Draft paper: do not quote or reference without authors' permission

Who Does the Numbers? The Role of Independent Technology Assessment to Inform Health Systems' Decision Making about the Funding of Health Technologies

Marco Barbieri[1]
Neil Hawkins[2]
Mark Sculpher[2]

1. Universitat Pompeu Fabra, Barcelona, Spain. 2. Centre for Health Economics, University of York, UK.

Abstract

There is an increasing number of health care systems which use health technology assessment, including economic evaluation, to inform decisions about the funding/reimbursement/coverage of (mainly new) health technologies. In most systems, the technology assessment is undertaken by the manufacturer of the technology but, in a few, an independent technology assessment group is responsible for this research. In the UK, the National Institute for Health and Clinical Excellence (NICE) used independent assessment for all its technology appraisals between 1999 and the end of 2005. After this point, a new Single Technology Assessment (STA) programme was instituted to cover a large proportion of technologies (probably most pharmaceuticals) on the basis of technology assessments completed by the manufacturers but critically reviewed by an independent evidence review groups. This paper addresses the general question of the role of independent assessment in this from of decision making. It starts by reviewing the requirements of technology assessment in general, and economic evaluation on particular, to support decision making, and considers the extent to which manufacturer-based or independent assessment is likely to meet these requirements. It is concluded that, if a system relies on manufacturer assessment, there is a risk of misleading analysis which may not be fully addressed by a rigorous independent critical review. This is particularly the case regarding the identification and synthesis of relevant evidence, the consideration of heterogeneity and the characterisation of uncertainty. The paper also addresses the question of whether the two forms of technology assessment differ in their impact on decision making using a comparison of the decisions made by NICE (under their conventional multiple technology appraisal system) and the Scottish Medicines Consortium (SMC) which relies on manufacturer analyses. The comparison shows that, in terms of the detailed decisions, there are important differences between the two bodies, with NICE generally placing more restrictions of the use of technologies. It is concluded that there are important ways of making manufacturer-based assessments more informative to decision making including detailed and prescriptive methods guidelines and the placement of the 'burden of proof' on manufacturers, but there remain some compelling reasons for the use of independent technology assessment.

Paper presented at the Health Economists' Study Group, Birmingham, January 2007

1. Introduction

Many health care systems now require technology assessment as a key input into a formal decision-making process about whether to reimburse/fund/cover new health technologies.[1 2] The process of 'technology assessment' typically takes the form of a review of relevant clinical and economic evidence and the development of a cost-effectiveness model. In most systems the focus of technology assessment is to inform decisions about new pharmaceutical products: the publicly-funded systems of Australia [3] and Ontario, Canada[4] initiated the use of technology assessment for this purpose, but now many systems have followed suit. In some systems – notably the National Institute for Health and Clinical Excellence (NICE) in the UK – technology assessment support decisions on other technologies such as medical devices and surgical procedures, in addition to pharmaceuticals.

As these technology assessment and decision-making approaches embed themselves in health care systems, and are considered for use in others, an important issue to address is who undertakes the technology assessments. In most systems, the technology assessment (i.e. the systematic reviews and modelling) is undertaken by the manufacturer although it is usually critically reviewed by experts within the system or by an independent group. In a small subset of systems, however, there is a role for technology assessment independent of both the manufacturer and decision-maker – that is, conducted by a group (usually academic) commissioned at arms length by the health care system or the decision-making agency and with no association with manufacturer. Such arrangements exist with, for example, NICE in the UK and the Medical Services Advisory Committee which considers non-pharmaceutical technologies for the Australian health care system. There is an issue regarding the meaning of the term 'independent' in this context, and the contractual positions of the various groups doing this sort of work internationally are likely to vary. However, it seems reasonable to suggest that the status of 'independent' would require comparable arrangements to those which usually exist in publishing research in peer-reviewed journals: freedom to undertake and publish research without constraints imposed by the funder (whether health system or commercial).

The role of independent technology assessment in this form of decision making has recently been brought to the fore following NICE's decision to limit its use in their technology

appraisal process. From its advent in 1999 until the end of 2005, NICE's standard process involved an independent academic group undertaking a technology assessment report (TAR) to provide the key input into the Appraisal Committee's deliberations. The TAR consists of a review of relevant clinical and economic evidence, a critical assessment of one or more manufacturer submissions (which includes the manufacturers' own reviews and models) and, usually, the development of a cost-effectiveness model. These arrangements still exist for some technologies as part of NICE's multiple technology appraisal (MTA) process. However, from 2006, many technologies (particularly newly licensed pharmaceuticals) have entered a new single technology appraisal (STA) process.[5] These arrangements are similar to those in many systems internationally in that the only reviews and analysis informing decisions-making are undertaken by the manufacturer, although these are accompanied by a critical review of the latter's submission by an independent evidence review group (ERG).

It is important to address the question of whether independent technology assessment should be a feature of decision-making processes regarding health technologies. In doing so, this will inform decisions as to whether health care systems should require independent technology assessment when developing new decision-making processes.. It will also inform the specific debate in the UK about the balance between the MTA and STA processes at NICE.

A key issue to consider in addressing this question is the evidence regarding whether independent assessment makes a difference to decisions – that is, are decision-making authorities more or less likely to support the use of a technology if they have access to an independent assessment? In time, some evidence will be provided by studying the decisions of the NICE Appraisal Committee and comparing those made under the MTA system with those from STAs. However, such a comparison will inevitably be hampered by the fact that different technologies have been considered under the two sets of arrangements. Arguably, a more informative comparison is facilitated by what is effectively a natural experiment between the standard (MTA) arrangements at NICE (covering England and Wales) and the process adopted by the Scottish Medicines Consortium (SMC) in Scotland. Full details of the SMC's assessment and decision-making processes, in the comparison with those of NICE, are provided elsewhere.[6] In general terms, it shares many of the features of the

NICE STA process (and similar ones internationally), in particular the absence of independent technology assessment. Such a comparison is provided here for those technologies and indications considered by both organisations and this updates and extends the comparison undertaken by Cairns.[6]

A key component of the technology assessment is an economic evaluation and the next section of the paper describes the necessary features required for economic evaluation to inform decision making about health care technologies. Section 3 considers the implications of independent assessment for these features. Section 4 presents the comparison of NICE (MTA) and SMC decisions for a range of technologies. Section 5 provides a discussion which considers implications for health care systems internationally.

2. The requirements of economic evaluation to support decision making

The advent of economic evaluation studies to inform explicit decisions about the use of health care technologies in particular populations by specific decision makers at a point in time has raised a number of distinct issues about research methods. Various papers have suggested some general requirements of economic evaluation in this context.[7 8] Some of these are particularly relevant to the consideration of the role of independent assessment to inform decision making, and are summarized below.

- *Appropriate specification of the decision problem.* Any economic analysis informing decisions needs to be clear about which patient population(s) and indications are being considered. Once this is clear, all relevant options need to be defined – that is, the technologies of interest together with the full list of comparators. The latter should include not just the 'mostly widely used' or 'the most effective' but all options that could feasibility be used in the system. The definition of 'alternative options' may well need to include the specification of alternative sequences of interventions or diagnostic tests and different stopping and starting rules for treatments.
- *All relevant evidence.* In order to inform social decisions about resource allocation, the available evidence base needs to be identified and synthesised in a systematic manner. The inclusion of only a subset of relevant evidence represents a partial analysis with

potentially misleading results. This relates not only to measures of relative treatment effect, but also to the full range of other parameters in an economic analysis. This requirement is consistent with the principles of evidence-based medicine [9].

- *Appropriate characterisation of uncertainty.* Uncertainty in economic evaluation studies, in terms of parameters and structural features, needs to be quantified and expressed in ways relevant to decision making. Given a system's objective of maximizing population health gain subject to a budget constraint, there is little role for conventional statistical inference and similar approaches based on arbitrary error probabilities despite a continued role for these methods when new pharmaceuticals are licensed.[10] It does not follow, however, that uncertainty is unimportant in decision making, and the key question is whether there is sufficient evidence to support a decision to fund a particular technology. 'Sufficiency' rests on whether additional research would be potentially efficient (the value of further research exceeds the cost of that research). Decision making should consider the extent of decision uncertainty and the potential value of additional research, although the specifics of the ultimate decision will vary because agencies differ in the control they exert over public research funding and in the remit they have to require manufacturers to undertaken further research.[11]
- *Focusing technologies on sub-groups in which cost-effectiveness is maximized.* Patients inevitably vary in the potential costs and benefits they derive from medical interventions. To some extent, this variation can be 'explained' on the basis of patients' characteristics which are known at the point at which treatment decisions are taken. This can lead to a heterogeneity in the costs, benefits and, hence, cost-effectiveness of interventions between different sub-groups of patients. This heterogeneity is not just a result of differences in the effectiveness of interventions, it can also reflect variation in the underlying risks of clinical events, prognosis conditional on a clinical event, costs and preferences. In principle, heterogeneity can be reflected in decisions by specifying those sub-groups of a patient population in which a particular intervention is cost-effective (this may often restrict use from that indicated in a pharmaceutical's marketing authorization). Indeed, assuming such restrictions can be operationalised in routine practice, failing to reflect heterogeneity in decisions can impose 'costs' in terms of health benefits to other patients forgone and/or resources used inefficiently.[12]

In addition to these considerations which largely relate to methods, there is a further requirement for technology assessment, and that is to be timely. That is, the assessment needs to be available to inform decisions at the point at which they are taken. With respect to new pharmaceuticals, health systems have to consider whether they will provide funding/reimbursement as close as possible to launch following a marketing authorization being granted for the product. Although the regulation of other technologies (e.g. medical devices and surgical procedures) differs from pharmaceuticals, decisions regarding their adoption and the need for additional research will typically be required as soon as manufacturers start marketing their products to clinicians. It can be argued that the timing of decisions about technologies which are already provided is somewhat more flexible. However, such decisions are (or should be) inextricably linked to decisions about new technologies given the need to identify appropriate technologies to 'displace' from existing budget commitments to 'make room' for the funding of new technologies.

3. What are the implications of (the absence of) independent assessment?

It seems reasonable to consider the extent to which independent technology assessment will increase or decrease the chance of meeting the requirements detailed in Section 2. This is done below for each requirement.

3.1 Specifying the decision problem

In the context of decision making about new technologies, the decision problem should be specified in advance and be explicit in any analysis. In the many systems, this is left to the manufacturer to define within their economic analysis. Although, in the case of pharmaceuticals, the license defines the relevant patient population, there is considerable flexibility for the manufacturer in specifying their decision problem even if they are made aware of the principles outlined in Section 2.

There seems to be a strong case, however, for the health system having control of the specification of the decision problem. Firstly, the system should be able to form a view on how a new technology would be used based on its licence and clinical guidelines and advice in that system. Secondly, the system should have the best information about the existing

ways of managing the particular patient group (i.e. the comparators) in terms of individual management options, sequences and strategies. In the UK, the SMC and NICE (STA) form an interesting comparison in this regard. The SMC leaves the specification of the decision problem (as long as it is within license) to the manufacturer with the vague guidance that "the comparator the SMC is interested in is the one that will most likely be replaced if the medicine under consideration is accepted by the SMC for use in Scotland". NICE uses a scoping phase in an attempt to reach a shared view with the manufacturer on the appropriate question.

What implications does this have for independent assessment? It is arguable that, given freedom to specify the decision problem, manufacturers will have the incentive to select comparators which reflect favourably on their product. Of course, this can be countered with an effective review process which correctly identifies a potentially misleading analysis on the basis on an inappropriate decision question. However, the risk of misleading analysis can be reduced by the system clearly specifying the appropriate decision problem at the outset and only accepting analyses from the manufacturer which are consistent with it. The only situation where there may be a case for independent assessment with respect to an appropriate specification of the decision problem is when the decision maker is unwilling or unable to accept responsibility for defining the decision problem (in which case independence may increase the likelihood of an appropriate set of comparators).

3.2 Evidence base

The appropriate evidence base for an economic evaluation cannot be specified before the work is undertaken as it is an inextricable feature of the research itself. Therefore, whether the manufacturer or an independent group undertake the assessment, there is inevitably some degree for judgement on the part of the researchers regarding the identification, extraction and synthesis of the evidence. It may be argued that the manufacturer has an interest in selecting evidence which, when incorporated into a model, is likely to bolster the cost-effectiveness of their product. However, this risk can be ameliorated in two ways which fall short of independent assessment. The first is by the decision maker, through its methods guidelines, providing a clear and ambiguous statement of the principles relating to

the evidence – namely, the need for a complete, transparent and reproducible systematic review of all evidence, and an explicit synthesis. The second is for the decision maker to undertake or commission a rigorous review of all aspects of the manufacturer's assessment including their evidence base and to identify research which falls short in implanting the principles outlined in the methods guidelines.

There may be a case that, unless the decision maker is able to commission an independent systematic review, it will never know whether the manufacturer's evidence base is reliable. However, there is perhaps a more compelling argument for independent assessment in the context of identifying and synthesising the relevant evidence base. For some assessments, more than one new technology will be included in the comparison. This might be the case, for example, if two new pharmaceuticals within a class are licensed for the same patient group at a similar point of time. For example etanercept and efalizumab were licensed for the treatment of moderate to severe psoriasis in the UK are a similar time. In this situation, it is very likely that the full extent of the evidence base relating to the new technologies is not in the public domain. In which case the manufacturer is the only source of this unpublished evidence, but only for their own product. In other words, in the case of two new technologies, neither manufacturer has access to the full extent of the available evidence to include in their economic evaluation. Indeed, this may not just apply to new technologies. Even when a comparator technology has been used in the system for a while, if its manufacturer has any data relating to the product which is not in the public domain, it will not be available to the manufacturers of new intervention: inevitably only a fraction of the available data, running to tens of thousands of pages and many datasets, are published. In these circumstances, if the decision maker requires each manufacturer (new and existing technologies) to submit details of all its evidence (even if on a 'commercial in confidence' basis), then an independent assessment group commissioned by the decision maker can be given assess to this information for use in its analyses. As such, the decision maker becomes a clearing house for information.

3.3 The handling of uncertainty

There are many sources of uncertainty in cost-effectiveness analyses. In addition to the uncertainties surrounding estimates of relative treatment effects and other individual model

parameters, cost-effectiveness analyses typically require a number of extrapolations, each of which may be associated with considerable uncertainty. These extrapolations may include extrapolation of estimates from clinical trials to the general population, extrapolation in the form of indirect comparisons where direct estimates comparing all relevant treatments are not available or sufficient, extrapolation from reported trial endpoints to all relevant costs and effects and extrapolation over the relevant time horizon.

It is difficult to quantify the combined uncertainty arising from the imprecision with which parameters are estimated and these extrapolations, and the full uncertainties are generally not fully incorporated into final estimates of decision uncertainty. In principle, the 'structural' uncertainties associated with extrapolations, could be parameterised, dealt with using probabilistic sensitivity analysis and summarised as a measure of decision uncertainty and the value of additional research.[13] In reality, structural uncertainty is likely to be reflected in 'scenario analysis' in which cost-effectiveness is presented under different assumptions about extrapolation. The appropriate characterisation and quantification of these different types of uncertainty is a key element of decision making in that it is central to defining when there is sufficient evidence to support the use of a technology and when additional research should be undertaken.

If is it accepted that analyses provided by the manufacturers may seek to exaggerate the cost-effectiveness of their products, then one aspect of this is that they will tend to under-estimate uncertainty. This is most likely to arise through the failure to reflect all the structural uncertainties in analyses. As for the specification of the decision problem and details of the evidence base, a failure to provide a reasonable characterisation of uncertainty may be identified through thorough critical review by the decision maker. However, reviewing such features of an analysis is not straightforward, and there are inevitably judgements required about, for example, the selection and specification of scenarios.

3.4 Heterogeneity

As discussed in Section 2, the cost-effectiveness of treatments may vary between identifiable patient sub-groups. For any health care system seeking to maximise some population health gain from available resources, there is a strong case to reflect this heterogeneity in decisions

about the technologies by restricting their use to sub-groups in which they have been shown to be cost-effective. When manufacturers are undertaking economic evaluation to inform decisions about their products, there is every incentive to identify cost-effective sub-groups when the product is not cost-effective when averaged across the range of different sub-groups in a patient population. This is because, if they ignore heterogeneity in this context, they risk a negative decision for all patients. However, if heterogeneity exists, it should always be reflected in economic analyses by reporting the cost-effectiveness of the options being compared separately for each sub-group of patients. In situations where the manufacturer feels able to show their product is cost-effective when averaged across patients, they will have no incentive to undertake a thorough analysis of heterogeneity to identify sub-groups in which their product is *not* cost-effective. Again, thorough review of a company submission may identify potential heterogeneity which has apparently not been investigated. However, sources of potential heterogeneity are most reliably identified using individual patient data from trials and other studies and, in most instances, these will only be available to the manufacturers. There may, therefore, be good reason to think that an independent assessment, based on the companies' evidence, will provide an impartial assessment of heterogeneity.

3.5 Timing

The above suggests that some of the requirements of economic evaluation for decision making are unlikely to be satisfied for all technologies when manufacturers are the sole source of technology assessment. However, it can be argued that the use of an independent assessment introduces time delays which hinder the undoubted need to inform decisions in a timely manner. This seems to be a particularly compelling reason why NICE has moved to their STA process for a proportion of technology appraisals. However, for several reasons, it is not clear that the use of an independent assessment inevitably prolongs the period before a decision. Firstly, there is no evidence that the assessment research itself is undertaken more rapidly by a manufacturer than a independent group, it is just that the company is likely to start work earlier. But this is essentially a process feature: an independent group could be commissioned as soon as the company notifies the health care system that it is preparing to launch the product. Indeed, if the health system's review of

company submission identifies limitations, the iterative process of seeking re-submissions may well lengthen the process.

Secondly, there is no 'correct' period of time for an independent assessment. Arguably, at least in some instances, an independent assessment lasting as little as 3 months, in addition to a company submission, may be preferable to a situation where the only evidence informing decision makers comes from the company together with a short period to review that submission. Thirdly, the time duration until a decision is probably more a function of process factors such as the degree of consultation with manufacturers and other stakeholders, the opportunity for manufacturers to appeal and the use of expert advisors. In the case of the STA process at NICE, there is little evidence to date to suggest that the time to a decision is markedly shorter than under the former (MTA) process.

4. Does independent assessment make a difference to decisions?

4.1 Background and methods

In considering the role of independent assessment in health service decision making, it is important to address the question as to whether the use of such research makes a difference to the ultimate decisions. It is useful to compare the results of SMC and the standard NICE (MTA) technology appraisals as they represent comparable examples of the two options of basing technology appraisal on submissions by manufacturers or independent reviews commissioned by decision-makers. As they are making decisions for the same populations covered by similar health-care systems, any differences in recommendations may be ascribed to differences in process.

A similar study, in which the SMC and NICE (MTA) decisions were compared, has recently been published by Cairns.[6] This concluded that 'there are important differences between the approaches of SMC and NICE relating primarily to the timing of the review on clinical and cost-effectiveness' and 'where direct comparison between NICE and SMC recommendations is meaningful, the advice has been similar'. In order to extend the Cairns study, an attempt is made to give more details on the guidance compared, focusing on the potential differences in restrictions and the importance of patient heterogeneity. To achieve this, the main

recommendations for each drug have been categorized on the basis of Raftery's[14] classification in order to analyse the key elements of the final decision. To update the Cairns analysis, new appraisals for the same product and indication, issued both by NICE and SMC, are included.

4.2 Summary of results

A total of 25 cases have been identified where it is possible to make a comparison between the SMC and NICE (MTA) guidance. These include a total of 22 drugs and 18 indications. The comparative guidance is detailed in Table 1, and more details of the decisions are given in Appendix 1. There was a general agreement between NICE and SMC in terms of whether to accept or reject a medication. Out of the 25 cases where comparison is possible, both agencies accepted the intervention in 23 cases (although with restrictions in many circumstances) and both rejected the technology in 1 case (anakinra for rheumatoid arthritis). There has been only one case where the two institutes reached a different conclusion - that is, pimecrolimus for atopic dermatitis which was recommended as 2nd line treatment by NICE and rejected by SMC.

4.3 Differences in restrictions

Although, in the vast majority of comparisons, the two institutes ended up with similar general recommendations, there were some important differences with respect to restrictions on use and the level of detail given regarding which patient sub-groups should receive treatment. Differences in restrictions in use occurred in at least 10 comparisons: in 7 cases NICE appeared to be more restrictive, while in 3 cases SMC appeared more restrictive (Table 1). For example, in the assessment of risendronate for osteoporosis in post-menopausal women, NICE recommended the use of this medication only for secondary prevention of fractures and states "that choice between risedronate and other bisphosphonates should be based on tolerability". The SMC recommendation is less restrictive for risendronate which is approved for general use in treatment and prevention of osteoporosis in post-menopausal women, and it is considered convenient to the patient compared to other treatments.

In the assessment of imatinib for gastro-intestinal stomal tumours (GIST), NICE recommended continuation only when a response at 12 weeks is achieved, while SMC did not impose any restriction on duration of treatment. In the case of pegylated interferon (alfa-2a and alfa-2b), NICE recommended its use as combination therapy for patients with moderate or severe hepatitis C, while monotherapy is accepted only for individuals with mild hepatitis C. SMC did not make this distinction and pegylated interferon is approved both as montherapy and as combination therapy. Similarly, NICE appeared more restrictive in its recommendations for mycophenolate (renal transplant), olanzapine (bipolar disorder and schizophrenia) and teriparitide (osteoporosis). On the other hand, SMC imposed potentially more restrictions in the recommendations for methylphenidate (attention deficit and hyperactivity disorder), imatinib (chronic myeloid leukaemia) and clopidogrel (acute coronary syndrome). For example, SMC recommended that clopidrogel is only initiated during an inpatient stay (not stated by NICE) and that methylphenidate is only used as second line treatment in exceptional circumstances.

4.4 Use of the Raftery classification[14]

A more detailed analysis of the NICE and SMC guidance is presented in Table 2, where Raftery's classification has been used to categorise the main recommendations for each product. These categories are: use as detailed in the Summary of Product Characteristics (SPc) (i.e. without additional restrictions); use as 2nd line; use only if intolerant to another treatment; continue if response is achieved by a specified duration; confine to most severe patients; use least costly option; monitoring required; use by specialist only; do not use due to insufficient evidence (of clinical effectiveness); and do not use due to poor cost-effectiveness. Using this classification, the differences between NICE and SMC recommendations appear more evident and there are many cases where NICE and SMC do not fall into the same group. There are more circumstances in which NICE recommendations impose additional restrictions over SPc indications (in comparison with SMC), including a requirement for additional monitoring, or a more limited use (only if intolerant to other treatment or confine to sub-group of patients). In particular, only NICE makes specific recommendations regarding restricting use to patients intolerant to another treatment, continuing only if a defined response is obtained and to use the cheapest option.

4.5 Heterogeneity and sub-groups

These findings suggests a key distinction between SMC and NICE guidance. In general, NICE gives more details on patient sub-groups in which a product is (not) cost-effective. And there are situations where this sub-group analysis has a strong impact on the cost-effectiveness of the medication, given heterogeneity between patients. Using the example of medications for osteoporosis (alendronate, risendronate, teriparitide), a case is highlighted where patient heterogeneity is reflected in the NICE recommendations but not in the SMC guidance. Furthermore, the importance of this difference is quantified in terms of health to patients and resource costs on the NHS.

The independent assessment group informing the NICE decision developed a Markov model to estimate the cost-effectiveness of drugs for osteoporosis separately for women with different age and different risk factors.[15] It was found that, for all the assessed drugs, the incremental cost per quality-adjusted life year (QALY) gained improved substantially with increasing patient age and risk factors. For example, for women with a T-score (the number of standard deviations from the average bone mineral density of healthy young women) of -2.5 SD, the incremental cost-effectiveness ratio (ICER) ranged between £32,937 (age 50), £36,595 (age 60), £12,191 (age 70) and dominating (age 80) in the case of alendronate; £37,030 (age 50), £38,645 (age 60), £15,067 (age 70) and dominating (age 80) in the case of risendronate; and £91,657 (age 50), £102,418 (age 60), £43,827 (age 70) and £30,687 (age 80) in the case of teriparitide.

Lower ICERs were found for women with double or quadruple risk factors, for all ages and all drugs. Thus, medications for osteoporosis may be cost-effective for some sub-groups of women (mainly older and with higher baseline fracture risk) and not so for other patients. This was reflected in the NICE guidance that recommended the use of alendronate and risendronate for women 65 years and older or for women aged less than 65 years but with some additional age-independent risk factors. Similarly, teriparatide was recommended as a treatment option in women aged 65 years and older who have had an unsatisfactory response or intolerance to bisphosphonates. The SMC guidance, on the other hand, does not make a distinction between patient sub-groups and recommends these medications for all post-menopausal women (although teriparatide is confined to the most severe cases).

The independent assessment group informing NICE estimated that approximately 940,000 women aged 50 or more suffer from osteoporosis in the UK. The lifetime risk of vertebral fracture was estimated to be about one in three (about 310,000 women). Of these, 5% occur in the age group 50 to 59 and 15% in the age group 60-69. Thus, there will be approximately 15,500 osteoporotic women aged 50 to 59 with previous fractures and about 46,500 women aged 60 to 69. On the basis of the results in terms of incremental costs and incremental QALYs for alendronate and risendronate obtained from the NICE assessment group (Matt Stevenson, personal communication), we have compared the expected net monetary benefit[16] of treating all women (STA decision) and restricting treatment to specific sub-groups (NICE decision).

Table 3 shows the net monetary benefit results for women with previous fractures and T-score –2.5. Assuming a monetary value of £30,000 per QALY, if all women were treated, both alendronate and risendronate would generate positive net monetary benefits (£16,351 and £6,459 per patient, respectively), while teriparatide would be associated with an expected negative net benefit (-£132,581). If treatment is restricted to those sub-groups where treatment is cost-effective, both alendronate and risendronate would generate positive net monetary benefits (£18,183 and £9,350 per patient, respectively). Thus, if heterogeneity is ignored it would be concluded that both alendronate and risendronate should be given to all women suffering from established osteoporosis. This would, however, result in a loss of *populaton* net monetary benefit, compared to more selective use, of £1,831,790 and £2,891,184, for alendronate and risendronate, respectively.

Similar conclusions can be obtained in the case of women with higher risk factors for subsequent fractures (Table 4). In this situation, while alendronate and risendronate are associated with positive net monetary benefits for all age groups, the net monetary benefit (positive or negative) with teriparitide depends on the age group considered. If heterogeneity is ignored, teriparitide would not be recommended. If heterogeneity is taken into account, teriparitide would be recommend for patients aged over 70 with an increase in population net monetary benefit of £10,000,000. Although these values are likely to be overestimated (not all women with osteoporosis and previous fractures will be candidates for

alendronate or risendronate) and they do not take account of the distribution of patients between treatments, the potential importance and the impact of taking heterogeneity into account is evident. All these issues are explicitly addressed in the NICE guidance, while less attention on patient heterogeneity is given in the SMC recommendations.

5. Discussion

This paper has argued that manufacturers face an incentive to undertake economic analysis such that their choice of comparators, selection of evidence, consideration of heterogeneity and treatment of uncertainty can exaggerate the cost-effectiveness of their product. What evidence exists to support this viewpoint? Perhaps the most compelling is the Miners *et al* analysis which considered the cost-effectiveness analyses undertaken as part of 27 NICE technology appraisals in which is was possible to compare the ICER generated by the independent assessment group with that of the manufacturer.[17] Of the 54 pair-wise comparisons, 25 showed higher ICER estimates from the independent assessment group, 29 had the same ICERS from both the assessment group and the manufacturer and none had higher ICERs from the company.

Even if it is accepted that manufacturers can submit potentially misleading economic evaluations to decision-making bodies, it can be argued that this can be identified through a process of thorough critical review of the submission. This approach is used by many health care systems and is a feature of the NICE (STA) process where the critical review is undertaken by an independent Evidence Review Group. It is possible to identify factors which will increase the likelihood that such arrangements will be successful in limiting the number of potentially misleading analyses. The first is for the system to produce detailed and prescriptive methods guidance which define, as fully as possible, how it wants manufacturers to undertake economic evaluation, thus giving little room for judgement on the part of companies. The critical review of the manufacturer's submission is then undertaken against this methods guidance. This would be supported by the decision maker being as transparent as possible in describing its decisions and how they relate to the (lack of) evidence. Currently, however, most methods guidelines internationally are quite general and non-prescriptive,[2] and few systems provide complete details of how their decisions were made. The second is for decision makers to encourage a relationship with the manufacturers

which, in the context of a particular piece of analysis, facilitates an iterative line of communication between the two about appropriate analytic methods.

A third approach a decision maker could use to limit the risk of misleading analyses from companies is to change the balance of incentives by placing the 'burden of proof' about cost-effectiveness on the manufacturer. This is something approaching a UK criminal court's 'beyond all reasonable doubt' standard of proof. Consistent with this, the SMC note in their guidance to manufacturers that 'it is the responsibility of the manufacturer to clearly demonstrate the case for the cost-effectiveness of medicines submitted to the SMC'. In contrast, where decisions are based on independent technology assessment but commissioned by the decision-maker, the burden of proof that the final decision is 'appropriate' will lie with the decision-maker and something closer to a UK civil court's 'balance of probabilities' standard of proof is likely to be applied — NICE (MTA) note in their guidance to the appraisal process that "the evidence is considered by the Institute's Appraisal Committee, which reaches a judgement as to whether, on balance, the technology can be recommended as a cost-effective use of NHS resources in general". Placing the burden of proof on the manufacturer can influence the incentives they face by indicating that, if an analysis falls short of a convincing demonstration of cost-effectiveness, a negative decision may follow.

Setting up these types of arrangements to limit the risk of misleading analyses from manufacturers has it limitations, however. The first is that an iterative relationship between the decision maker and the manufacturer in which problems in the latter's analysis can be communicated and corrected takes time and threatens the requirement for timeliness in decision making. The second relates to the use of a thorough review of the company's submission against a detailed set of prescriptive methods guidelines, where such arrangements may make it difficult to maintain the burden of proof with the manufacturer. For example, if the decision maker (or a group its commissions) takes its critical review of the manufacturer's analysis to the point of altering parameters or even re-structuring the model, the distinction is blurred between who is doing the assessment and who is reviewing it. The third limitation of these arrangements is that, if the burden of proof falls with the manufacturer and this is assured through a rigorous review of their submission, genuinely

cost-effective technologies (which have been poorly analysed by their manufacturer) are rejected thus imposing costs on patients and the health care system.

These problems suggest that independent technology assessment has some clear advantages for decision making. Specifically, the independence of the group should ensure an impartial analysis which should provide more confidence in the analytical basis for decision making. Of course, a given independent analysis may be wrong (and the manufacturer's analysis may be more appropriate). For this reason, the decision maker should never abrogate their responsibility to think critically about all analyses it is presented with. However, over time, it would be expected that independent assessment would provide more reliable analyses for the reasons outlined in Section 2.

The comparison between the SMC and NICE in Section 4 suggests that independent assessment can make a difference to decisions, in particular due to a fuller consideration of heterogeneity. The are some limitations to this analysis which should be acknowledged. First, few comparisons were found (25, including 22 drugs and 18 indications), so this sample may be too small to drive any significant conclusion about the comparability of the two approaches. Second, the timing of NICE and SMC guidance is often different, and a gap in published appraisal of more than 2 years was found for some products (see Appendix 1 for more details). Thus, new evidence might have been available during this gap of time and could partly explain some differences in the recommendations. Third, while NICE publishes its full guidance on the website, the SMC only provides a short version and some details could not be available to the public. Nevertheless, a trend for the NICE guidance to be more restrictive and to give more information on patient sub-groups is evident.

So why do so few systems use independent assessments? Furthermore, why has one of the few examples of independent assessment been partly removed (through the initiation of STA at NICE)? In the case of NICE's changes, the need for more rapid decision-making was probably the main impetus for introducing the STA process. Reasons for doubting whether removing independent assessment will speed up the process were considered in Section 2, and time will tell whether NICE is able to make decisions more rapidly given that many aspects its standard (MTA) process are still in place. More generally, it might be felt that

relying solely on manufacturer analyses as the basis of decision making is more conducive to a process of negotiation and 'give and take' between the manufacturer and the health care system, in particular in those systems where negotiation over price is permitted.

Perhaps a more convincing case against independent assessment is its resource cost – not just the financial cost, but the limited numbers of researchers with appropriate training and experience to undertake such work. This factor is likely to be particularly compelling for health care systems in smaller countries, and in systems which consider all newly licensed pharmaceuticals. Furthermore, resource limits for independent assessment contrasts with the financial strengths of pharmaceutical companies who, at least in part, often undertake their analyses at a global level rather than separately for individual markets. It could be argued, however, that the limited capacity for independent assessment could be countered by using it more efficiently. In particular, different health care systems could fund and commission independent assessment collaboratively. It should also be noted that not all technologies are manufactured by affluent global corporations, and many manufacturers will not have the wealth, staff or wherewithal to undertake analyses satisfying the requirement of decision-making. In which case, independent assessment may provide the only means of undertaking appropriate analysis.

6. Conclusions

International health care systems have adopted different processes regarding technology assessment and its incorporation into decision-making, but most of these are based on the technology assessment being undertaken by the manufacturer. This position is unlikely to change in the short-term. There are ways in which the risk of manufacturers providing misleading analyses can be reduced including the use of thorough critical reviews against a clear, detailed and prescriptive sets of methods guidelines, and conferring the burden of proof at decision making on the manufacturer. Some systems are yet to initiate technology assessment and appraisal systems, and they should carefully consider the advantages and limitations of independent assessment. In the case of NICE, it seems as if the MTA and STA processes (with and without independent assessment, respectively) will co-exist for the time being. An important consideration is, therefore, which technologies are channelled into

which process. The nature of the comparators, the availability of the full evidence base and the likely presence of heterogeneity are factors which should influence this decision.

References

Formatted: Bullets and Numbering

1. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines - similarities, differences and some implications. *Value in Health* 2001;4:225-250.
2. Tarn TY, Smith MD. Pharmacoeconomic guidelines around the world. *ISPOR Connections* 2004;10(4):5.
3. Commonwealth Department of Health Housing and Community Services. Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: AGPS, 1992.
4. Ministry of Health. Ontario Guidelines for Economic Analysis of Pharmaceutical Products. Ontario: Ministry of Health, 1994.
5. National Institute for Health and Clinical Excellence (NICE). *Guide to the Single Technology Appraisal Process*. London: NICE, 2006.
6. Cairns J. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy* 2006;76:143-143.
7. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet* 2002;360:711-715.
8. Sculpher M, Claxton K, Akehurst R. It's just evaluation for decision making: recent developments in, and challenges for, cost-effectiveness research. In: Smith PC, Ginnelly L, Sculpher M, editors. *Health Policy and Economics. Opportunities and Challenges*. Maidenhead: Open University Press, 2005.
9. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *British Medical Journal* 1996;312:71-72.
10. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of heath care technologies. *Journal of Health Economics* 1999;18:341-364.
11. Griffin S, Claxton K, Palmer S, Sculpher M. *Dangerous Omissions: The Consequences of Ignoring Decision Uncertainty. Paper presented at HESG, summer, York*, 2006.
12. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness anaylsis: a framework for establishing efficient limited use criteria. *Health Economics* 2003;12:421-427.
13. Bojke L, Claxton.K., Palmer S, Sculpher M. *Defining and characterising structural uncertainty in decision analytic models. Centre for Health Economics (CHE) Research Report 9*. York: CHE, University of York, 2006.
14. Raftery J. Review of NICE' s recommendations 1999-2005. *British Medical Journal* 2006;332:1266-1268.
15. Stevenson M, Lloyd Jones M, De Nigris M, Brewer N, Davis S, Oakley J. A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis. *Health Technology Assessment* 2005;9(22).
16. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998;18 (suppl):S68-S80.

17. Miners AH, Garau M, Fidan D, Fischer AJ. Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. *British Medical Journal* 2005;330:65.

Table 1: Comparison between SMC and NICE guidance

Drug	Indication	Comparison
Alendronate	Osteoporosis	Both recommend for specific patient groups, but with different details ***NICE*** *gives more details on patients subgroups that might benefit from the treatment.* ***NICE*** *states that the choice between alendronate and other bisphosphonates should be based on tolerability*
Anakinra	RA	Both do not recommend
Capecitabine	Breast Cancer	Both recommend as 2nd line treatment
Capecitabine	Colon Cancer	Both recommend as adjuvant therapy
Clopidogrel	ACS	Both recommend as combination therapy (with aspirin), but with different restrictions ***SMC*** *appears more restrictive (given initially only to inpatient)*
Docetaxel	Breast Cancer	Both recommend as adjuvant therapy
Drotrecogin	Sepsis	Both recommend for severe patients
Etanercept	Psoriatic Arthritis	Both recommend for patients who failed other treatments, but with different details ***NICE*** *gives much more details on patients who could benefit from use*
Imatinib	GIST	Both recommend but with different restrictions ***NICE*** *appears more restrictive (recommends continuation only if response at 12 weeks achieved)*
Imatinib	CML	Both recommend for subgroup of patients, but with different restrictions and details ***NICE*** *gives much more details on subgroup of patients who could benefit from treatment* ***SMC*** *appears more restrictive and states that more clinical evidence is needed*
Insulin Glargine	Diabetes	Both recommend for patients with Type I diabetes
Levetiracetam	Epilepsy Adults	Both recommend, but with some differences: ***NICE****: as monotherapy for those who had failed older antiepileptic drugs and in combination or adjuvant therapy only when monotherapy fails.* ***SMC****: as adjunctive therapy in the treatment of partial seizures with or without secondary generalization*
Methylphenidate	ADHD	Both recommend, but with different restrictions ***SMC*** *appears more restrictive: use as 2nd line under specific circumstances*
Mycophenolate	Renal Transplant	Both recommend, but with difference restrictions ***NICE*** *appears more restrictive (confined to subgroups of patients and if intolerant to other treatments)*
Olanzapine	Bipolar Disorder	Both recommend, but for different phases of bipolar disorder and different restrictions: ***NICE****: treatment in acute manic phase (more restrictive)* ***SMC****: treatment both in acute manic phase and in stable phase to prevent recurrence of manic episodes*
Olanzapine	Schizophrenia	Both recommend as first line, but with different restrictions ***NICE*** *states that the choice between olanzapine and other treatments should be based on side effects (more restrictive)*
Oxaliplatin	Colon Cancer	Both recommend as combination therapy
Peg Interferon (alfa-2a and alfa-2b)	Hepatitis C	Both recommend, but with different details and restrictions. ***NICE*** *gives more details on treatment duration based on Hepatitis C virus genotype.* ***NICE*** *appears more restrictive recommending combination therapy for moderate or severe Hepatitis C and monotherpy only in the case of intolerance to ribavirin or for mild hepatitis.* ***SMC*** *recommends both as monotherapy and in combination with ribavirin*
Pimecrolimus	Atopic Dermatitis	NICE recommends (only as 2nd line, not 1st line), SMC do not recommend
Risedronate	Osteoporosis	Both recommend, but with different restrictions ***NICE****: focuses more on secondary prevention and states that choice between risedronate and other bisphosphonates should be based on tolerability (more restrictive)* ***SMC****: recommended for treatment and prevention*
Risperidone	Schizophrenia	Both recommend, but with different details ***NICE*** *gives much more details on use*
Rituximab	NHL	Both recommend as 1st line therapy in combination
Tacrolimus	Atopic Dermatitis	Both recommend as 2nd line and not as 1st line
Teriparitide	Osteoporosis	Both recommend, but with different details and restrictions ***NICE*** *only for secondary prevention, confined to subgroup of patients (more restrictive)*

		NICE *gives more details on patients subgroups that might benefit from the treatment.*
Topirimate	Epilepsy Adults and children	Both recommend, but with different details ***NICE*** *gives more details on formulation (monotherapy initially, combination only if monotherapy fails)*

Table 2. Recommendations of NICE and SMC guidance on the basis of Raftery's[14] categories.

Treatment	*Indication*	*Recommendation*									
		No additional restriction over SPc	***Use as 2nd Line***	***Use only if intolerant to other treatment***	***Continue only if response***	***Confine to specific sub-group***	***Use least costly option,***	***Monitoring Required,***	***Use by Specialist Only,***		
Alendronate	Osteoporosis			N		NS		N			
Anakinra	Rheumatoid Arthritis									N	
Capecitabine	Breast Cancer		NS					N	NS		
Capecitabine	Adjuvant treatment Following Surgery for Colon Cancer	NS						N	NS		
Clopidogrel	Acute Coronary Syndrome	NS							S		
Docetaxel	Early Breast Cancer	NS						S			
Drotrecogin	Sepsis	NS				NS			NS		
Etanercept	Psoriatic Arthritis	S	N						N		
Imatinib	GIST				N				NS		
Imatinib	CML	N				NS		S	S		
Insulin Glargine	Diabetes					NS					
Levetiracetam	Epilepsy Adults	S	N		N						
Methylphenidate	ADHD		S				N		NS		
Mycophenolate	Renal Transplant	S		N		N			S		
Olanzapine	Bipolar Disorder	S						N			
Olanzapine	Schizophrenia	S		N			N				
Oxaliplatin	Colon Cancer	NS							S		

Treatment	*Indication*	*Recommendation*									
		No additional restriction over SPc	*Use as 2nd Line*	*Use only if intolerant to other treatment*	*Continue only if response*	*Confine to specific sub-group*	*Use least costly option,*	*Monitoring Required,*	*Use by Specialist Only,*	.	
Peg Interferon (alfa-2a and alfa-2b)	Hepatitis C		N			NS					
Pimecrolimus	Atopic Dermatitis		N							S	
Risedronate	Osteoporosis	S				N		N			
Risperidone	Schizophrenia			N		S	N		S		
Rituximab	NHL	NS							NS		
Tacrolimus	Atopic Dermatitis		NS						NS		
Teriparitide	Osteoporosis					NS					
Topirimate	Epilepsy Adults and children		NS								

Note: N=NICE, S=SMC

Table 3. Incremental costs, QALYs and net benefits for osteoporosis drugs stratified by age (women with p and T-score –2.5)

Drug	Age	Incremental costs (£)*	Incremental QALYs*	Population	Total net benefit (λ=£30,000)*
Alendronate					
	50	149	0.0044	15500	-248,372
	60	139	0.0035	46500	-1,583,418
	70	95	0.0056	248000	18,183,608
	80	5	0.0065		0
All Patients		**104**	**0.0052**	**310000**	**16,351,818**
Risedronate					
	50	145	0.0034	15500	-654,581
	60	135	0.0029	46500	-2,236,604
	70	104	0.0047	248000	9,350,592
	80	26	0.0053		0
		110	**0.0044**	**310000**	**6,459,408**
Teriparitide					
	50	558	0.0025	15500	-7,505,782
	60	553	0.0021	46500	-22,857,215
	70	530	0.0039	248000	-102,218,160
	80	487	0.0040		0
		535	**0.0036**	**310000**	**-132,581,157**

* compared with no treatment in women with sufficient calcium and vitamin D intakes

Table 4. Incremental costs, QALYs and net benefits for osteoporosis drugs stratified by age (fracture risk d alendronate and risendronate, quadrupled for teriparitide)

Drug	Age	Incremental costs (£)*	Incremental QALYs*	Population	Total net benefit (λ=£30,000)*
Alendronate					
	50	128	0.0089	15500	2,131,018
	60	111	0.0070	46500	4,572,624
	70	33	0.0112	248000	75,209,472
	80	-131	0.0130		0
All Patients		**50**	**0.0105**	**310000**	**81,913,114**
Risedronate					
	50	128	0.0069	15500	1,206,861
	60	112	0.0058	46500	2,908,668
	70	92	0.0094	248000	47,293,848
	80	-80	0.0106		0
		97	**0.0088**	**310000**	**51,409,377**
Teriparitide					
	50	526	0.0098	15500	-3,593,040
	60	511	0.0083	46500	-12,239,684
	70	431	0.0158	248000	10,215,616
	80	268	0.0158		0
		448	**0.0143**	**310000**	**-5,617,107**

Formatted: Bullets and Numbering