Guest Editorial

Conducting and interpreting multi-national economic evaluations: the measurement of costs

Anita Patel

Economic evaluations involving more than one country are generally carried out to accelerate recruitment, increase statistical power, improve generalisability and/or gather information relevant to regulatory requirements of particular countries. While increased speed and size of recruitment and regulatory compliance are usually achieved, meeting the goals of increased statistical power and generalisability can bring challenges.

Between-country variations in national-, local- and patient-level factors affect clinical outcomes and the supply of and demand for health care, for example, the overall health care budget, the mix of health care goods and services that are bought with that budget, relative prices of resources and the epidemiology of the disease. As theories of welfare production would suggest, such factors are not independent of each other. For instance, the costs of resources can affect the availability of services, or pricing of services may affect their demand, and this may in turn affect patient outcomes. In pursuit of a larger sample size to increase the robustness of statistical inferences, it is common to pool data together and thereby ignore the effects of such between-country differences. There are also more pragmatic factors to consider: language and terminology differences need to be accounted for in the collection of both economic and outcomes data; differences in accounting systems may affect the collection of unit cost data; and costs for each country are likely to be in different currencies (even if they share a currency, as in the case of the Euro in various European countries, its value will vary).

The fundamental challenges when carrying out economic evaluations across multiple countries are to consider whether there are any between-country variations in the economic or outcomes data and, if so, how to take account of these to ensure that the conclusions drawn from the evaluation are valid and meaningful both within and across the countries involved. This necessitates taking particular consideration throughout the process of the evaluation: when choosing a study perspective; when collecting resource use data; when collecting unit cost data; when imputing missing unit costs; when choosing an analysis perspective; when standardising costs to a common currency; when linking costs with outcomes; and when interpreting findings.

This editorial can only touch on the surface of this area. There are now several good comprehensive reviews of the issues involved and the way in which previous studies have dealt with them (for example, Pang, 2002; and Mogyorosy and Smith, 2005). This editorial uses empirical data to simply highlight some of the issues that may be encountered when choosing a study perspective and when collecting unit costs.

Choosing a study perspective

While it has been suggested that the selection of a study perspective is a surmountable problem in multi-national evaluations because a disaggregated presentation of results can allow them to be re-estimated differently (Pang, 2002), decisions nevertheless need to be made at the design stage of a study regarding what resources will be measured.

One of the factors to account for in this decision is potential differences across countries in the balance of care between different sectors of a health or other support system. This may be particularly prominent in mental health and elderly care contexts, where patterns of service use can be complex. Therefore, a broad study perspective may be needed to ensure coverage of all relevant cost drivers. A good example of this is provided by the QUATRO study (Thornicroft et al., 2005; Patel, 2006), a randomised controlled trial of two interventions for people with schizophrenia carried out across four European countries (England, Germany, Italy and the Netherlands; one centre per country).

The study planned to estimate cost-effectiveness through analysis of pooled data. This has been a common approach in several previous multi-national economic evaluations, especially when countries have geographical proximity, perhaps with the implicit assumption that they share cultural, social and political features. Does the virtue of all QUATRO participating countries being western European countries demonstrate this approach to be appropriate? The star plots shown in Figures 1 and 2 would suggest not.

Figure 1: Percentage of participants using each type of resource
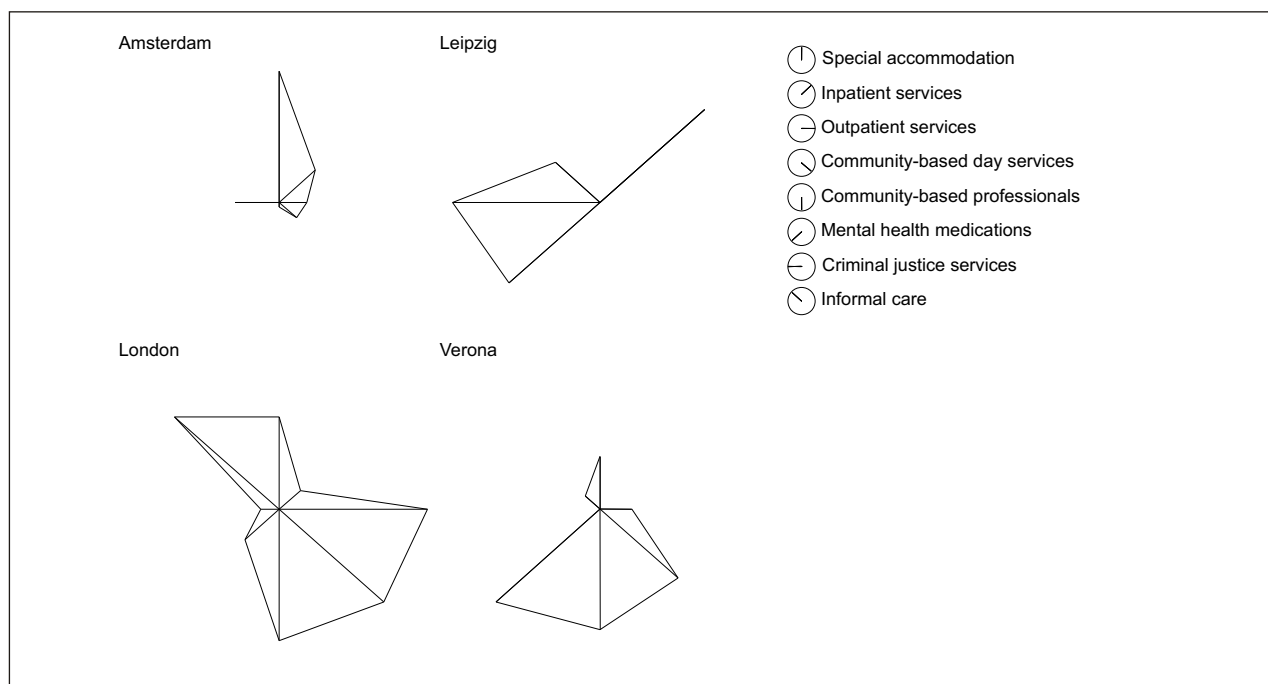
Figure 2: Mean amount of use for each type of resource
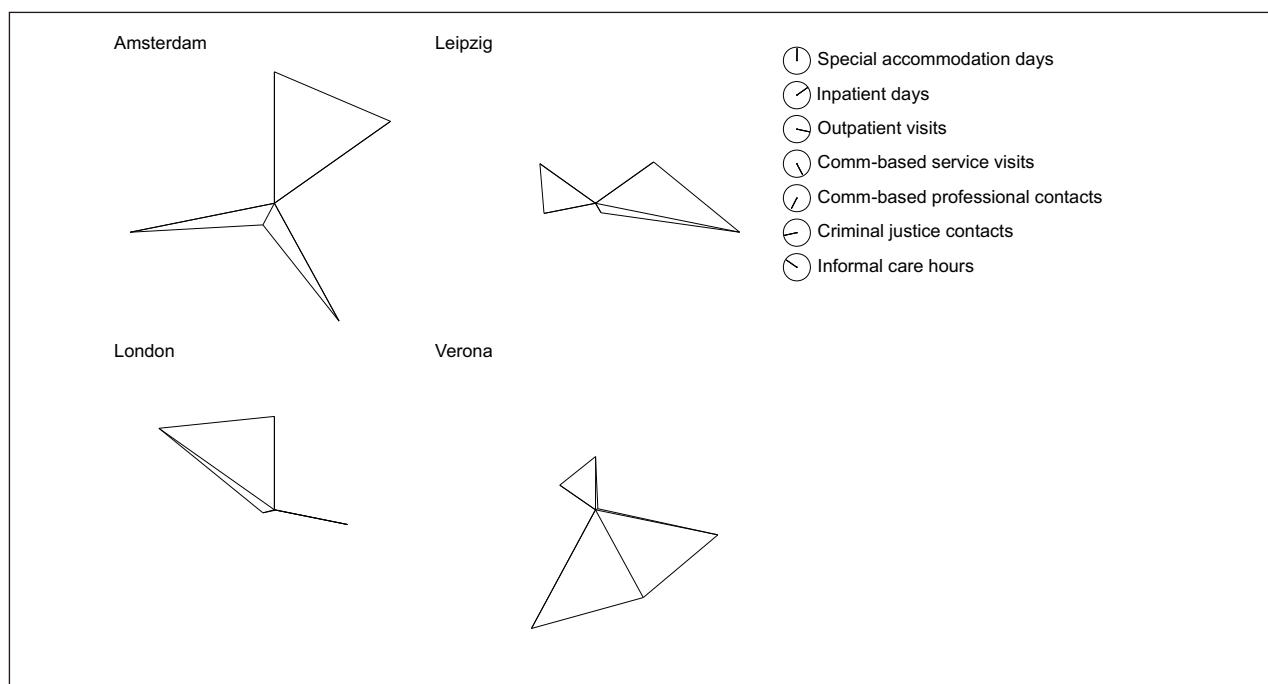
These star plots are a graphical representation of treatment patterns of participants in each country at baseline. Each country's star plot is formed of a series of lines stemming from the origin. Each line represents a particular resource and is differentiated from others by the angle at which it stems from the origin (see the keys provided in each Figure for explanation of which resource each line represents). The length of each line represents a quantity (in this instance, either the proportion of each sample that used a resource, or the mean amount of use). For example, in Figure 1, a vertical positive line represents the proportion of participants using special accommodation services, and in Figure 2, it represents the mean number of days spent in special accommodation. If the balance of resources used were comparable across the countries, they would have the same shaped star plots. If the number of resource users/resource quantities were comparable, then the size of the plots would be similar (note that values for each variable are re-scaled to fit the start plots).

Neither the shape nor size of the star plots is consistent across the four study sites. The proportion of participants in each country using secondary care services varied between 28-76 per cent per cent and average length of stay was equally variable (19-88 days). The balance of care between different care sectors also varies. Take-up of community-based services was more consistent across the countries. These simple star plots clearly show that overall service use patterns differed by site and raise questions about the suitability of pooling these data together. Such variations could be a result of any number of factors. It is also important to note that the trial included only one centre per country and it is unclear whether those centres were representative of national patterns in those countries. This highlights the fact that the choice of centres in multi-national trials may be less determined by the representativeness of a centre and more on the basis of ease of data collection (Johnston et al., 1999). Indeed, the QUATRO centres were wholly determined by the localities in which the existing research collaboration group were based (although the formation of the collaboration may originally have accounted for broader issues of generalisability).

Estimating unit costs

Estimating the unit costs of the resources that have been measured presents a further set of considerations. The availability, quality and structure of financial information may vary between countries, as a result of different accounting practices for example. Attention is needed to ensure that the cost components that contribute to the estimation of unit costs are standardised across countries. Definitions of a particular resource input or type of cost may vary between countries, for example, capital overheads may have different names which are not directly translatable, and the semantic meaning of a particular cost element may vary from country to country, for example, 'overheads' may variably include/exclude capital costs. Also, different efforts may be required to collect particular unit costs as the cost-driving items may vary from country to country e.g. due to differences in balance of care.

Limited research resources often necessitate pragmatic approaches so, in reality, many evaluations resort to using a mixture of newly-computed costs and readily available information, which in turn can be a mixture of top-down and bottom-up costing methods, which themselves can lead to variable estimations. The approaches taken to date vary from one country's unit costs being applied to all countries to centre-specific costs being calculated for each country. However, using one country's unit costs does not allow for the possible between-country variations in the relationships between inputs and outputs. In fact, this illustrates that researchers often simply strive for an aggregate composite 'score' for a disparate series of resource use data, rather than to truly explore cost implications. Using one country's unit costs will not be appropriate if the aim is to reliably inform decisions about the allocation of scarce resources within each participating country. Unfortunately, collecting country-specific costs for multi-national studies can be a huge challenge, and one that may be magnified in evaluations measuring a broad range of services across several service sectors.

Although the QUATRO study aimed to standardise the process of collecting unit costs in each centre through the use of a protocol and templates specifically developed for this purpose (based on Chisholm, 2001), it was not possible to adhere exactly to these due to study resource constraints, variations in the availability of information and limitations to the level of commitment that we could reasonably expect from collaborators to carry out this onerous task. Therefore, the unit costs were ultimately based on the most readily available sources, formed of a mixture of local accounts and national published estimates, which in turn were a mixture of costs and charges. Several adjustments were needed to address missing overhead costs, capital costs and activity data.

Table 1 shows, for each centre, the ratios of unit costs of various resources against a comparison resource. The comparators were chosen on the basis of being in the same service sector and all centres having a unit cost estimate available for it (which was not possible in the case of accommodation services). While there are similarities in ratios for some services across the four centres (for example, the unit costs for 24-hour staffed overnight accommodation facilities were approximately a third of the unit cost of acute psychiatric wards in Leipzig, London and Verona), there were also wide variations for other services. The unit cost for a psychologist is two-thirds of the cost of a psychiatrist in Amsterdam and Verona, but ratios are much lower at 0.43 and 0.31 in Leipzig and London respectively. Ratios for the cost of a primary care physician vary greatly from 0.51 in London to 2.27 in Verona. There appear to be no systematic patterns in these ratios across the centres and accounting for purchasing power did not explain them. Although these ratios are obviously dependent on the choice of comparator, it is likely that they represent

variations in the constituent components of the unit costs estimates and/or between-country differences in relative prices.

Table 1: Average ratios of unit costs

	Amsterdam	Leipzig	London	Vienna
Accommodation — ratio calculated against cost of acute psychiatric ward				
Overnight facility, 24 hours staffed	NA	0.33	0.31	0.35
Overnight facility, staffed (not 24 hours)	NA	0.33	0.06	0.06
Overnight facility, unstaffed	NA	NA	0.06	NA
Hospital inpatient services — ratio calculated against cost of acute psychiatric ward				
Acute psychiatric ward	1	1	1	1
Psychiatric rehabilitation ward	0.42	1	1	NA
Long-stay ward	0.16	1	0.76	NA
Emergency/crisis centre	NA	1	2.04	NA
General medical ward	0.58	1.42	1.52	0.92
Hospital outpatient services — ratio calculated against cost of acute psychiatric ward				
Psychiatric outpatients	0.1	NA	0.52	0.18
Non-psychiatric outpatients	0.11	NA	0.51	0.04
Day hospital	0.29	0.24	0.41	NA
Community-based services — ratio calculated against cost of community mental health centre				
Community mental heatlh centre	1	1	1	1
Day care centre	1	NA	0.14	2.6
Group therapy	NA	NA	0.16	2.2
Sheltered workshop	NA	NA	0.12	16.1
Specialist education	NA	NA	0.17	NA
Community-based professionals — ratio calculated against cost of psychaitrist				
Psychiatrist	1	1	1	1
Psychologist	0.65	0.43	0.13	0.67
Primary care physician	0.82	2.09	0.51	2.27
District nurse	0.39	0.4	0.22	0.51
Community psychiatric nurse/care manager	0.48	0.39	0.29	0.51
Social worker	0.48	0.39	0.44	0.37
Occupational therapist	0.48	0.39	0.18	NA
Home help/care worker	0.29	0.44	0.05	0.31

Note: NA = Not available

Future directions

The variations in resource use and unit costs illustrated here raises the question of how relevant the pooled findings would be for the individual study sites. Although the appropriateness of pooling data can be tested (for example, by testing for treatment-by-country interactions as suggested by Cook et al. (2003)), few studies are powered to conduct such tests. Country-specific analyses are also likely to lack statistical power and would anyway contradict one of the common rationales for such studies i.e. to maximise statistical power.

There is still insufficient consensus about the appropriateness of various available approaches to multi-national economic evaluation. This generates uncertainty, not only for evaluators who face such challenges in the day to day process of conducting the research, but also for policy makers who need to ensure that correct inferences are drawn from such evaluations. It seems sensible for those embarking on multi-national economic evaluations to begin under the assumption that there will be differences and that contextual information will need to be collated as part of the study to explore the presence and potential causes of such variations. Although country-specific analyses may lack power compared with pooled

analyses, they may nevertheless be a useful exploratory approach to examine whether pooled conclusions are at least broadly applicable to each study country. There is now a growing interest in developing methods to directly address these challenges. For example, multi-level modelling has been suggested as a middle-ground option to choosing between country-specific and pooled analyses because it accounts for the hierarchical nature of pooled data (Manca et al., 2005). However, it may not be a suitable approach for studies with small numbers of countries and/or centres. Pinto et al. (2005) and Willan et al. (2005) propose bayesian shrinkage estimation as another compromise option between pooled and country-specific analyses because it provides country-specific estimates that lack the variability associated with country-specific or sub-group analyses.

Despite great headway in this area in recent years, there are still few direct comparisons of different approaches and their impacts on trial results. Sensitivity analyses clearly have an important first line role to play in multi-national economic evaluations. Alternative approaches may have a quantitative impact on the findings, even if they do not change the overall conclusions. The importance of quantitative impacts will vary according to context. There may be acceptable levels of divergence between different approaches that evaluators or policy makers would be willing to accept (for example, quantitative differences may not matter as long as they are moving in the same direction). Conversely, there may be unacceptable levels of divergence, prior to the point at which the direction of the difference changes. Such questions are important because although quantitative impacts of alternative approaches may not change the conclusion of an evaluation, they suggest different budgetary impacts. It would also be important to know how representative the study sample is of the local populations and/or conditions.

This editorial has focused on economic evaluations conducted across more than one country. While geographical variations and their consequences are more prominent in multi-national studies, they are also present in multi-site/multi-centre evaluations carried out within a single country. Therefore, all of the issues discussed here (and their potential solutions) may also hold for single-country multi-centre evaluations. (It follows that multi-national economic evaluations which involve multiple centres within each country may present even greater complexities).

In considering these issues, we must remember the rationale for carrying out economic evaluations in the first place: to inform decisions about allocating scarce resources as efficiently and equitably as possible in order to maximise the welfare of a society. Geographical variations in various factors necessitate decision-making at localised levels. Sensitivity to context, in both conduct and interpretation, is therefore essential in multi-national evaluations. It could be the case that the 'summary measure' of total costs may be less useful in a multi-national context and that ever more sophisticated approaches to the analysis of multi-national data may prove less useful to decision-makers than simple breakdowns of data that can be reconstituted to suit their needs.

References

Chisholm, D.H. (2001) Cross-cultural mental health care evaluation: developing and demonstrating methods for economic analysis, PhD thesis, Institute of Psychiatry, King's College London, University of London.
Cook, J.R., Drummond, M., Glick, H. and Heyse, J.F. (2003) Assessing the appropriateness of combining economic data from multinational clinical trials, *Statistics in Medicine*, 22, 1955-1976.
Johnston, K., Buxton, M.J., Jones, D.R. and Fitzpatrick, R. (1999) Assessing the costs of health care technologies in clinical trials, *Health Technology Assessment*, 3.
Manca, A., Rice, N., Sculpher, M.J. and Briggs, A.H. (2005) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models, *Health Economics*, 14, 471-485.

Mogyorosy, Z. and Smith, P. (2005) The main methodological issues in costing health care services. A literature review, CHE Research Paper 7, Centre for Health Economics, University of York.

Pang, F. (2002) Design, analysis and presentation of multinational economic studies: the need for guidance, *Pharmacoeconomics*, 20, 75-90.

Patel, A. (2006) Issues in multi-national health economic evaluation, PhD thesis, Institute of Psychiatry, King's College London, University of London.

Pinto, E.M., Willan, A.M. and O'Brien, B.J. (2005) Cost-effectiveness analysis for multinational clinical trials, *Statistics in Medicine*, 24, 1965-1982.

Thornicroft, G., Bindman, J., Becker, T., Tansella, M. and Schene, A. (2005) Quality of life following adherence therapy for people disabled by schizophrenia and their carers. Report to the European Union (Quality of life and management of living resources programme 1998-2002), Contract number QLG4-CT-2001-01734.

Willan, A.R., Pinto, E.M., O'Brien, B.J., Kaul, P., Goeree, R., Lynd, L. and Armstrong, P.W. (2005) Country specific cost comparisons from multinational clinical trials using empirical Bayesian estimation: the Canadian ASSENT-3 economic analysis, *Health Economics*, 14, 327-338.

Acknowledgements