely that major adaptations would be necessary (e.g. changing components of the model), or that a new decision analytic model would have to be built, as outlined above. The modeling effort would have to rely on data published in the literature, for the jurisdiction of interest or other jurisdictions. Manca and Willan [9] outline the major challenges, which include relating the model to clinical practice in the jurisdiction of interest, assessing whether the baseline event rates in the trials conducted elsewhere are relevant to the jurisdiction of interest, assessing whether the relative risk reductions estimated from the trials are related to baseline risk and incorporating jurisdiction-specific data on resource use and cost (for a good example of how these issues can be tackled, see the study by Palmer et al. [35]).
Whether or not some IPD are available, it is clear that a major barrier to conducting economic evaluations in some jurisdictions is the lack of local data. Therefore, in order to address issues of transferability, investments need to be made in the collection of epidemiological and demographic data, plus data on clinical practice patterns, resource use, costs and health state valuations.

Smaller jurisdictions with similar health care systems and clinical practice patterns (e.g. the Baltic countries) may be able to develop partnerships to develop relevant regional databases and registries. Over time it may be possible to assess whether or not there are natural groupings of similar jurisdictions, which may increase our confidence in transferring cost-effectiveness results from one jurisdiction to another.

**Analysis**

The final stage of decision modeling relates to the use of the results of the model as a direct input into jurisdiction-specific decision making. It is clear from international methods guidelines that different jurisdictions require different analytical methods. For example, some want full probabilistic sensitivity analysis, whereas some require only one-way sensitivity analysis [2]. Some of these variations seem reasonable – for example, the discount rate on future costs and benefits ought to reflect local economic conditions. However, many of the other differences between decision making agencies in their preferred analytic methods seem less justifiable [36]. These include the principles for selecting comparators, the choice of health outcome measure, the principles for determining time horizon and the methods for characterizing uncertainty. Nonetheless, for those submitting analyses to decision making authorities, their defined analytical requirements cannot easily be ignored. However, the analyst should always consider supplementing these ‘required’ analyses with additional ones, including methods that are considered more scientifically appropriate.

**Trials versus models: the false dichotomy**
Largely for presentational reasons, this paper has separated two analytical paradigms in economic evaluations – analysis of IPD data from (usually randomized) studies and decision analytic modeling. However, this is increasingly a false dichotomy, as more studies effectively use a combination of these methods. Decision analytic models are widely used to support decision making and provide a powerful framework within which to incorporate a full range of evidence and to assess the importance of structural assumptions and particular elements of evidence. However, trials provide key evidence for these models, not just on (intended and unintended) treatment effects, but also parameters such as baseline risk, resource use and health-related quality of life. Furthermore, access to IPD (as opposed to summary results) from trials allows statistical techniques to be used to provide more suitable parameter estimates for models. Advantages of access to IPD for this purpose include the ability to control for covariates, sub-group analysis and assessment of time trends in, for example, baseline risks. The use of meta-analysis based on IPD from several trials has many advantages including the use of covariate adjustment. In the context of this paper, the methods described in Section 3.2 to analyze potential heterogeneity in cost-effectiveness between jurisdictions in trials, such as fixed and random effects regression methods, can also be used in estimating specific parameters for decision models (e.g. the costs of particular events). These methods can be used with data from a single trial or a synthesis of IPD. An example of where such methods have been used was described in Box 4.

CONCLUSIONS AND RECOMMENDATIONS

The work of the Task Force has confirmed that there are several important methodological and practical issues surrounding the transferability of economic data. There is some evidence of the variability in the cost-effectiveness of health technologies between locations. Also, many international guidelines for economic evaluation make references to problems concerning economic data transferability and include requirements for jurisdiction-specific data or methods. Against this background, the Task Force makes the following recommendations.

Developing Guidelines for Economic Evaluation
Those developing guidelines for economic evaluation should fully justify the need for local data or methods, since this increases the burden on those undertaking studies in multiple jurisdictions.

**Interpreting Existing Studies**

Where a study already exists that is relevant to the decision in the jurisdiction of interest, consideration should be given (using criteria outlined in this report) to whether it can be used with simple adaptation (e.g. substituting prices only). Expert opinion and the existing literature can be used to assess whether the setting for the study is sufficiently similar to the jurisdiction of interest.

**Analysis of IPD from Multinational Studies**

Prior to statistical modeling, simple descriptive statistics should be used to examine potential differences among jurisdictions in incremental costs and effects. Jurisdiction-specific point estimates, with confidence intervals, should be reported.

Heterogeneity should be explored using some form of statistical analysis (e.g. test of interaction, multivariate regression).

The level of sophistication of the subsequent statistical modeling (i.e. fixed effects versus random effects) should be guided by the following criteria 1) number of jurisdictions (e.g. countries, clinical centers); 2) exchangeability or non-exchangeability of data; and 3) the availability of covariates (e.g. at center and country level). With more jurisdictions, partial exchangeability of data and greater availability of covariates, hierarchical modeling is to be preferred.

**Addressing Transferability through Decision-Analytic Models**
Analysts should carefully consider which parameters need to be jurisdiction-specific, wherever possible justifying assumptions empirically. Current evidence suggests that prices and, in some instances, baseline risk probably need to be jurisdiction specific, whereas treatment effect/relative risk reduction, may be more generalizable. However, where possible this should be demonstrated. The evidence on the generalizability of clinical practice patterns (resource use) and health state valuations (utilities) is mixed and therefore the use of data from other jurisdictions needs to be justified.

If trial-based data are available, analysts should use the methods typically applied in the analysis of individual patient data to estimate jurisdiction-specific model parameters.

Analysts should use scenario analysis (a form of multi-way sensitivity analysis) to explore the implication of different assumptions about economic data transferability.

**Undertaking Further Research into Economic Data Transferability**

More research should be undertaken into those sources of local differences that affect economic data transferability. This would help justify jurisdiction-specific data requirements and inform the selection of jurisdiction-level covariates in statistical models.

Analysts should evaluate whether certain smaller countries or other jurisdictions are sufficiently similar (in incremental costs and effects of new health technologies) such that they can be grouped for the purposes of analysis of economic data transferability.

Analysts should consider issues of economic data transferability when designing multinational trials that include economic data capture. This includes the recruitment of a representative sample of clinical centers in each jurisdiction and an overall balance of patient enrollment in the jurisdictions included in the study. It also includes the collection of data on jurisdiction and center-level covariates for subsequent statistical modeling.
There should be more investment in data collection for those parameters that are thought to differ most from place to place. This would include the establishment of observational studies to estimate baseline risk and the development of local cost databases (although true of all jurisdictions, this recommendation is particularly pertinent to smaller countries).

ACKNOWLEDGMENTS

We are grateful to ISPOR for financial support and to the many ISPOR members who commented on an earlier version of this report: Michael Adena, Renée Arnold, Muhammad Ashar Malik, Federico Augustovski, Sixten Borg, Julia Bottomley, Lance Brannman, Tom Burke, Anita Burrell, Isabelle Chabot, Glenn Davies, Joseph DiCesare, Pero Draganic, Janeen DuChane, Isabelle Durand-Zaleski, Ancilla Fernandes, Anna Filonenko, Holger Gothe, Carlos Gouveia Pinto, Wolfgang Greiner, Gijs Hubben, Margaret Hux, Anna Karin, Peter Kaskel, Bengt Liljas, Andrea Manca, Doreen McBride, Mike McClure, Jacqui Miot, Jan Oltvoort, Pedro Plans-Rubio, Wolf Rogowski, Giuseppa Rudelli, Miguel Ruiz, Erik Sarpong, Mark Schulz, Paul Scuffham, Tomas Sechser, D Eldon Spackman, Hans-Martin Späth, Rosanna Tarricone, George Torrance, Peter Verplanken, F Randy Vogenberg, Adele Weston, Richard Willke, Michael Wonder, Shawn Xie and Hanna Zowall. However, we alone are responsible for any remaining errors.
REFERENCES


Box 1.

Steps for determining appropriate methods for adjusting cost-effectiveness (CE) information

- **CE information available?**
  - Yes
  - No
  - Relevant to the decision problem and sound methodology? (See Weibe et al criteria)
    - Yes
    - No
      - Simple adaptation, e.g. using an appropriate price vector
      - Consider other data and modeling
    - CE reported for comparable treatment patterns?
      - Yes
      - No
      - CE based on multi location trial, data from the jurisdiction included?
        - Yes
        - No
          - Consider analysis of individual patient data
          - Consider other data and modeling
        - Consider other data and modeling
Box 2. Representation of Qualitative and Quantitative Interactions by Gail and Simon\textsuperscript{12}

<table>
<thead>
<tr>
<th>Quantitative Interaction</th>
<th>Qualitative Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>- Net Benefits</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>- Net Benefits</td>
<td>+</td>
</tr>
</tbody>
</table>
Box 3. ‘Pooling’ and ‘splitting’ versus hierarchical modeling for multinational trials: estimating country-specific mean difference in cost

**Pooling and splitting analyses**

**Hierarchical modeling results**

*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.
### Box 4. Examples of decision modeling studies seeking to inform decision making in a particular jurisdiction using evidence from a range of locations.


This study considered the cost-effectiveness of early intervention (e.g. using coronary stents) compared to best medical management in patients with non-ST-elevation acute coronary syndrome. In the primary analysis, it estimates all clinical, utility and cost parameters using data from the only trial undertaken in the jurisdiction of interest (the UK). As a secondary analysis it incorporated relative treatment effect data from the full range of international clinical trials, using a meta-analysis. The advantage of the primary analysis was that the data are known to relate only the jurisdiction of interest with no need to ‘adjust’ data from other sources. Its limitation is that it effectively gives a zero weight to other evidence that exists outside the jurisdiction of interest.


This study used data from a multi-national trial on baseline risks, relative treatment effects, utility and resource use data. It used regression methods using data from the whole trial to estimate costs for the jurisdiction of interest (UK). Its strength is that it uses all available data from clinical trial but models the effect of country of randomization on costs. However, it assumes all clinical and utility data are fully generalizable across jurisdictions.


This analysis used a meta-analysis of relative treatment effects from all trials regardless of where they were undertaken; baseline risks, resource use and utility data from jurisdiction of interest. It assumed that relative treatment effects were independent of baseline risks, but also assessed the validity of this assumption using meta-regression. However, it still assumes that relative effects are fully generalizable across jurisdictions and that resource use and quality of life data are specific to country of interest.