Addressing prospectively the generalisability of economic evaluation results from multi-centre RCTs

Adrian Gheorghe, MSc[1, 2], Tracy Roberts, PhD[1], Melanie Calvert, PhD[2], Sue Wilson, PhD[2]

[1] - Health Economics Unit, University of Birmingham, UK
[2] - Primary Care Clinical Sciences, University of Birmingham, UK
Correspondence to: Adrian Gheorghe, Primary Care Clinical Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom. E: axg986@bham.ac.uk T: +44 (0)121 4143951

Background

This paper outlines the initial thoughts for the methodological component of my PhD (started October 2010). It addresses the argument for conducting research in this direction, brings forward a proposed research plan and discusses the potential applications of the investigation.

The document is structured as follows: the Introduction section summarises briefly the current developments concerning the generalisability of economic evaluation results. Two research questions are formulated and these form the basis of the Rationale section, where knowledge gaps are identified. It is argued that centre selection should play an important role when addressing generalisability of economic results from randomised controlled trials (RCTs). A Research plan is further proposed to strengthen the available knowledge on the topic. The paper ends with a discussion of the potential applications of this research.

Introduction

Economic evaluations of medical interventions have long been conducted alongside randomised controlled trials (RCTs). The rationale and the relevant methods have been described in the literature(1;2). RCTs are particularly expensive to run in terms of time and monetary resources, therefore studies are often designed so that patients are recruited from several clinical sites located in various locations within or across jurisdictions[1] for various, mostly pragmatic reasons e.g. to meet recruitment targets, to ensure that the study

[1] 'Jurisdiction' is referred to throughout the document as the administrative space to be informed by the findings of the trial (e.g. NHS England & Wales, NHS Yorkshire and the Humber)

population is representative to a certain (often unspecified) extent or, in trials supported by the pharmaceutical industry, to obtain evidence of effectiveness in countries where reimbursement decisions are envisaged(1;3). The issue of applying results from a RCT conducted in a specific geographical setting to another setting arises naturally. From a clinical perspective, RCTs are often conducted across jurisdictions (this refers especially to multinational RCTs) without acknowledging that trial results may vary with location(4).

In a landmark paper on the topic published more than 20 years ago, Drummond et al(5) first suggested that economic evaluation results were not completely transferable between jurisdictions. Their main findings were that cost-effectiveness results were significantly different between countries; and the main drivers of variability were the cost variations and the patterns of care. The focus group research of Hoffmann et al(6) pointed out that the issue of generalisability was of particular concern for decision makers.

A terminology note should be made. 'Generalisability', 'transferability', 'portability' have all been used to describe the extent to which economic evaluation results are applicable from one geographical setting to another. Boulenger et al(7) suggested that 'transferability' may be a broader concept than 'generalisability' as it encapsulates both the intrinsic value of the results and the methods available to assess their applicability in various settings. Other categorisations have been proposed(8). The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force has referred to economic evaluation results that are applicable to other settings without adjustment as being 'generalisable', while the term 'transferable' has been used when results were adapted to other settings(9). Furthermore, Sculpher et al(3) made the distinction between generalisability over location and generalisability over time. The focus of this document is entirely on generalisability over location. For the purpose of this paper, no formal distinction has been made between 'generalisability' and 'transferability'.

Current research

Two research directions have been intensively explored in the literature: the factors which influence transferability; and the methods that address the transferability of centre-level or study-level results.

Several papers investigated the factors that affect transferability, mostly in the context of multinational studies. Welte et al(10) identified three broad categories of such factors: methodological characteristics of the economic evaluation; healthcare system characteristics; and patient characteristics. Barbieri et al(11) concluded that variations of cost-effectiveness results between countries were not systematic. The main determinants were found to be the country-specific resource use, major cost drivers (hospitalization costs, drug costs), and the willingness-to-pay for an additional unit of effect. Sculpher et al(3) comprehensively reviewed the determinants of generalisability and identified four groups of factors influencing the geographical variability of economic evaluation results: patient factors; clinician factors; health care system factors; and wider socio-economic factors. Their review was, again, particularly relevant to a multinational context. Goeree et al(12) identified in a more recent systematic review no less than 77 factors affecting transferability falling into five categories: patient; disease; provider; health care system; methodological. Unit costs, practice patterns and clinical efficacy were highlighted as the most influential factors.

Two broad categories of methods have been suggested to aid decision makers in assessing the transferability of economic evaluation results. The first category refers to methods looking at economic evaluation studies as a whole. Boulenger et al(7) introduced a checklist and a sub-checklist of relevant factors in order to aid comparison between studies conducted in different countries. A transferability index was then constructed based on the number of completed items on the checklist and their score. Welte et al(10) proposed a decision maker-friendly transferability chart: following general knock-out criteria and specific knock-out criteria, an assessment of whether modelling adjustments were needed could be made. If that was not the case, the results were judged to be transferable. Another checklist-like approach was employed by Spath et al(13), who used five transferability criteria to assess the eligibility of economic evaluations for transfer to the French health care system. Turner et al(14) described the process of developing a toolkit for adaptation of health technology assessment results from one country to another. Building on the transferability factors identified by Boulenger et al(7) and Welte et al(10), Antonanzas et al(15) combined a number of objective and subjective factors to create a global transferability index and specific transferability indices applicable to economic evaluation

studies, respectively. Their framework was applied in assessing the transferability of 27 economic evaluations conducted in Spain.

A second category of methods use primary data from multi-centre[2] and/or multinational RCTs to refine local cost-effectiveness estimates. Willke et al(16) used cost and outcome regressions with country-specific interaction terms. Coyle and Drummond(17) used regression methods and one-way sensitivity analysis (ANOVA) to point out that cost-effectiveness estimates vary between centres in an RCT-based economic evaluation conducted in the UK. Multilevel modelling (MLM) has been suggested as a particularly useful method to refine centre-specific cost-effectiveness estimates, particularly in multinational RCTs; nevertheless, an important assumption made in MLM methods is that the participating centres are randomly selected, which has been acknowledged as an intrinsic limitation(18;19). Bivariate hierarchical modelling has also been suggested (20;21). Both MLM and bivariate hierarchical modelling use trial-wide information to adjust the centre-specific cost-effectiveness estimates by means of a shrinkage factor. Goeree et al(22) reported a framework allowing the selection of the number of hospitals (one or more) from which unit costs should be collected for economic evaluations along multi-centre studies. A recent and comprehensive synthesis of the analytical approaches presented above together with their limitations has been given by Manca et al(4).

Objective of the paper

This paper aims to argue that centre selection for multi-centre RCTs is relevant when considering the generalisability of trial-based economic evaluation results. The argument to be presented in the following section originates from two observations based on the evidence summarised in the Introduction section. First, the majority of studies mentioned above addressed transferability mostly from an international perspective by referring to multinational RCTs and acknowledged that the proposed adjustment methods can also be applied to single-country studies. While between country differences are arguably obvious, there remains the question **whether the available evidence is sufficient to justify the use of generalisability techniques in refining centre-specific economic evaluation results of single-country RCTs**.

[2] For the purpose of this paper, 'multi-centre RCTs' refers to single-country trials recruiting from more than one centre

Second, a series of general recommendations for further research in the area of generalisability have been made (3;9;23). As far as the recruiting centres are concerned, two directions were suggested: selecting representative centres in economic evaluations conducted alongside RCTs so that generalisability is ensured; and collecting as much centre-specific data as possible in order to be used for statistical modelling. Both these recommendations lead to a new research question: given that the majority of the available analytical techniques that serve generalisability are employed retrospectively using trial-wide results, **what is the scope for a prospective quantitative methodology (applicable at the trial design stage) to support generalisability**?

These two relevant questions will be addressed in the following section and their implications will be discussed.

Rationale

1. Are economic evaluation results expected to differ significantly between within-country centres?

A reasonable starting point for the first question is that extrapolating the economic evaluation results of an RCT interests the decision makers, who want to know whether and the extent to which a particular intervention is cost-effective in their jurisdiction (e.g. country). A question that arises is the following: **are there reasons to believe that economic evaluation results vary between centres within the same country**? Sculpher et al(3) mentioned in their review several studies to have detected differences in cost-effectiveness estimates between centres as a result of differences in unit costs and practice variation. Only one of these studies was UK-based. Nevertheless, they suggested that obtaining centre-specific cost-effectiveness results required further exploration in the direction of their usefulness for local policy makers. Goeree et al(12) acknowledged that "the little evidence that does exist suggests that hospital cost variation may be as large within countries as it is between countries" (p. 565). The immediate thought would be that variables intrinsic to the patient (e.g. age, co-morbidities), the healthcare inputs (e.g. qualification of surgeons, number of CT scanners) and the healthcare system (e.g. financing streams) may explain reasonably well the source of these variations. Even in healthcare systems like the NHS where hospital reimbursement relies on largely fixed tariffs (Payment

by Results), hospital-specific costs are expected to vary (24;25). Recent evidence focused on obstetrics departments in UK hospitals has shown that the hospitals' cost-efficiency may still vary even when patient- and centre-level characteristics were accounted for(26). Coding inaccuracies, apportioning shared costs and managerial inefficiency were all indicated as potential explanations for the observed differences. In the light of this evidence and given that the interdependence between costs and outcomes is often difficult to quantify, there are reasons to expect a potentially significant variation in cost-effectiveness between centres at the very onset of the RCT. This should lead to a proportionate interest from local decision-makers in accounting for as much of this variation as possible in economic analyses of interventions.

The following step along the reasoning line is thus: **can it be ascertained *a priori* that economic evaluation results actually vary between the centres involved in the RCT**? Unfortunately this question can only be reliably answered in retrospect, once the trial results have been analysed. Gail and Simon(27) have described such tests for heterogeneity to test the influence of centre on any parameter of interest (e.g. incremental cost-effectiveness ratio ICER or net monetary benefit NMB), may it be in direction (qualitative interaction) or magnitude (quantitative interaction). Their methods were further discussed and applied by Cook et al(28) in the context of a multinational RCT. The problem with heterogeneity tests is that they are usually underpowered and it has been recently suggested that heterogeneity tests should not be used at all(4). Nevertheless, once heterogeneity has been ascertained and the need for adjustment acknowledged, the methods outlined in the Introduction section can be used to refine the cost-effectiveness estimates for each of the participating centres. Refining should be understood in this context as adjusting the centre-specific economic evaluation results based on the trial-wide estimate. For instance, Manca et al(21) accounted for patient-level and jurisdiction-level covariates simultaneously using bivariate hierarchical modelling. It must be made clear that the existing methods refer to data analysis and are retrospective in nature because they are only applicable when the trial results are complete.

It can thus be concluded that there are reasons to believe that the cost-effectiveness of health interventions can vary significantly between centres, although there is no methodology yet available to evaluate *a priori* the extent of this variation. It must be noted here that a 'significant' variation does not only refer to a different decision (invest/not

invest), but also to a magnitude which can be relevant to the local decision maker (see below).

2. What is the scope for a prospective quantitative methodology to support generalisability?

In order to address the second research question, the line of reasoning can be continued from the previous paragraphs: **are there limitations inherent in the retrospective approach of these methods that question the validity of the adjusted cost-effectiveness estimates**? Two observations must be made. First, neither of the existing methods makes any verifiable assumption regarding the centres included in the analysis. In other words, pooling data from centres from within the same jurisdiction is assumed to reliably lead to a representative cost-effectiveness estimate for the entire jurisdiction. Intuitively, at least, this assumption should hold if the centres are representative for the jurisdiction they represent. This could be achieved in two ways: either centres are deliberately chosen based on a number of covariates which recommended them as representative at jurisdiction level; or the centres are randomly selected from the pool of available centres in the jurisdiction. There is no evidence in the literature to date that either of these two conditions have been satisfied. Moreover, given the host of factors influencing cost levels, it is expected that the definition of 'representative' is both complex and difficult to specify. Purposive selection of recruiting centres/sites has been previously mentioned (3;9;29) without further details. Drummond et al(23) suggested possible centre-level covariates and introduced the concept of minimum patients recruited from each centre, but no consistent method to address this suggestion has yet been developed. The issue of randomly selecting centres has been touched upon in the literature (18;19) rather as a limitation and an area where more research should be conducted. Furthermore, choosing an insufficient number of centres and corresponding sample sizes can only lead to biased mean estimates and large variances. The issue extends to the suitability of using average unit costs across all the centres in the trial: in the absence of unit costs missing completely at random (MCAR), the average unit cost will most likely misrepresent the centre-specific cost (30;31). This is a legitimate concern: a recent systematic review of economic evaluations conducted alongside trials funded by the UK Health Technology Assessment Programme revealed that only 52 of 95 reviewed studies used unit costs that were sourced locally (32).

Second, the existing generalisability methods still leave decision makers from jurisdictions that were not involved in the trial with difficulties in transferring the economic evaluation results. Building on the first limitation outlined above, this equally applies to centres that belong to a jurisdiction included in the trial but have not contributed with primary data. The authors of the suggested methods righteously advised that great caution should be taken when considering such extrapolations. If the generalisability refinements have not incorporated centre-specific covariates (e.g. patient case-mix), one potential approach would be to find a similar recruiting centre in terms of the covariates considered in the model and then simply use the economic evaluation result. However, there is no guarantee that such a centre exists and, more importantly and in relation with 'representativeness', there is no straightforward indication as to what exactly constitutes 'similarity'.

Manca et al(19) offered a comprehensive account of the analytical strategies available depending on the availability of individual-patient data (IPD) and participation in the trial. The proposed framework was designed to address multinational studies, but the authors suggested that it may be useful for within-country jurisdictions. In the absence of IPD and if the jurisdiction of interest did not participate in the trial, decision-modelling was the indicated option. Decision models usually offer cost-effectiveness estimates with confidence intervals around them according to the uncertainty and sensitivity analyses incorporated; if uncertainty around the point cost-effectiveness estimate is large (and the confidence interval is wide), the result is of little use and its applicability is restricted to jurisdictions which are assumed to have identical budgets and identical priorities. Of course, constructing a decision model for each centre would be impractical. An alternative solution would be to make the decision model available to all interested decision makers, who may adjust the parameters to their own needs. This would involve specifying a transparent and user-friendly decision model and circulating it to decision makers.

Another approach involves the use of a preliminary decision model. Glasziou et al(33) have used a preliminary cost-utility analysis to inform data selection and the required sample size in the LIPID trial. Sculpher et al(34) also included decision modelling as preceding trial-based exercises in gathering evidence on prospective interventions. The question at hand is the following: **when a preliminary decision model already gives an estimate of cost-effectiveness robust to sensitivity analyses, under what circumstances is**

it worth collecting prospectively additional centre-specific data alongside an RCT for further modelling? A host of answers can be thought of. First, the preliminary decision model may often be based on effectiveness and resource use estimates from outside the jurisdiction (e.g. another country) and the impact of the differences would be difficult to assess. Second, in relation to the concerns expressed in the previous paragraph, even if country-specific cost-effectiveness estimates exist, there is no guarantee that they reflect national practice if the process of centre selection (centres from which primary data were collected and the estimates have been calculated) has not been justified. Finally, let us assume that a country-specific decision model exists and wide sensitivity analyses around the base case estimates have proved virtually every scenario to be cost-effective. It would thus be expected that the intervention is cost-effective in any given jurisdiction within the country and a yes/no decision can be made. However, taking the UK example for a moment, where increased decentralisation is about to be implemented and commissioning devolved to GP consortia, this is unlikely to be enough in the near future. The magnitude of the cost-effectiveness estimates is expected to increasingly matter for local resource allocation and there will be less and less scope for simple yes/no decisions.

The discussion centred on the two questions above attempted to identify the knowledge gaps associated with retrospective methods concerned with IPD analysis from RCTs. Two main themes emerged, both giving reason for concern: first, no explicit methods of selecting centres and their corresponding sample sizes have been mentioned in the methodological literature, although this has been suggested in the literature at a conceptual level (9;19;23). This limitation may hamper the validity of a series of computations, from heterogeneity tests to adjusted cost-effectiveness estimates. And second, there is no reliable tool available for decision makers representing centres and/or jurisdictions which did not participate in an RCT to relate to the trial-wide results when having to make decisions in their own settings. These themes suggest that centre selection has not been addressed in the literature and it may matter in deriving centre-specific cost-effectiveness estimates, however its impact has not yet been estimated.

The following section outlines a research plan that aims to address this gap.

Research plan

The **objective** of this research is to establish the impact of systematic centre selection at trial design stage in multi-centre RCTs coupled with economic evaluations in obtaining accurate centre-specific cost-effectiveness estimates. 'Systematic' is employed in this context to acknowledge that centre selection obviously takes place in real practice; however the process of identifying centres and quantitatively integrating the centre-specific factors is unclear.

A number of successive **research questions** addressing this objective have been formulated:

1. What is the current practice in selecting centres for economic evaluations alongside RCTs in the UK?
2. How can centre-level characteristics be identified and quantitatively integrated at the RCT design stage to select centres so that the sample is representative for the jurisdiction?
3. Is there a real benefit from selecting a representative sample of participating centres compared with existing modelling techniques in obtaining accurate estimates of centre-specific cost-effectiveness estimates?

The following paragraphs outline how each research question will be addressed.

Research question 1: *What is the current practice in selecting centres for economic evaluations alongside RCTs in the UK?*

As no method of selecting centres has yet been identified in the methodological literature, there is an interest to establish how centres are selected in the current real practice of RCTs coupled with economic evaluations. Two methods will be employed to provide an answer to this question: a systematic review and a survey.

A. A systematic review will be conducted to investigate how centres were selected in economic evaluations alongside RCTs in the UK published after 2005. Two questions will be answered: 1) did authors report the rationale for selecting and including centres in the RCT? And 2) did authors discuss generalisability by location when reporting the trial-wide economic evaluation results?

The following databases will be consulted: Medline, Embase, NHS Economic Evaluations Database, and Health Economic Evaluations Database. The review will only include papers reporting trial-based economic evaluations (either solely or as a subsection) and RCT protocols that describe a prospective economic evaluation. It can be assumed that the economic evaluations are published later than the trial results, so every effort will be made to retrieve the trial protocol for inclusion in the review, subject to availability. The focus of the review is whether generalisability has been addressed **prospectively**, at the onset of the RCT. The 2005 threshold has been chosen as such because the first recommendations on centre selection with respect to trial-based economic evaluations were published in 2004(3) and 2005(23).

The following information will be extracted: study authors; publishing year; acronym of RCT; intervention; control; type of economic evaluation; rationale for centre selection; and discussion of generalisability by location. The review will thus be able to summarise the extent to which centre selection is reported and the rationale (if any) used in the selection process. Furthermore, the review will also investigate whether researchers discussed generalisability by location of the trial's economic evaluation results.

B. Building on the results of the systematic review, particularly on the various rationales reported to drive the selection of centres, a questionnaire will be elaborated and circulated to all 49 UK CRC Clinical Trials Units. The objective of this survey will be to investigate how centre selection for RCTs coupled with economic evaluations are reflected in the views of the professionals involved in the trials units (trial managers, health economists, statisticians, epidemiologists). The survey will collect information on:

- what are the strategies, methods and criteria currently employed for identifying and selecting centres (e.g. hospitals, community practices);
- whether and how generalisability of economic evaluation results is accounted for in data analysis; and
- the interests that various professionals have in selecting centres.

The results of the survey are meant to inform on the current practice of selecting centres for economic evaluations and on establishing the professionals involved in RCTs who make decisions with this respect. Evidence for neither of these questions has yet been

identified in the literature. Moreover, the results of the survey will be compared against the current assumption of generalisability techniques presented in the Introduction section i.e. that centres are representative for their jurisdictions.

The systematic review and the survey will offer an account on how centres are selected for multi-centre RCTs in the UK. The scope of the research was restricted to trials conducted in the UK for two reasons: 1) analysing economic evaluations published in other countries involves a workload that exceeds the resources allocated to the project; and 2) administering the questionnaire to UK clinical trials units offers the unique opportunity of validating and potentially enriching the findings of the systematic review.

Research question 2: *How can centre-level characteristics be identified and quantitatively integrated at the RCT design stage to select centres so that the sample is representative for the jurisdiction?*

A number of centre-level characteristics have been suggested in the literature(23) and centre selection at the trial design stage has been advocated, but there is no evidence yet that a consistent technique has been elaborated (the systematic review detailed above attempts to clarify this issue to a certain extent).

The research is expected to unfold in two stages:

1. Identifying and selecting relevant centre-specific characteristics

These variables will have to fulfil a series of basic requirements:

- there is evidence in the literature of them influencing either costs or outcomes of the respective intervention;
- they can be plotted onto a numerical scale;
- their values are easily accessible.

It is not necessary for these covariates to be applicable to all interventions. For these reasons, it is expected there will be a set of general covariates (e.g. hospital size and case-mix) and a set of disease-specific or intervention-specific covariates, given that no significant correlation between specific and general covariates will be identified. A thorough systematic review of the literature is likely to reveal the most important such factors. One issue to be resolved will be selecting the variables to be accounted for based on the strength of available evidence.

2. Quantitatively integrating centre-level covariates in a prospective **generalisability index (GIx)**. Such an index would have to be calculated at jurisdiction and centre level, respectively, and then become the basis of an algorithm of selecting centres in the design stage of the RCT. It is expected that the centre-specific and global GIx's will be derived from a combination of a 2*xn* vector (a_{ij}) and the centre-specific sample size, where:

- n is the number of selected covariates;
- a_{1j} is the point estimate of the covariate;
- a_{2j} is the weight to be attached in the aggregation process (to allow for differential weighting of these factors). This weight may possibly be related to the impact on cost-effectiveness results, the strength of evidence and the uncertainty around a_{1j}. The process of assigning this weight may well be controversial and is subject to discussion.

The research aims to explore how the centre-level GIx's can be brought together in the context of sampling techniques in one formula so that the combination of *m* GIx's (where m is the number of envisaged centres) falls as close as possible (with specifiable uncertainty) to the jurisdiction GIx, calculated based on the jurisdiction level estimates (e.g. nationwide) and the trial-wide sample size. This approach would thus address the suggestion of Drummond et al(23) that centres could be selected as to be representative and a minimum sample size for each should be specified.

Data collection for the selected covariates will ideally be made based on publically available sources (e.g. Office for National Statistics, Public Health Observatories, Hospital Episode Statistics) and local institutions records (e.g. hospitals, PCTs).

Several issues will need to be addressed, such as defining a numerical scale for the GIx and investigating its statistical properties.

Research question 3: *Is there a real benefit from selecting a representative sample of participating centres compared with existing modelling techniques in obtaining accurate estimates of centre-specific cost-effectiveness estimates?*

Building on research question 2 above, it must be established whether the systematic centre-selection method is valid and feasible. Feasibility will be judged against the burden of collecting centre and jurisdiction-specific data and the computational burden. Validity testing will most likely involve investigating the relationship between centre-specific

cost-effectiveness estimates (possibly obtained using the available methods) and the centre-specific values of the GIx using primary data from at least two RCTs. Exploring this correlation would provide first-hand evidence that the GIx can be used at the design stage.

Anticipated results

This research aims to investigate a knowledge gap in the methods addressing the generalisability of economic evaluation results from RCTs. There is currently little to no published evidence on how centre selection is actually carried out. Recommendations towards centre selection have been made but there is no evidence to suggest that they have been developed or that they have been applied in practice. A consistent method will be elaborated and its merits will be assessed alongside the current modelling techniques relying on primary data.

This research may be beneficial from a multitude of angles. The immediate consequence refers to the current methods of addressing generalisability of economic evaluation results. As highlighted earlier, these retrospective methods make the assumption that centres are representative for the jurisdictions they represent. The results of the systematic review and the survey will provide evidence towards the strength of this assumption. A prior investigation of centre representativeness will thus become necessary in order to warrant the use of these methods. In perspective, the methodology based on the generalisability index is meant to assist retrospective modelling techniques in assessing the external validity of the trial *as it was designed* and in pursuing more and more precise cost-effectiveness estimates, to a reasonable level.

From a clinical perspective, it would allow an advance in clinical trials recruitment by providing a method able to state which centres are of more interest than others in terms of extrapolating economic evaluation results. While the object of this research is not RCTs per se, such a method may also inspire the centre selection process in multi-centre studies that do not necessarily have an economic evaluation component.

From a policy perspective, decision makers are often faced with making judgements based on evidence resulting from trials conducted in settings that have little relevance to

their jurisdictions. It is thus important to provide a quantitative tool allowing them to plan the need for future RCTs so that their results will be as relevant as possible to the policy context. The method outlined above would compel decision makers to think in advance of the locations where they are interested in applying the economic results of a clinical trial and also of how health services could be re-shaped in particular regions so that an intervention becomes locally cost-effective. Moreover, it is expected that the generalisability index (GIx) may assist not only the design of the study, but also (conversely) the refinement the cost-effectiveness estimates. Moreover, it may well prove to be useful to decision-makers from centres that did not contribute with primary data in the RCT in relating quantitatively to the trial-wide results by offering a working framework against which 'similarity' can be better described.

Providing evidence that centre selection makes a difference in estimating local cost-effectiveness results and that the systematic selection process is a feasible task would also have implications on the reporting manner of RCTs in general and of economic evaluations conducted alongside RCTs in particular. Current reporting guidelines(35;36) do not mention centre selection as an item and the outcome of this research may turn into an advocate for considering the introduction of this reporting requirement.

The successful development of the research and methodology presented in this paper would bring a set of valuable norms in a research area dominated by empiricism, with strong implications for clinical trials methodology and policy making.

Thank you for agreeing to discuss this material in the HESG. This research is only at its beginning and quite a number of methodological and practical issues are to be resolved (some have been touched upon above), not to mention unforeseeable challenges. Thoughts and feedback would be highly welcome and valued, especially given the early stage of the research. I would be particularly interested on your thoughts regarding the extent to which integrating centre-specific covariates can become truly quantitative to assess generalisability and the potential validation approaches for the generalisability index (GIx).

Reference List

(1) Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press, 2007.

(2) Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press, 2005.

(3) Sculpher M, Pang F, Manca A, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Asses 2004;8(49).

(4) Manca A, Sculpher MJ, Goeree R. The Analysis of Multinational Cost-Effectiveness Data for Reimbursement Decisions: A Critical Appraisal of Recent Methodological Developments. Pharmacoeconomics 2010;28(12).

(5) Drummond MF, Bloom BS, Carrin G, et al. Issues in the Cross-National Assessment of Health Technology. International Journal of Technology Assessment in Health Care 1992;8(04):670-82.

(6) Hoffmann C, Stoykova BA, Nixon J, et al. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value in Health 2002 Mar;5(2):71-8.

(7) Boulenger S, Nixon J, Drummond M, et al. Can economic evaluations be made more transferable?. [Review] [36 refs]. European Journal of Health Economics 2005 Dec;6(4):334-46.

(8) Mason JM, Mason AR. The generalisability of pharmacoeconomic studies: issues and challenges ahead. [Review] [36 refs]. Pharmacoeconomics 2006;24(10):937-45.

(9) Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value in Health 2009 Jun;12(4):409-18.

(10) Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics 2004;22(13):857-76.

(11) Barbieri M, Drummond M, Willke R, et al. Variability of Cost-Effectiveness Estimates for Pharmaceuticals in Western Europe: Lessons for Inferring Generalizability. Value in Health 2001 Jan;8(1):10-23.

(12) Goeree R, Burke N, O'Reilly D, et al. Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. [Review] [104 refs]. Current Medical Research & Opinion 2007 Apr;23(4):671-82.

(13) Spath HM, Carrere MO, Fervers B, Philip T. Analysis of the eligibility of published economic evaluations for transfer to a given health care system: Methodological approach and application to the French health care system. Health Policy 1999 Nov;49(3):161-77.

(14) Turner S, Chase DL, Milne R, et al. The health technology assessment adaptation toolkit: description and use. International Journal of Technology Assessment in Health Care 2009 Dec;25:Suppl-41.

(15) Antonanzas F, Rodriguez-Ibeas R, Juarez C, et al. Transferability indices for health economic evaluations: methods and applications. Health Econ 2009 Jun;18(6):629-43.

(16) Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from multinational clinical trials. Health Econ 1998;7(6):481-93.

(17) Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluations. International Journal of Technology Assessment in Health Care 2001;17(2):155-63.

(18) Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models.[Erratum appears in Health Econ. 2005 May;14(5):486]. Health Econ 2005 May;14(5):471-85.

(19) Manca A, Willan AR. 'Lost in translation': accounting for between-country differences in the analysis of multinational cost-effectiveness data. [Review] [112 refs]. Pharmacoeconomics 2006;24(11):1101-19.

(20) Willan AR, Pinto EM, O'Brien BJ, et al. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. Health Econ 2005;14(4):327-38.

(21) Manca A, Lambert PC, Sculpher M, Rice N. Cost-effectiveness analysis using data from multinational trials: the use of bivariate hierarchical modeling. Medical Decision Making 2007 Jul;27(4):471-90.

(22) Goeree R, Gafni A, Hannah M, et al. Hospital Selection for Unit Cost Estimates in Multicentre Economic Evaluations: Does the Choice of Hospitals Make a Difference? Pharmacoeconomics 1999 Jun;15(6):561-72.

(23) Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: Recommendations for the design, analysis, and reporting of studies. International Journal of Technology Assessment in Health Care 2005;21(02):165-71.

(24) Street A, Maynard A. Activity based financing in England: the need for continual refinement of payment by results. Health Economics, Policy, & Law 2007 Oct;2(Pt:4):4-27.

(25) Malcomson JM. Hospital cost differences and payment by results. Health Economics, Policy, & Law 2007 Oct;2(Pt:4):4-33.

(26) Laudicella M, Olsen KR, Street A. Examining cost variation across hospital departments--a two-stage multi-level approach using patient-level data. Social Science & Medicine 2010 Nov;71(10):1872-81.

(27) Gail M, Simon R. Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets. Biometrics 1985;41(2):361-72.

(28) Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. Statistics in Medicine 2003;22(12):1955-76.

(29) Reed SD, Anstrom KJ, Bakhai A, et al. Conducting economic evaluations alongside multinational clinical trials: Toward a research consensus. American Heart Journal 2005;149(3):434-43.

(30) Grieve R, Cairns J, Thompson SG. Improving costing methods in multicentre economic evaluation: the use of multiple imputation for unit costs. Health Econ 2010 Aug;19(8):939-54.

(31) Raikou M, Briggs A, Gray A, McGuire A. Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. Health Econ 2000 Apr;9(3):191-8.

(32) Ridyard CH, Hughes DA. Methods for the Collection of Resource Use Data within Clinical Trials: A Systematic Review of Studies Funded by the UK Health Technology Assessment Program. Value in Health 2010;13(8):867-72.

(33) Glasziou PP, Simes RJ, Hall J, Donaldson C. Design of a cost-effectiveness study within a randomized trial: The LIPID trial for secondary prevention of IHD. Controlled Clinical Trials 1997;18(5):464-76.

(34) Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ 2006;15(7):677-87.

(35) Drummond MF, Jefferson TO. Guidelines For Authors And Peer Reviewers Of Economic Submissions To The BMJ. BMJ: British Medical Journal 1996 Aug 3;313(7052):275-83.

(36) Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Jan 1;340.