

Increasing the generalizability of economic evaluations: Recommendations for the design, analysis, and reporting of studies

Michael Drummond, Andrea Manca, Mark Sculpher

University of York

Objectives: Health technology assessment (HTA) is increasingly an international activity, and HTA agencies collaborate to avoid unnecessary duplication of effort. However, the sharing of the results from HTAs raises questions about their generalizability; namely, are the results of an HTA undertaken in one country relevant to another?

Methods: This study presents recommendations for increasing the generalizability of economic evaluations. They represent an important component of HTAs and are commonly thought to have limited generalizability.

Results: Recommendations are given for studies using patient-level data (i.e., evaluations conducted alongside clinical trials) and for studies using decision analytic modeling.

Conclusions: If implemented, the recommendations would increase the value for investments in HTA.

Keywords: Cost-effectiveness analysis, Clinical trials, Modeling studies, Decision making

Health technology assessment (HTA) is increasingly an international activity, and many jurisdictions have established HTA agencies. The various agencies collaborate extensively, both informally and formally, through the International Association of Health Technology Agencies (INAHTA). This collaboration includes discussions about the prioritization of topics to avoid unnecessary duplication of effort, sharing of resources and expertise, as well as exchanging the results of particular assessments. The sharing of the results of HTAs raises issues about their generalizability. That is,

This work was funded by the NHS Research and Development Health Technology Assessment Programme. The views expressed are those of the authors and do not necessarily reflect those of the Department of Health. Mark Sculpher is the recipient of a career scientist award from the Department of Health in the United Kingdom, and Andrea Manca is the recipient of a Wellcome Trust Training Fellowship in Health Services Research. We are also grateful to those colleagues who worked on this study (Francis Pang, Sue Golder, Hege Urdahl, Linda Davies, and Alison Eastwood) for helpful comments and advice.

are the results of an HTA undertaken in one country relevant to another? Also, within a large country, do the results of a given HTA apply in all regions?

Undoubtedly, particular HTAs are of most use in the setting where they are conducted and could probably not be used without adaptation in another location. However, the value from investments in HTA would be greatly increased if studies could be made more generalizable. This strategy would facilitate closer collaboration between agencies and would be of particular benefit to smaller countries, which may not have the resources to conduct HTAs on a wide range of topics.

Several components of HTAs could be considered to be fairly generalizable. For example, the data from clinical trials are normally assumed to be generalizable with respect to setting (i.e., homogeneous among settings studied in the trial and relevant to other settings beyond those studied in the trial). On the other hand, data from economic evaluations are normally considered *not* to be generalizable, due

to factors, varying from place to place, that might alter the cost-effectiveness results (7). These factors include differences in the availability of health-care resources, clinical practice patterns, and relative prices. For example, a new health technology that potentially reduces the need for or length of hospitalizations, is likely to be more cost-effective in settings where there is adequate availability of community care. Differences in clinical practice patterns can themselves lead to differences in utilization of important resource items such as hospitalizations or the length of hospital stay. Also, the prices of major cost-drivers such as drugs can lead to changes in the relative cost-effectiveness of treatment regimens.

To some extent, issues of generalizability in economic evaluations can be addressed by using data from different locations. For example, a decision analytic model can be populated by data (i.e., unit costs or prices) from a range of settings. However, this may only help up to a point, as there may be situations where differences between settings may suggest a different model structure. In trial-based economic evaluations, where patient-level data on resource use, quality of life, and (possibly) utilities, unit costs or prices are collected as part of the trial protocol; the problems of lack of generalizability are often more serious and intractable. For example, where resource use varies by setting due to (say) different practice patterns, the pooled result from the trial (for resource use) may not apply to any individual location. This may make it quite difficult to deliver relevant data to decision-makers in a wide range of settings.

Recent reviews of the literature indicate that existing economic studies vary in the extent to which issues of generalizability are recognized and explored. For example, in some trial-based economic evaluations, the pooled resource use data are applied to all countries, without consideration of whether there are important differences in the impact of treatment on resource use from country to country. Similarly, in some modeling studies, the same data may be used to populate the model for more than one country (e.g., rate of hospitalizations), without adequate consideration of whether different estimates should be used. (See Barbieri et al. [2] for a recent review of current practice with respect to studies of pharmaceuticals in Western Europe.)

Given these problems, the National Health Service (NHS) Health Technology Assessment Programme in the United Kingdom commissioned a study of the generalizability of economic evaluations in time and place. The objectives of the study were to undertake a systematic review of the literature relating to generalizability in economic evaluation in health care and to undertake a series of case studies relating to multilocation trials and decision analytic models.

The full report has been published recently by the HTA Programme and is available on its Web site (18). This paper draws together the main recommendations for the design, analysis, and reporting of economic evaluations. Recommendations are made separately for trial-based studies (i.e.,

economic evaluations using patient-level data) and decision analytic models, because although many of the issues are common to both types of study, some are not. The objective of this study is to stimulate an international debate about the problems associated with the lack of generalizability of economic evaluations and to make recommendations on how the generalizability of studies can be increased in the future.

RECOMMENDATIONS FOR INCREASING THE GENERALIZABILITY OF ECONOMIC EVALUATIONS

Economic Evaluations Using Patient-Level Data

Economic evaluations using patient-level data, particularly conducted alongside randomized controlled trials, continue to provide an important source of data on cost-effectiveness. In these studies, the economic analyst has three opportunities to increase the generalizability of his or her study. First, at the *design* stage, the need for generalizability of findings can be anticipated. Second, in the *analysis* of results, qualitative and quantitative approaches can be used to produce findings relevant to a range of settings. Finally, in the *reporting* of results, attempts can be made to accommodate the needs of users/decision-makers in different geographical locations. This section discusses these issues in the context of patient-level data, with a focus on trial-based studies, and makes several recommendations for improved practice. Recommendations in relation to design, analysis, and reporting are given separately below, although recognizing that these issues are often interlinked.

Recommendations in Relation to Design of Studies

Clinical trials are primarily designed to estimate clinical parameters, for which generalizability has traditionally been considered less of an issue. Therefore, the economist seeking design changes for the purpose of increased generalizability will probably have to compromise on the range of changes sought. Nevertheless, several changes are possible, and it may be argued that some serve the clinical objectives of the trial, as well as the economist's need for increased generalizability of findings (e.g., the selection of one or more comparators that are widely used in several countries). Much depends on the body funding the study and its objectives. For example, a multinational pharmaceutical company may be seeking to appeal to several jurisdictions, whereas NHS Research and Development funders may only be interested in implications for the NHS in the United Kingdom.

Selection of Study Sites (i.e., Centers and Countries). For purposes of generalizability, selection of study sites would ideally focus on those that are representative of the jurisdiction(s) for which economic data are

Table 1. Possible Higher-Level Covariates To Use in a Multi-level Model

Jurisdiction level	Center level ^a
Percent GDP spent on health care	Volume of relevant cases treated
Reimbursement system for hospitals	Bed occupancy
Payment method for physicians	Hospital type (e.g., teaching or nonteaching)
Type of health-care system (e.g., public or private)	Range of clinical specialties
	Geographical location (e.g., large city or small town)

^a These covariates would be relevant if the “center” were a hospital. However, in principle, the “center” could be a region, health authority, or individual clinician.

GDP, gross domestic product.

required. In principle, this selection could be a single site, but is more likely to be several sites to reflect the variation in health care provision within and between different jurisdictions. The use of regression models, such as multilevel modeling, potentially provides a rigorous means of modeling variation in cost-effectiveness between centers or countries (13). These models recognize the hierarchical structure of the data, where patients may be “clustered” in centers and the centers, in turn, clustered in countries.

If the intention is to apply multilevel modeling techniques in the analysis of the economic data, it would be useful to collect data on center characteristics that could be used as covariates in the multilevel model. The same would apply to jurisdiction characteristics if the trial were being performed in more than one jurisdiction. These covariates will increase the efficiency of trial-wide cost-effectiveness estimates and, by looking at interactions with treatment, facilitate subgroup analysis by location characteristics. A statistical analysis plan should define covariates at all levels for which data are to be collected, and proposed analytical methods should be clearly stated. (This approach has been used recently in the UK Endovascular Aneurism Repair (EVAR) Trials (3) and is discussed in more detail later.)

More research is required to determine which characteristics are most useful as covariates in such a multilevel model, so initially it would be wise to collect a wide range of data. Some initial suggestions are given in Table 1. It would improve the efficiency of the model if the centers were selected randomly from the relevant population and a reasonable number (i.e., 15–20) were included in the trial. It would also be ideal to have a minimum number of observations (patients) in each center to make sure that the cluster characteristics are adequately represented.

Inclusion/Exclusion of Patients. It is typical, in clinical trials, to have criteria for inclusion and exclusion of patients. For an economic evaluation based on the trial to be generalizable, the patients included should reflect the normal clinical caseload. Therefore, there would be con-

cerns if a large percentage of patients were excluded from the trial. Another threat to generalizability would be if the “normal” caseload varies from place to place. This variation could arise if participating centers differ in respect of their catchment populations. In such situations, it would be important to have a wide range of centers in the trial. It would also be important to collect several patient-level variables that could be used as covariates in a multilevel model. These covariates could include age, gender, socioeconomic status, and previous medical history. This strategy would facilitate more efficient estimates of trial-wide treatment cost-effectiveness and allow the estimation of cost-effectiveness relating to subgroups based on patient characteristics (10). Patient-level variables are typically collected in trials already, so this should not impose any additional data collection burden.

There also may be instances where the centers’ characteristics determine their typical caseload. For example, centers of excellence typically treat more serious cases than normal general hospitals. In these situations, there is likely to be an interrelationship between center characteristics and patient characteristics. Hence, it becomes crucial to collect both patient-level and center-level variables.

Selection of Comparator Therapy. The comparator selected needs to be relevant to the jurisdictions in which the study is going to be used. Therefore, a threat to generalizability could exist if “current practice” varies from place to place. In some cases, it may be possible to agree on one or more compromise comparator(s) that reflect(s) normal practice in a wide range of settings. The alternative approach would be to let the clinician or center select their own comparator therapy (20). In this case, it would be important to ensure that the trial includes a representative sample of centers and/or physicians, and again there is value in using multilevel modeling to explore variation in cost-effectiveness by location.

Perspective of the Study. The various international guidelines for economic evaluation have differences with respect to study perspective. Some recommend adopting a societal perspective, whereas others focus on government expenditure or a particular budget (e.g., the drugs budget) (9). Therefore, the recommended approach, bearing in mind the need for generalizability, would be to adopt a broad societal perspective while retaining the capability to present costs and benefits by a range of different perspectives. There are also strong normative reasons for adopting the societal perspective (11).

Collection of Resource Use and Cost Data. The main recommendation here is to collect resource use data (e.g., hospital days, intensive care unit days, community nurse visits) separately from the unit costs or prices of those resources. The reasons for this approach are obvious. First, decision-makers considering a study undertaken in another location need to assess whether the practice patterns (and

resulting resource use) observed in the study apply in their own setting. Second, decision-makers in other locations may wish to apply their own prices to the units of resource use. Within the context of a clinical trial, these data can easily be collected through a combination of case report forms, patient diaries, and locally administered questionnaires (3).

Health State Preference Values. Health state preference values can be obtained from the literature, estimated directly on patients in the trial, or derived by using a generic instrument (e.g., EQ-5D, Health Utilities Index; 6). The generic instruments use a questionnaire, administered during the trial, to classify patients into health states. The set of values for states (i.e., the tariff) is then provided with the instrument, having been obtained from a community survey. For the purposes of generalizability, the health state valuations would ideally be relevant to the population(s) under study. For example, in the United Kingdom the National Institute for Clinical Excellence (NICE) guidance to manufacturers and sponsors of health technologies states that “health states should be measured in patients using a generic and validated classification system for which reliable UK population preference values, elicited using a choice-based method such as the time trade-off or standard gamble, are available” (p. 25 of reference 15).

Recommendations in Relation to Analysis of Results

Two approaches to the analysis of variability in cost-effectiveness by location, using data from multicenter or multinational trials, have been reported in the literature. First, Cook et al. (5) recommend a test of interaction approach to explore the level of homogeneity in the data. This method mirrors the approach frequently followed in the analysis of clinical data from multicenter trials. Namely, if no interaction exists between center and treatment effect, the data can be pooled, thereby giving a more precise estimate of treatment effect. Second, Willke et al. (23) have used a fixed effect regression approach, based on separate regressions for cost and outcomes, whereby country dummy variables are introduced alongside other explanatory variables.

The further development proposed here is to use multilevel modeling and, although further methods research is needed to identify the best way of applying these methods, this approach should be considered as part of the analysis of multilocation trials (13). The advantage of the use of multilevel modeling is that, if patient-level data are clustered by location, it will provide more appropriate estimates of the uncertainty around an intervention's cost-effectiveness; it can also facilitate location-specific estimates of cost-effectiveness. At the least, the approach can be used to consider the degree of clustering in costs and outcome data between locations and, hence, the extent to which this finding should be reflected in the full analysis.

Several further research issues arise in the context of multilevel modeling. These issues include the overall specification of the models; selection of patient- and location-level covariates and the specification of their interaction with treatment; the appropriate multilevel modeling approach, when there are several levels in the data hierarchy (e.g., patients, surgeons, centers, countries); appropriate methods when there are few locations in the trial; and the use of Bayesian approaches to multilevel modeling.

There has also been some use of econometric methods, such as selection models and instrumental variables, to adjust observational data sets for selection bias (12), and some consideration of those methods to increase the generalizability of randomized trials (14), especially in the context of comprehensive cohort analysis (16). Further research is justified into the principles and application of these methods. Although the greater use of formal statistical methods, such as multilevel modeling, is warranted in trial-based studies, there will remain an important role for sensitivity analysis in exploring the implications of variation in some parameters (e.g., unit costs and preference values).

Recommendations in Relation to the Reporting of Results

Even if it has not been possible to address fully all the issues of generalizability at the design or analysis stage, the needs of study users can still be partly accommodated during the reporting of results. The recommendations are summarized in Table 2. The general objective is to help the users of studies decide whether or not a given study is relevant to their own setting. One way to achieve this determination would be to report a table showing the characteristics of each site (country), so that the reader can assess whether these findings apply to his or her jurisdiction. Clearly, these additional reporting suggestions will be constrained by the limitations of space, particularly by journals. There is, therefore, an argument for greater use of more detailed technical reports to be made available as supporting documents, perhaps on journal Web sites. An important area of further research relates to the policy-relevance of the location-specific estimates of cost-effectiveness which multilevel modeling facilitates. Although the value of these results may be clear for individual countries in a multinational trial, the decisions that might be made, given different center-specific estimates of cost-effectiveness, are less obvious. In centrally funded health-care systems like those existing in Northern Europe, it is unlikely that policy-makers will differentiate between locations in making treatments available because of equity considerations. However, in more decentralized systems like that in the United States, where payers bargain locally with providers, the knowledge that an intervention is more cost-effective in Hospital A than in Hospital B may improve the efficiency of resource allocation decisions, although it may have implications for equity.

Table 2. Recommendations for Reporting the Results of Economic Evaluations Alongside Randomized Trials

Study element	Reporting recommendations
Study sites (centers)	Describe the characteristics of the centers participating in the trial. If these are from different countries, also report the relevant features of the various health-care systems.
Patient enrollment	Report the types of patients excluded from the trials and the percentage of the normal caseload that these represent. Comparison with the relevant patient population outside the trial centers.
Treatment alternatives	Describe the alternatives in detail, so that study users can assess the relevance to their own setting.
Perspective(s)	Report costs and benefits by each relevant perspective.
Resource use and costs	Report quantities separately from prices/unit costs.
Health state preference values	Report the source of the values and any instrument used.
Analysis of variability	Provide details of quantitative analysis of variability by location. Ideally, this will be based on statistical analysis (such as multilevel modeling), but should at least incorporate standard sensitivity analysis.
Other analytical issues	Provide details on the extent of incomplete observations (i.e., missing and censored data). Detail the characteristics of patients with incomplete data. Describe the methods used to address the problem.

Economic Evaluations Using Decision Analytic Modeling

The use of a single patient-level data set, such as a randomized trial, as a vehicle for economic evaluation frequently has several limitations. These limitations include the partial nature of the comparisons undertaken, short-term follow-up, use of intermediate rather than ultimate measures of health outcomes, and unrepresentative patients, clinicians, and locations. Given the increasing need for policy-relevant cost-effectiveness research to inform particular decisions about the funding and reimbursement of health-care interventions, these shortcomings of trial-based analyses will need to be addressed. The decision model represents an important analytic framework to generate estimates of cost-effectiveness based on a synthesis of available data and the explicit representation of uncertainty (4). Several recommendations for decision model-based economic evaluation, and suggestions for further research, flow from this. Again, these points are arranged under design, analysis, and reporting.

Recommendations in Relation to Design of Studies

Given the focus on a *decision*, any analysis should be clear about two important features of the research. The first is the specification of the decision problem. That is, the explicit statement of the options the cost-effectiveness of which is being compared and the patient group(s) for which the options are relevant. This key feature of the design of a decision model is a feature of most general guidelines in the area (17;22). The second important feature is less frequently identified in these guidelines, and it relates to the decision-maker(s) and jurisdiction(s) whose decision the model is designed to inform. In some cases, a specific decision-maker might be specified, such as the National Institute for Clinical Excellence, which issues guidance on the use of health technologies for England and Wales. For other decision models, a more general focus may be suitable such as individual Primary Care Trusts in England and Wales.

Once these features have been defined, an important next stage is to ensure that the overall analytical approach and structure are appropriate to the relevant decision-maker(s). This will rely on the latter having made a clear statement about factors such as the perspective of the analysis (e.g., health service or societal) and the relevant objective function (e.g., generic health gain such as quality-adjusted survival or disease-specific outcomes). Sometimes there will be a lack of clarity about these factors, or they will vary between decision-makers when the model is targeted on more than one. In these circumstances, there is value in adopting the broadest perspective and objective function, which will allow the results to be presented in several different ways.

Recommendations in Relation to Analysis of Results

The data that are used to populate a decision model should be justified given the stated target decision-maker(s) or jurisdiction(s). This justification will apply not just to unit costs but to resource use, effectiveness, and preference value data. Where several appropriate sources of data exist for a particular parameter, these sources should be appropriately pooled in such a way that the uncertainty relating to their precision and their possible heterogeneity is reflected in the model. This will involve standard meta-analysis (21), or more advanced methods of multiparameter synthesis (1;8). When sources of evidence are available from within, *as well as from outside*, the target jurisdiction, an important issue is whether and how the latter should be incorporated. Further research is needed to develop methods of evidence synthesis that combine data from a range of jurisdictions and allow for the additional uncertainty in this process.

When only data from outside the target jurisdiction are available, it is important to assess whether these findings can be assumed exchangeable across locations. In the clinical, as well as economic evaluation, fields, *relative* treatment effectiveness is often assumed to be exchangeable across locations and patient subgroups, whereas baseline event rates

are not. Given available data, the reliability of this assumption can be assessed empirically (19). For preference values, available evidence suggests little systematic variation between locations (e.g., countries) in mean values, indicating that location-specific estimates may not be essential (16). In the case of resource use and costs, it would be expected that location-specific data would be required, given their known variability. Further research would be valuable to consider the issues around using the same approach for resource use, costs, and preference values as for effectiveness data; that is, taking a relative treatment effect as exchangeable across locations and the baseline as location-specific. An important feature of such research would again be to reflect the uncertainty associated with the assumptions regarding the location-related exchangeability in the decision model.

In any decision model, there will be a range of different types of uncertainty to deal with explicitly and to reflect in the overall results and interpretation of the analysis. In this process, it is important to distinguish parameter uncertainty, which relates to the imprecision with which a parameter is estimated due to there being a finite sample, from variability or heterogeneity, which is concerned with how parameter estimates vary across contexts. These “contexts” could be patient subgroups or, as is the focus here, locations. As suggested above, parameter uncertainty may need to include the implications of taking data from sources other than the main jurisdiction of interest—further research will illuminate how this might be implemented. Probabilistic models, where data inputs are incorporated as uncertain variables, are the appropriate means of handling parameter uncertainty.

When a model is targeted at more than one decision-maker/jurisdiction, an important aspect of the analysis will be to assess the *variability* in results between locations. This strategy is feasible, using sensitivity or scenario analysis, as long as alternative parameter estimates exist for individual locations. These methods will be important for multinational analyses as well as multilocation studies within a given country.

Recommendations in Relation to the Reporting of Results

The level of detail and complexity involved with many decision models means that communicating all aspects of model

structure, assumptions, and data inputs can be a major task. Some general guidelines for this process have been published elsewhere (17;22). As noted above, the more extensive use of technical reports to support journal articles is likely to be very important for comprehensive communication. A key feature of reporting models is to be able to establish that each parameter input is appropriate for target decision-maker(s)/jurisdiction(s). This feature is part of the more general reporting task of justifying all assumptions and parameter values, but it is recommended that this is clearly related to the target customer. Again, explaining the methods that have been used to “pre-analyze” data inputs so they are suitable for incorporation into models (e.g., meta-analysis) is part of the general reporting process for decision models. This input should include any pre-analysis that was undertaken to adjust parameters estimated from the location in which they were measured to that which is relevant to the model.

CONCLUSIONS

Two checklists for assessing the generalizability of economic evaluations are presented in Tables 3 and 4 for trial-based and modeling studies, respectively. The checklists are intended to be useful for those decision-makers using economic evaluations, in particular those undertaken as part of a health technology assessment. The checklists may also be useful for those planning or undertaking economic evaluations. If the principles suggested in the checklists are followed by those conducting studies, it is likely that, over time, more economic evaluations will produce generalizable results.

Policy Implications

The value of undertaking HTAs would be greatly increased if these assessments could produce results that are generalizable beyond the setting in which the HTA is undertaken. The biggest threat to the generalizability of most HTAs is the economic evaluation component of the assessment, because there are several reasons why economic data may not be transferable from location to location. This study contains recommendations for the design, analysis, and reporting of economic evaluations that, if implemented, will increase their generalizability. It, therefore, offers governments and other

Table 3. Checklist for Assessing the Generalizability of Trial-Based Studies

-
1. Are study sites representative of the jurisdiction(s) for which data are required?
 2. Are study sites (centers) randomly selected?
 3. Can data on center characteristics be collected (e.g., bed occupancy levels)?
 4. Does the trial include a high proportion of the normal clinical caseload?
 5. Does the comparator therapy (to the technology of interest) represent current practice in the settings concerned?
 6. Is a wide range of user perspectives represented in the study?
 7. Are prices (unit costs) being collected separately from resource use data?
 8. Is a widely used generic instrument being used for quality of life (e.g., utility) measurement?
 9. Can regression-based techniques be used to obtain center-specific measures of cost-effectiveness?
 10. Are the reporting recommendations in Table 2 being followed?
-

Table 4. Checklist for Assessing the Generalizability of Modeling Studies

-
1. Are the decision problem, the relevant settings, and audiences (i.e., decision-makers) clearly specified?
 2. Does the overall analytical approach incorporate the relevant perspectives (e.g., health service or societal) and relevant objective functions (e.g., maximizing health gain)?
 3. Are the data used to populate the model relevant to the target audiences (i.e., decision-makers) and settings?
 4. Where data from different sources are pooled, is this done in a way that the uncertainty relating to their precision and possible heterogeneity is adequately reflected?
 5. If data from other settings are used, have these been assessed for relevance in the settings of interest?
 6. Is uncertainty (i.e., parameter uncertainty and heterogeneity) adequately reflected in the model?
 7. Are results reported in a way that allows the assessment of the appropriateness of each parameter input and each assumption in the target settings?
-

fundlers of HTAs the opportunity to increase the value from these investments in knowledge.

CONTACT INFORMATION

Michael Drummond, DPhil (chedir@york.ac.uk), Director and Professor of Health Economics, **Andrea Manca**, MSc, Research Fellow, **Mark Sculpher**, PhD, Professor of Health Economics, Centre for Health Economics, University of York, Heslington, York, North Yorkshire YO10 5DD, UK

REFERENCES

1. Ades AE, Cliffe S. Markov Chain Monte Carlo estimation of a multi-parameter decision model: Consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making*. 2002;22:359-371.
2. Barbieri M, Drummond M, Willke R, et al. Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: Lessons for inferring generalisability. *Value Health*. 2005;8(1):10-23.
3. Brown LC, Epstein D, Manca A, et al. The UK Endovascular Aneurism Report (EVAR) Trials: Design, methodology and progress. *Eur J Vasc Endovasc Surg*. 2004;27:372-381.
4. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet*. 2002;360:711-715.
5. Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. *Stat Med*. 2003;22:1955-1976.
6. Drummond MF. Introducing economic and quality of life measurements into clinical studies. *Annal Med*. 2001;33:344-349.
7. Drummond MF, Pang F. Transferability of economic evaluation results. In: Drummond MF, McGuire A, eds. *Economic evaluation in health care: Merging theory with practice*. Oxford: Oxford University Press; 2001.
8. Eddy DM. The confidence profile method: A Bayesian method for assessing health technologies. *Oper Res*. 1989;37:210-228.
9. Hjelmgren J, Berggren F, Anderson F. Health economic guidelines: Similarities, differences and some implications. *Value Health* 2001;4:225-250.
10. Hoch JS, Briggs AH, Willan A. Something old, something new, something borrowed, something BLUE: A framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*. 2002;11:415-430.
11. Johannesson M, O'Connor RM. Cost-utility analysis from a societal perspective. *Health Policy*. 1997;39:241-253.
12. McClellan M, Newhouse JP. The marginal cost-effectiveness of a medical technology. A panel instrumental-variables approach. *J Econ*. 1997;77:39-64.
13. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: The use of multilevel models. *Health Econom*. In press.
14. Manning R, Claxton K. *Experimental and econometric solutions to selection bias. Paper presented to the Health Economics Study Group*. York; 1997.
15. National Institute for Clinical Excellence. *Guide to the methods of technology appraisal*. London: NICE; April 2004.
16. Olschewski M, Schumacher M, Davis KB. Analysis of randomised and nonrandomised patients in clinical trials using comprehensive cohort follow-up study design. *Control Clin Trials* 1992;13:226-239.
17. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models: A suggested framework and example of application. *Pharmacoeconomics* 2000; 17:461-477.
18. Sculpher MF, Pang FS, Manca A, et al. Generalisability in economic evaluation studies in health care: A review and case studies. *Health Technol Assess*. 2004;8(49):iii-iv,1-19.
19. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: Comparison and development of approaches. *Stat Med*. 2000;19:3251-3274.
20. Simon GE, VonKorff M, Heiligenstein JH, et al. Initial antidepressant choice in primary care. Effectiveness and cost of fluoxetine vs tricyclic antidepressants. *JAMA*. 1996;275:1897-1902.
21. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song TA. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
22. Weinstein MC, O'Brien B, Hornberger J, et al. Principles for good practice for decision analytic modelling in health care evaluation. A Report of the ISPOR Task Force on good research practices—modelling studies. *Value Health* 2003;6:9-17. Available at: <http://www.ispor.org/workpaper/healthscience/TFModeling.pdf>. Accessed: February 2002.
23. Willke RJ, Glick HA, Polsky D, Schulman KA. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Econ*. 1998;7:481-493.