Systematic reviews of economic evaluations: utility or futility?

Rob Anderson, PhD

Peninsula Technology Assessment Group, Peninsula Medical School, Universities of Exeter and Plymouth, Exeter.

(Version: 17th December)

To be discussed at

Health Economics Study Group meeting,

at UEA Norwich, January 2008

DRAFT: PLEASE DO NOT CITE WITHOUT AUTHOR'S PERMISSION

Disclaimer(s): none

Correspondence about the paper to:

Dr R Anderson,
PenTAG
Peninsula Medical School
Noy Scott House
Royal Devon & Exeter Hospital
Barrack Road
Exeter EX2 5DW
United Kingdom

Tel: +44 1392 406967
Fax: +44 1392 406401
e-mail: rob.anderson@pms.ac.uk

Abstract

Systematic reviews of studies of effectiveness are the centrepiece of evidence-based medicine and policy making. Increasingly, systematic reviews of economic evaluations have also become an expected input into much evidence-based policy making, with some health economists even calling for 'an economics approach to systematic review'.

This paper questions the value of conducting systematic reviews of economic evaluations to inform decision making in health care. I argue that the likely value of systematic reviews of economic evaluations is usually undermined by two things. Firstly, the inherently more limited generalisability of resource use and cost data and cost-effectiveness results over time and between health systems and service settings. Secondly, the fact that one of the two main forms of economic evaluation – decision analytic modelling – is now itself a well-developed method of evidence synthesis, should largely obviate the need for comprehensive systematic reviews of previous economic evaluations of a particular health technology or policy choice.

To encourage more useful incorporation of insights and knowledge from previous economic studies in evidence-based policy making, I propose a limited and specific range of reasons for conducting systematic reviews of economic evaluations. Furthermore, I conclude that apparent 'meta-analytic expectations' - for clear and unambiguous cost-effectiveness conclusions from systematic reviews of numbers of economic evaluations - are optimistic and largely futile.

key words: systematic review, economic evaluation, economic studies, evidence

Acknowledgements:

I thank current and past colleagues, and others, for comments on an earlier paper on the appropriate role of economic evidence in evidence-based policy-making more generally, from which the arguments in the current paper have emerged. I am also grateful to Ian Shemilt (of the Campbell Cochrane Economic Methods Group, and UEA Norwich) for sharing pre-publication versions of the *In press* chapter for the Cochrane Handbook on 'Incorporating economic evidence' in Cochrane Reviews, and which has been a useful stimulus to thought.

Introduction

Evidence-based policy making cannot rely on single empirical studies, and must generally seek to synthesise evidence from a systematic selection of the most relevant and rigorous existing research for a given decision problem. This is partly so that politicians and service managers can be seen to be impartial, "unbiased" and rational in their consideration of "all relevant evidence" about alternative policies. However, it is also for the more fundamental and simple reason that it is nearly always impractical to commission, in advance, the specific empirical research that would inform any given policy choice, since the results would not be available in time to inform the decision. Instead, policy makers and the researchers that support them must usually find a way of making best use of the findings from the many partially relevant studies that are already available.

In the health care field, given the variety of criteria for policy evaluation (efficacy, effectiveness, safety, cost-effectiveness, feasibility, acceptability etc.), and corresponding variation in the methods used to evaluate these goals, there is a vibrant ongoing debate about the appropriate methods of research synthesis for informing policy and practice (Dixon-Woods et al. 2005;Lavis et al. 2005;Popay 2006). In health economics, despite recurrent concerns about the lack of use of 'economic evidence' in health policy making - and the publication of a key book in 2002 (Donaldson, Mugford, & Vale 2002) - this debate has been comparatively limited. The form of research synthesis which has attracted the most methodological attention from health economists is the decision analytic model as a basis for cost-effectiveness analysis. At the same time, however, systematic reviews of economic studies have become a key feature of many policy making processes (especially in national health technology assessment processes) and also a common form of published study in certain health economics and medical journals.

A little over a decade ago, it was observed that health economists "have not yet developed a formal methodology for reviewing and summing up evidence from individual economic evaluations … or indeed for assessing whether systematic reviews are possible in this context" (Jefferson, Mugford, & Gray 1996). Nowadays, however, those who conduct systematic reviews of economic evaluations have available to them: guidelines on how to conduct such reviews (Carande-Kulis et al. 2000); checklists to enable more consistent

appraisal of the quality of included studies (Evers et al. 2005); graphical tools to summarise the results of numbers of economic evaluations (Nixon, Khan, & Kleijnen 2001); and dedicated bibliographic databases (such as NHS EED) and standardised search filters (also from CRD at York) to ease the task of identifying economic evaluations and other economic studies. The question is no longer 'are they possible?', but 'are they worth the effort?'

This paper therefore questions the fundamental value of conducting systematic reviews of economic studies to inform decision making in health care. Others have voiced similar concerns (Drummond 2002;Mugford 2002). Indeed, in relation to conducting systematic reviews of economic evaluations, Mike Drummond has observed that "it is not clear that the motivation to produce an authoritative statement of cost-effectiveness is as strong as it appears to be for clinical effects size" and he further asserts that "there is widespread recognition amongst economists, and possibly amongst decision makers, that whether or not a particular intervention is cost-effective depends on the local situation" (p.150). In a review of studies concerning the generalisability of economic evaluations, 26 different factors were identified which authors believed could make cost-effectiveness results vary from location to location or over time (Table 4, p.10; (Sculpher et al. 2004)). Also, more recently, studies reporting substantial between-country differences in public preferences for the same health states, have again raised the question of whether *any* cost-effectiveness findings can be confidently transferred from one place to another (Russell 2007). Despite Drummond's assertion, the findings of the Sculpher et al. HTA review, and more concrete evidence that local context does indeed impact on cost-effectiveness estimates, health economists continue to expend much effort on conducting systematic, and often quite exhaustive reviews of economic evaluations.

This paper therefore develops and extends some of these earlier observations and arguments about why systematic reviews of economic evaluations might very rarely allow valid and widely applicable statements concerning cost-effectiveness or the efficiency of policies or treatments. It attempts to ground these arguments in the broader context of the information needs of evidence-based policy making, the limits to the generalisability of the results of economic studies, and with recognition of the increasing role of model-based economic evaluation – itself, an increasingly sophisticated form of evidence synthesis.

Finally, I conclude by proposing a limited and specific range of reasons for conducting useful systematic reviews of economic evaluations, and make some more general suggestions about the limits to the generalisation of findings across numbers of economic studies.

Systematic reviews *of* or *for* economic evaluations?

There are various possible roles for economic evidence and systematic reviews in policy-making.

a. Systematic reviews *of* economic evaluations
b. Systematic reviews *for* (model-based) economic evaluations
c. Economic evaluation *alongside* systematic reviews (of effectiveness)
d. "Economic methods for use in systematic reviews"
e. "Economics components" of effectiveness (eg. Cochrane) reviews

In this paper, I am not directly concerned with systematic reviews for economic evaluations; that is, primarily, the role of systematic reviews in informing the parameters of (model-based) economic evaluations. The arguments that justify the need for model-based analyses to inform decision making, together with the increasingly rigorous standards for conducting and reporting them, are well documented elsewhere (Buxton et al. 1997;Sculpher, Claxton, & Akehurst 2005). (Of course, there remain ardent critics of the typical primary outcomes of such analyses: the incremental cost-effectiveness ratio and the quality-adjusted life-year (Birch 2002;Birch & Gafni 2006;Oliver 2003).) Also – largely on the grounds that their exact meaning is as yet rather unclear - I am also putting to one side any consideration of conducting economic evaluations *alongside* systematic reviews, and the use of *economic methods in systematic reviews*.

As already noted, systematic reviews of numbers of economic studies (of a particular technology, or in a particular patient group), have become a more common type of journal paper, and also a more common requirement within formal processes for evidence-based policy making. Some journals, notably *Pharmacoeconomics* and *Expert Reviews in Pharmacoeconomics and Outcomes Research*, specifically invite systematic reviews of economic studies. A review of the NHS EED database suggests that between 100 and 200

reviews or systematic reviews of "economic studies" are published each year (Figure 1). Meanwhile, many government agencies for the appraisal of health technologies or public health policies currently require systematic reviews of the relevant economic literature (see Table 1). Why?

Since systematic reviews, even of particular types of study, can be conducted for a wide variety of purposes (Petticrew 2001), it is worth briefly examining what the review questions of systematic reviews of economic evaluations appear to have been. Although a great many have been conducted specifically to assess the methodological quality of economic evaluations or other methods in a given area (Jefferson, Vale, & Demicheli 2002), many more appear (implicitly at least) to be based around some version of the following review question: "What is the cost-effectiveness of intervention X (compared with Y or Z)?" Can such questions be credibly answered by a systematic review of economic evaluations? Furthermore, if we believe that economic evaluations are primarily meant to be a decision-informing method of analysis (and therefore *should* be inherently jurisdiction- and time-specific) does it even make sense to seek a generalisable or average answer to such questions?

Limits to generalising from numbers of economic studies

There are a large number of reasons why economic evidence may not be transferable between different places and times (Drummond & Pang 2001;Sculpher, Pang, Manca, Drummond, Golder, Urdahl, Davies, & Eastwood 2004). I would argue that, for the same reasons, it may not be realistic to expect economic evidence synthesised from many different places and times (i.e. a number of economic evaluations) to usefully inform the likely cost-effectiveness of an intervention in a particular decision context.

What is 'economic evidence'

In this paper I refer to economic evaluations as a specific subset of health economic studies, in which comparative data is collected or reported on both costs and effects (however measured), and this permits an incremental analysis (Drummond et al. 2005). I also refer to a broader class of research as 'economic evidence' or 'economic studies' which, in addition to economic evaluations, may include comparative cost comparisons, or non-comparative

studies such as cost analyses, cost-of-illness studies, or even resource use or service utilisation studies which do not include any valuation of the resources/services consumed. Such studies usually still shed some light on the resource (and cost) consequences of different treatments or associated with the different chronic or acute health states of a particular patient group.

The main arguments of this paper are meant to be a critique of systematic reviews of economic evaluations. However, when I come to describe what might be more legitimate 'economic questions' that are worthy of systematic review, it will become clear that the range of economic study types that may need to be reviewed is somewhat wider.

Methodology and standards

Clearly, one major reason that the findings of economic evaluations (of the same intervention comparison) vary is that they have used different detailed methods. This is both due to a continuing lack of standardisation of methods, and a typical lack of compliance with those standards that have been established (Drummond 2002). On the other hand, it is also acknowledged that there are a considerable number of methodological considerations which feed into economic evaluations for which inter-country or regional variations can be expected and justified (Sculpher & Drummond 2006). Therefore, even with complete compliance with jurisdiction-specific methods standards, there is still considerable scope for variation in the analysis and results of economic evaluations. However, even assuming that all 'unjustified' methodological variations could be corrected, there are other more powerful reasons for variations in cost-effectiveness which would still remain.

Intervention context and intervention costs

The most compelling reason for questioning the value of systematic reviews of economic evaluations is that, as a specific type of outcome, costs and resource use are highly likely to vary from country to country, in different regional or service settings, as well as over time. It is widely stated that a key reason for this is differences in unit costs (e.g. between countires, and over time due to inflation). However, intervention context may also substantially impact on the costs of an intervention (or its comparator) through interventions being provided with different levels and combinations of particular resources

in different health systems (e.g. with different clinical grades, or different average length of hospital stay for a key procedure) or in different service settings (e.g. a different balance between primary and secondary care). Of the many conceptual factors that have been identified which impact on the variability of cost-effectiveness estimates, six are explicitly associated with cost (absolute/relative costs, economies of scale, exchange rates, (different combinations of) healthcare resources, financial incentives, and opportunity cost) (Sculpher, Pang, Manca, Drummond, Golder, Urdahl, Davies, & Eastwood 2004).

Intervention context and intervention effects

The impact of context on the cost component of the cost-benefit equation, is further compounded by other reasons that the effectiveness or health benefits of interventions are also likely (and legitimately) to vary from place to place and over time. These have been well documented elsewhere, so I do not expand on them at length here (Jackson, Waters, & for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce 2005;Kraemer, Frank, & Kupfer 2006;Kravitz, Duan, & Braslow 2004). However, it is worth noting that health economists have contributed to this debate, in terms of effectiveness being an interplay between the changes introduced by an intervention and the existing mix of the underlying determinants of health or disease in a population (Birch 1997;Birch & Gafni 2003). Health system or service context, including professional cultures, are also believed to be a major factor in determining the success or failure of using different financial incentives or "economic interventions" (Kristiansen & Gosden 2002).

For many clinical questions about a treatment's *efficacy* it may be reasonable to presume, on the basis of well controlled studies in humans in different places, that the findings (e.g. effect sizes) will transferable to similarly diagnosed patients elsewhere. However, this reasonable assumption about the generalisability of clinical study findings is only made on the basis of a well-defined (and accurately diagnosed) patient group, in whom usually) a well explained biological disease process that is taking place, and for which the intervention has a theoretical mode of action on this process, which is also well supported by a body of scientific evidence. It may ultimately also be empirically supported by a lack of heterogeneity, or easily explained heterogeneity of results in different studies.

Contrast this with a decision-focused form of evaluation in which the real-life effectiveness of interventions, as used in routine practice, is the object of interest. There are very few (if any) health interventions or technologies which are effective in and of themselves; there is nearly always some assumed mixture and level of staffing, and a minimum degree of physical and organisational infrastructure which enables the intervention to work at a particular level. These intervening variables to do with the adequacy of the service setting, the skill-levels of the practitioner, or the motivations and capabilities of patients to properly adhere to treatment, can all impact on the ultimate effectiveness of the intervention. Furthermore, if there are a variety of clear ways in which a new technology or treatment can be provided to patients, this also expands the full range comparators that could (and some argue, should) be evaluated.

The context of the decision

As well as the context of the *intervention*, economic evaluations to inform decision making will usually have a particular *decision context*. Critically, in any particular decision-making context, this will determine what resource use does or does not have an opportunity cost. Thus, the cost-effectiveness of the same intervention in two places may be different even in circumstances where the incremental resource use and health effects are identical, because the alternative uses of the resources used are different. For example, in some situations – perhaps where an operating theatre or a ward is being used to full capacity – there will be an alternative use for operating theatre slots or beds on a ward, but in other hospitals that are working under-capacity there may not really be any benefits foregone due to those slots or beds going unused. Similar issues concerning how the scope and scale of service changes, and particularly whether new services were primarily intended to expand (i.e. supplement) service capacity or re-locate (i.e. substitute) capacity, can substantially impact on their estimated cost-effectiveness were well illustrated in Coast et al.'s insightful review of four economic evaluations of hospital at home programmes (Coast et al. 2000).

Two main types of economic evaluation

Although earlier texts on economic evaluation mainly focused on the three-way classification of studies as cost-benefit analyses, cost-effectiveness analyses, and cost-utility analyses, it is now more fully recognised that there is also a key distinction between

decision-model (simulation) based analyses, and empirical economic evaluations which collect patient-level data on costs and outcomes as part of a clinical trial (or other comparative outcome evaluation). They are now such different and specialised forms of economic evaluation that they have dedicated chapters and separate textbooks to explain the methods, and some have recently questioned the value of conducting empirical economic evaluations (Sculpher et al. 2006).

At the very least, in order to retain comparability amongst reviewed studies, this means that any systematic review of economic evaluations should become two systematic reviews: one of empirical (or trial-based) economic evaluations, and one of decision model-based economic evaluations. Trial-based and model-based economic evaluations of an intervention will invariably not be comparable for the same reasons that people advocate conducting decision model-based analyses (e.g. including the full range of relevant comparators, a representative case-mix of patients, and following patients for a sufficiently long time for all significant cost and outcome differences to be captured) (Buxton, Drummond, van Hout, Prince, Sheldon, Szucs, & Vray 1997;Sculpher, Claxton, Drummond, & McCabe 2006).

Cost-effectiveness not cost-efficacy

Because economic evaluations are primarily intended to inform decisions, they are more explicitly concerned with effectiveness rather than efficacy; that is, they seek to assess the incremental benefits and incremental costs implied by a given "real-world" choice between two interventions or programmes *as they would be resourced and implemented in routine practice.* The distinction between effectiveness and efficacy studies is in fact further recognition that context matters, in determining both benefits and costs. Essentially there is no point conducting cost-efficacy studies, in idealised controlled clinical settings and with very rigid protocol-driven care and specially selected patients, because it is widely expected that the costs and effectiveness of the "same treatment" will be different if delivered within routine care settings and across the whole case-mix of eligible patients. This lack of external generalisability of trial-based economic evaluations is a major element of the critique of trial-based economic evaluations and clearly has implications for the value of

conducting systematic reviews of such studies (Sculpher, Claxton, Drummond, & McCabe 2006).

Reasons for reviewing economic studies

I suggest that there are probably only three good reasons for conducting systematic reviews of economic evaluations: (1) to inform the development of a decision model; (2) to identify the most relevant single study to inform a particular decision, and; (3) to identify the key economic (causal) trade-offs implicit in a given treatment choice or disease area. I discuss each of these in turn below, including the likely review questions and the types of economic studies which might be included.

Reviews to justify and inform decision model development

When there is a plan to develop a new model for evaluating the cost and effectiveness of a set of health policy alternatives, some kind of systematic review of previous economic evaluations is at least necessary to rule out that there is not already in existence a recent, highly relevant and rigourously conducted analysis of essentially the same decision problem (and which would render the planned evaluation superfluous). This is simply the good academic practice of making sure that a piece of research will not be answering a research question which has, effectively, already been answered. It may therefore not go much further than a systematic search of the published literature and recent unpublished sources, in order to confirm that there is no recent economic evaluation of the same comparators in the same populations and jurisdiction.

If the planned model-based analysis is justified, there are various ways in which reviewing previous economic studies might usefully inform the development of a new decision model.

- First, previous decision model-based analyses might provide insights into some of the key trade-offs, clinical events, and changes in health states which are thought to determine how the types and levels of resource use implied by alternative interventions are associated with different outcomes. Previous models might not reflect all the important resource-outcome relationships implicit in a given decision problem, but they should provide an initial list of the key ones.

- If there are previous model-based analyses examining similar decision problems or treatements, these might indicate the strengths and weaknesses of different modelling approaches (e.g. simple decision trees vs Markov models vs discrete event simulations).
- Previous empirical economic studies, which have collected and reported resource use and/or effectiveness data for the same patients, may also usefully inform a new decision model. However, this will largely depend on whether the study is purely descriptive, or has attempted to explain how and why different types of patient are associated with different levels or mixes of resource use, or different levels and patterns of outcomes. Studies which merely report, for example, which types of resource use were measured and valued, but do not provide a breakdown of the incremental cost estimate by type of resource use would be less useful at helping decision modellers decide what resource use should be specified in any new model.

The use of reviews of model-based economic evaluations has been encouraged by Pignone on the basis of their experience of conducting reviews of economic evaluations for the US Preventive Services Task Force (Pignone et al. 2005). First, because model-based economic evaluations are themselves syntheses, it makes no sense consider traditional meta-analysis or the pooling of results from such studies. Instead, they suggest, reviews of model-based economic evaluations are "most useful for comparing and contrasting how different investigators have chosen to structure their models and estimate key variables" and can also "clarify how results differ between studies based on these different assumptions" (p.1073 in (Pignone, Saha, Hoerger, Lohr, Teutsch, & Mandelblatt 2005)). A recent example of such a review, in relation to the impact of structural assumptions in decision models, is that by Drummond et al. on models for rheumatoid arthritis (Drummond & Barbieri 2005).

Reviews to identify the most relevant study

There will be some decision-making situations where there are insufficient resources to conduct an original model-based evaluation of the specific decision problem being faced. In this situation, rather than not consider any economic evidence at all, it may be better to identify the best quality study (or two) which is most relevant to the decision being faced.

In this context, "best quality study" would have to include considerations of both internal validity (study design and methodological quality) and external validity (e.g. how long ago? similarity of comparators? similarity of health service/system settings?). If there happened to be several studies of similar quality, in relation to the current decision context, then it might be worth examining to what extent and why their cost-effectiveness estimates vary. However, this would primarily be with a view to contextualising the results of the most relevant high quality study, rather than to expect some average result to emerge. There may also be the possibility of updating and re-estimating the cost-effectiveness using local resource use prices, or perhaps through inflating and converting costs from past studies in other countries.

Reviews to understand the key economic trade-offs and causal relationships in a decision problem or disease or treatment area

This reason for conducting systematic reviews of economic studies is almost the antithesis of why many such published reviews seem currently to be undertaken; instead of (usually implicitly) seeking a generalisable empirical regularity, the aim of such a review is explanatory. Of course, the underlying 'explanation' of why a new intervention or treatment has a particular incremental cost may largely be that the new technology itself has a much higher per patient price. However, explaining *why* some interventions are more cost-effective than others will often be determined by a whole range of other trade-offs to do with downstream health effects, different adverse event rates, or different levels of patient compliance. With more complex interventions, it becomes even harder to explain how a specific bundle of intervention components (and their associated resource use), provided in a given context, have generated the levels and types of outcomes measured.

However, either as a preliminary to conducting a model-based economic evaluation, or as an exercise in developing theory, it would be useful to conduct systematic reviews of economic studies which seeks to answer a review question such as: "How do the level and configuration of resources involved in treatment/service design strategies P, Q and R appear to be related to the levels and types of outcomes observed, and what contextual factors affect these relationships?". Decision models, after all, are essentially a simplified expression of what these key trade-offs are presumed to be, and we often do not describe

clearly where these structural assumptions come from (Cooper et al. 2005). Such theory-building reviews may need to abandon the more protocol-driven and research design focussed processes typical of traditional narrative reviews of effectiveness evidence, and instead use approaches such as 'realist review' which use more iterative analysis and purposive sampling with which to build up a reliable picture of the key causal relationships at work in a given area of programme design or treatment (Pawson 2006;Pawson et al. 2005).

Although it might be useful to be able to explain how different configurations of resources appear to cause different levels and types of outcomes, the fact remains that few empirical economic evaluations have this explanatory intent. Therefore, so long as individual (empirical) economic evaluations remain almost exclusively descriptive in aim (that is, to measure and report the incremental costs and incremental effects for a particular comparison) then they will probably be rather uninformative material for a realist review. For these reasons, an explanatory or theory-building review of economic evidence should probably seek to include virtually all types of health economic study from (non-comparative) cost-of-illness studies through to regression-based cost and resource use analyses.

Discussion

Conducting high quality systematic reviews of any research question, in any topic area, requires a great deal of time and effort, and increasingly specialist research skills. In *some circumstances*, the 'pay-back' from such efforts can be generalisable and credible knowledge with wide probable application in a variety of policy and practice settings. It should also not be forgotten that systematic reviews also answer the more basic and useful questions of how much research has been conducted relating to a particular question, what is the quality of that research, and what are the major remaining 'knowledge gaps'.

The starting point for any consideration of whether systematic reviews of a particular type of research evidence ('economic evidence') or particular a class of studies (economic evaluations) are useful, should rest on whether the likely benefits will outweigh the research effort (and costs) involved. It might seem obvious but, on the benefit side, this critically depends on the number, quality and nature of studies which have addressed the

same or similar research questions. For example, a systematic review examining the research question "Is drug X more effective than drug Y (at improving outcomes P and Q) for patients with disease A?", will be ultimately pointless if there are not a number of clinical trials which have addressed this question. It follows that systematic reviews of economic evaluations can shed light on little more than whether the ratio of incremental costs and incremental effects for two or more policy or treatment comparators are consistent (and consistently favourable or undesirable) between times and places, since this is the only question that economic evaluations explicitly aim to answer. Yet, as this paper argues, there are many reasons why incremental cost-effectiveness will be likely to vary (and legitimately to vary) between different economic evaluations means that in most cases a consistent pattern of ICERs is unlikely to arise.

Therefore, at the centre of any discussion of the value of systematic reviews of economic evidence is a tension between economic evaluations as a primarily decision-informing or knowledge-generating method, and whether they can successfully be both. Many health economists view economic evaluations as explicitly decision-informing, and therefore view their findings as inherently jurisdiction-specific and time-limited. To this end they have increasingly advocated the decision model as a vehicle for such analysis. The contrasting assumption underlying most clinical effectiveness reviews is that the knowledge generated will have wide applicability to similarly diagnosed and managed patients elsewhere and in the future (for example, based on reliable knowledge of the underlying biological mechanisms). The overtly knowledge-generating ethos of the Cochrane Collaboration - indeed its very existence as a global rather than a national endeavour - illustrates this tension; Cochrane reviewers are encouraged to "be cautious about reaching conclusions about implications for practice and they should avoid making recommendations." (Cochrane Handbook) Thus, the applicability of evidence to real decision making situations - when faced with a particular patient group, a specific range of intervention comparators, and perhaps a given budget constraint and assumed willingness to pay for health benefits – is inherently limited when the goal of research synthesis is primarily to preserve some generalisable fact about the underlying effectiveness of a treatment.

In areas of health care like public health and health promotion, where effectiveness is more complex and contingent, for example upon on the specific combination of elements in an

intervention, and/or its interaction with different community and organisational contexts, it has become increasingly recognised that only asking whether an intervention "is effective" is almost pointless. Instead, it can so confidently be predicted that complex interventions will work in some places and not others, and with some individuals and not others, that it makes much more sense to ask from the beginning "how and why" an intervention is or is not effective in different circumstances (or when it its components are configured or implemented in different ways) (Jackson, Waters, & for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce 2005).

I have also set out three probably useful purposes for reviews of economic evaluations. The first is just good research practice; to see what methods others have used to answer the same or a similar research question. There may be sufficient evidence in such studies, particularly their sensitivity analyses, to indicate the key variables which appear to impact on cost-effectiveness (and conversely, those that do not at all) and thus inform the resource use data collection plans for a trial or the structure and parameters for a decision model.

An unavoidable weakness of my argument is that it has not itself been based on a thorough systematic review of systematic reviews of economic evaluations. Given the current numbers of such reviews, such a task would be enormous. However, there may be groups of systematic reviews, of economic evaluations on the same technology or patient group, and it would be interesting to see if there are sufficient consistencies in the review methods used, and in the reported results for them to reliably and convincingly inform decision makers. Funding will be sought to conduct a review of reviews in order to more properly test the "utility or futility" question of this paper.

Conclusion

It is practically a truism that evidence-based policy making cannot be driven by (and, indeed, cannot be seen to be driven by) the results of individual empirical studies. Instead, some kind of syntheses of evidence from numbers of research studies should be an integral part of the policy making process. At the same time, for those who recognise the inescapable reality of opportunity costs in policy choices, it is also essential that "economic evidence" plays some part in the decision making process. In health policy making, these two principles appear to have resulted in the widespread presumption that conducting

systematic reviews of economic evaluations provide a valuable basis either for decision making or, somehow, knowledge generation.

This paper has made the case that while increasingly popular, systematic reviews of economic evaluations are usually futile as an input to evidence-based policy making. This conclusion relies in part on the fact that, as practised in the health field, empirical economic evaluations have become a purely descriptive method (define comparators; measure and value inputs; measure and value outcomes; compare). If economic evaluations were to become more explanatory in their intent and methods, and sought more explicitly to explain *how* different levels and configurations of resources *cause* different levels and combinations of outcomes, then I believe systematic reviews of economic evaluations might generate more useful knowledge for policy makers. This call in fact closely echoes one of the conclusions of one of the first papers to discuss the feasibility of "secondary economic analyses", when they suggested that "Progress may require a more coherent *theoretical* framework linked to cost and production function theory" (emphasis added, p.163 of (Jefferson, Mugford, & Gray 1996)). It is also consistent with increasing methodological confidence in the merits of using more theory-driven and context-sensitive methods of systematic review for understanding the variable effectiveness of health interventions in different places and populations (Dixon-Woods, Agarwal, Jones, Young, & Sutton 2005;Greenhalgh & Kristjansson, V 2007;Pawson, Greenhalgh, Harvey, & Walshe 2005).

At the very least, commissioners of evidence syntheses which include a systematic review of 'economic evidence' need to be more explicit about its purpose (e.g. to inform a particular decision, to inform a decision model), its specific aims (review questions), and link these to the scope of the review (just economic evaluations or a broader range of economic study types), and consider whether resources might be better directly invested in a decision model based synthesis of evidence.

References

Birch, S. 1997, "As a matter of fact: evidence-based decision-making unplugged", *Health Economics*, vol. 6, pp. 547-559.

Birch, S. 2002, "Making the problem fit the solution: evidence-based decision making and "Dolly" economics," in *Evidence-based Health Economics: from effectiveness to efficiency in systematic review*, C. Donaldson, M. Mugford, & L. Vale, eds., BMJ Books, London, pp. 133-147.

Birch, S. & Gafni, A. 2006, "The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold", *Journal of Health Services Research and Policy*, vol. 11, pp. 46-51.

Birch, S. & Gafni, A. 2003, "Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results", *Health Policy*, vol. 64, pp. 207-219.

Buxton, M., Drummond, M. F., van Hout, B. A., Prince, R. L., Sheldon, T. A., Szucs, T., & Vray, M. 1997, "Modelling in economic evaluation: an unavoidable fact of life", *Health Economics*, vol. 6, pp. 217-227.

Carande-Kulis, V., Maciosek, M., Briss, P., Teutsch, S., Zaza, S., Truman, B., Messonnier, M., Papaioanou, M., Harris, J., & Fielding, J. 2000, "Methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services", *American Journal of Preventive Medicine*, vol. 18, no. 1S, pp. 75-91.

Coast, J., Hensher, M., Mulligan, J., Sheppard, S., & Jones, J. 2000, "Conceptual and practical difficulties with the economic evaluation of health services developments", *Journal of Health Services Research and Policy*, vol. 5, pp. 42-48.

Cooper, N., Coyle, D., Abrams, K., Mugford, M., & Sutton, A. 2005, "Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997", *Journal of Health Services Research and Policy*, vol. 10, pp. 245-250.

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. 2005, "Synthesising qualitative and quantitative evidence: a review of possible methods", *Journal of Health Services Research and Policy*, vol. 10, pp. 45-53.

Donaldson, C., Mugford, M., & Vale, L. 2002, *Evidence-based Health Economics: from effectiveness to efficiency in systematic review* BMJ Books, London.

Drummond, M. 2002, "Evidence-based medicine meets economic evaluation - an agenda for research," in *Evidence-based Health Economics: from effectiveness to efficiency in systematic review*, C. Donaldson, M. Mugford, & L. Vale, eds., BMJ Books, London, pp. 148-154.

Drummond, M. & Barbieri, M. W. J. 2005, "Analytic choices in economic models of treatments for rheumatoid arthritis. What makes a difference?", *Medical Decision Making*, vol. 25, pp. 520-533.

Drummond, M. & Pang, F. 2001, "Transferability of economic evaluation results," in *Economic evaluation in health care: merging theory with practice*, M. Drummond & A. McGuire, eds., Oxford University Press, Oxford, pp. 256-276.

Drummond, M., Sculpher, M., Torrance, G. W., O'Brien, B., & Stoddart, G. L. 2005, *Methods for the Economic Evaluation of Health Care Programmes* Oxford University Press, New York.

Evers, S., Goossens, M., de Vet, H., van Tulder, M., & Ament, A. 2005, "Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria", *International Journal of Technology Assessment in Health Care*, vol. 21, no. 1, pp. 240-245.

Greenhalgh, T. & Kristjansson, E. R., V 2007, "Realist review to understand the efficacy of school feeding programmes", *British Medical Journal*, vol. 335, pp. 858-861.

Jackson, N., Waters, E., & for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce 2005, *Guidelines for systematic reviews of health promotion and public health interventions*, Deakin University, for the Cochrane Health Promotion and Public Health Field, Melbourne, Australia.

Jefferson, T., Mugford, M., & Gray, A. 1996, "An exercise on the feasibility of carrying out secondary economic analyses", *Health Economics*, vol. 5, pp. 155-165.

Jefferson, T., Vale, L., & Demicheli, V. 2002, "Methodological quality of economic evaluations of health care interventions - evidence from systematic reviews," in *Evidence-based Health Economics: from effectiveness to efficiency in systematic review*, C. Donaldson, M. Mugford, & L. Vale, eds., BMJ Books, London, pp. 67-88.

Kraemer, H., Frank, E., & Kupfer, D. 2006, "Moderators of treatment outcomes: Clinical, research, and policy importance", *Journal of the American Medical Association*, vol. 296, pp. 1286-1289.

Kravitz, R., Duan, N., & Braslow, J. 2004, "Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages", *The Milbank Quarterly*, vol. 82, no. 4, pp. 661-687.

Kristiansen, I. & Gosden, T. 2002, "Evaluating economic interventions: a role for non-randomised designs?," in *Evidence-based Health Economics: from effectiveness to efficiency in systematic review*, C. Donaldson, M. Mugford, & L. Vale, eds., BMJ Books, London, pp. 114-126.

Lavis, J., Davies, H., Oxman, A., Denis, J.-L., Golden-Biddle, K., & Ferlie, E. 2005, "Towards systematic reviews that inform health care management and policy-making", *Journal of Health Services Research and Policy*, vol. 10, no. Suppl. 1, pp. 35-48.

Mugford, M. 2002, "Reviewing economic evidence alongside systematic reviews of effectiveness: example of neonatal exogenous surfactant," in *Evidence-based Health Economics: from effectiveness to efficiency in systematic review*, C. Donaldson, M. Mugford, & L. Vale, eds., BMJ Books, London, pp. 25-37.

Nixon, J., Khan, K. S., & Kleijnen, J. 2001, "Summarising economic evaluations in systematic reviews: a new approach", *British Medical Journal*, vol. 322, pp. 1596-1598.

Oliver, A. 2003, "Accounting for the missing opportunity costs in incremental cost-outcome analysis", *Applied Health Economics and Health Policy*, vol. 1, pp. 191-196.

Pawson, R. 2006, *Evidence-based policy: A realist perspective* SAGE Publications, London.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. 2005, "Realist review - a new method of systematic review designed for complex policy interventions", *Journal of Health Services Research and Policy*, vol. 10, p. S1:21-S1:34.

Petticrew, M. 2001, "Systematic reviews from astronomy to zoology: myths and misconceptions", *BMJ*, vol. 322, no. 7278, pp. 98-101.

Pignone, M., Saha, S., Hoerger, T., Lohr, K., Teutsch, S., & Mandelblatt, J. 2005, "Challenges in systematic reviews of economic evaluations", *Annals of Internal Medicine*, vol. 142, no. 12, pp. 1073-1079.

Popay, J. e. 2006, *Moving beyond effectiveness in evidence synthesis: Methodological issues in the synthesis of diverse sources of evidence*, National Institute for Health and Clinical Excellence, London.

Russell, L. B. 2007, "Is all cost-effectiveness analysis local?", *Medical Decision Making*, vol. 27, no. 3, pp. 231-232.

Sculpher, M., Claxton, K., Drummond, M., & McCabe, C. 2006, "Whither trial-based economic evaluation for health care decision making?", *Health Economics*, vol. 15, no. 7, pp. 677-687.

Sculpher, M. & Drummond, M. 2006, "Analysis Sans Frontieres: Can we ever make economic evaluations generalisable across jurisdictions?", *Pharmacoeconomics*, vol. 24, no. 11, pp. 1087-1099.

Sculpher, M., Pang, F., Manca, A., Drummond, M., Golder, S., Urdahl, H., Davies, L., & Eastwood, A. 2004, "Generalisability in economic evaluation studies in healthcare: a review and case studies", *Health Technology Assessment*, vol. 8, no. 49.

Sculpher, M., Claxton, K., & Akehurst, R. L. 2005, "It's just evaluation for decision-making: recent developments in, and challenges for, cost-effectiveness analysis," in *Health Policy and Economics: Opportunities and Challenges*, P. Smith, M. Sculpher, & L. Ginnelly, eds., Open University Press, Buckingham, pp. 8-41.

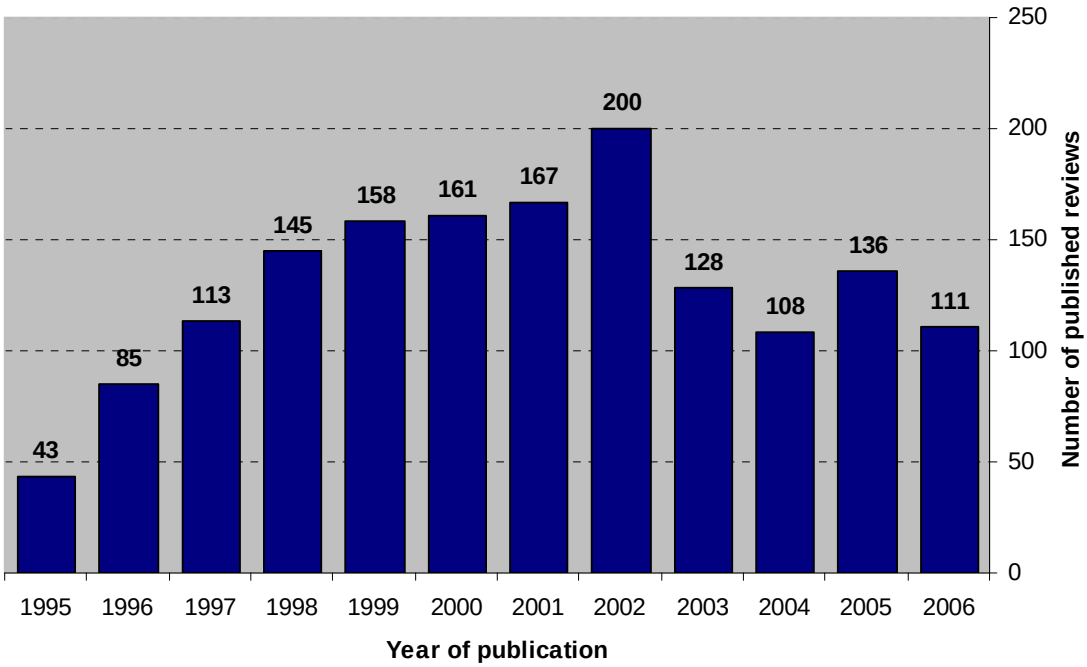
Table 1. Different health agency's requirements for reviews of economic studies

Agency	Requirement	Stated purpose or review question
For health technology appraisal:		
National Institute for Health and Clinical Excellence, UK	Evidence on cost-effectiveness "also includes the findings of existing published economics literature." (nb. That this is expected to be in the form of a systematic review is indicated in the required report sections for Technology Assessment Reports for NICE.)	None stated
Pharmaceutical Benefits Advisory Committee, Australia	"Present the results of a search of the literature for reports of economic evaluations of similar decision analyses (in terms of similarity to the treatment algorithm and/or the proposed and similar drugs). Where the submission's model is different from the literature-sourced models, explain the basis for the selection of the submission's approach." (Under section D3. "Structure and rationale for economic evaluation")[b]	Not stated, but implicitly to justify and contextualise whatever economic evaluation model and analytical approach (CUA, CEA or CBA) is adopted by the manufacturer's submission.
Canadian Agency for Drugs and Technologies in Health	"Discuss existing economic studies that address the same technology, and similar study question(s). Include a summary of methods and results of reviewed studies" and "Comment on the relevance and generalizability of the results of the reviewed studies to the target audience" Use NHS CRD Guidance for review methods	"To summarize the available knowledge in a systematic way that will be useful for decision makers and researchers"
Other health policy agencies:		

Agency	Requirement	Stated purpose or review question
National Institute for Health and Clinical Excellence (Centre for Public Health Excellence), UK	"A systematic review of the published economic literature should be carried out …" "A thorough systematic review should be attempted, but if there is a large amount of economic evidence, it may be necessary to limit the search." "Papers identified for inclusion should be critically appraised using a validated checklist [and] a commentary should be presented on the quality of each paper"	"… to ensure that no economic evaluations are missed during searches undertaken in the review of effectiveness" (sic) and to determine "if the published evidence is so reliable that further analysis would be superfluous"
US Preventive Services Task Force	"Systematic reviews of economic evaluations are completed for those interventions that are either Strongly Recommended or Recommended by the Task Force on Community Preventive Services"	

Sources:
a NICE: Guide to the Methods of Technology Appraisal, April 2004 (para 3.3.2, p.12).
b PBAC: Guidelines for preparing submissions to the PBAC, Section D on Economic evaluation for the main indication
c CADTH: Guidelines for the Economic Evaluation of Health Rechnologies: Canada (3rd Ed.), 2006
d NICE: Methods for development of NICE public health guidance, March 2006
e US PSTF: Guide to Community Preventive Services (Methods for Systematic Reviews of Economic Evaluations published in *Am J of Prev Medicine* in 2000.(Carande-Kulis, Maciosek, Briss, Teutsch, Zaza, Truman, Messonnier, Papaioanou, Harris, & Fielding 2000)

Figure 1. Published reviews of economic evaluations* in NHS EED 1995 to 2006



Source: *NHS Economic Evaluation Database* (at CRD, York); searched 24th August 2007, using search term "review", limited to 1985 to 2007. Papers initially included on the basis of a CRD reviewer having designated them as Record type "Review of economic evaluations". A second sub-selection was then made on the basis of paper titles (in order to exclude many which were clearly primarily reviews of effectiveness).

*May actually include reviews of a broader range of economic studies, such as costing studies or cost-of-illness studies.