For HESG/Birmingham, Jan 2007; work very much in progress; not for citation

Incorporating equity weights into cost effectiveness analyses – opening Pandora's box?

Aki Tsuchiya[1], Christopher McCabe[2], Allan Wailoo[1]

1. University of Sheffield
2. University of Leeds

Abstract

Historically, health care decision makers only considered issues of safety and efficacy when deciding whether a therapy should be available or not. Over the last two decades, the increasing cost of health care has led decision makers to consider issues relating to the efficiency of alternative uses of their limited health care budgets. Current practice is to assign equal social value to a unit of health improvement ("a QALY is a QALY is a QALY"). However, there is now substantial evidence suggesting that members of the public do not hold this view. Against this background it has become increasingly clear that decision makers need to consider issues of equity alongside that of efficiency. Amendments to the QALY model have been proposed which would allow the incorporation of explicit differential societal preferences for health gain to different groups within society – so called "equity weights". To date, research on equity weights has focussed on candidate criteria for equity weights and methods for estimating such weights. It has implicitly assumed that should legitimate, valid, and reliable equity weights become available, they would be incorporated through a simple process of re-weighting the denominator in the ICER. In the first part of this paper we explain why such an approach is likely to be appropriate in only a handful of cases and in many more cases it would lead to a misrepresentation of society's preferences for health and equity. The second part of the paper examines the hidden equity weights in current practice, and highlights some implications adopting an explicit 'disaggregated equity' strategy in cost effectiveness analyses.

Incorporating equity weights into cost effectiveness analyses – opening Pandora's box?

Background

Historically, health care decision makers only explicitly considered issues of safety and efficacy when deciding whether a therapy should be available or not. Over the last two decades, the increasing cost of health care has led decision makers to consider issues relating to the efficiency of alternative uses of the limited health care budgets. In response to this there has been a massive expansion in the utilisation of cost effectiveness analysis within health care decision making processes, accompanied by an explosion on the pace of development in the methods of cost effectiveness analysis.

One of the areas of greatest development has been in the measurement and valuation of the outcomes of health care, most notably the development of research around the Quality Adjusted Life Year (QALY) measure of health gain. The last 20 years or so has seen extensive research on its theoretical foundations (for example Pliskin et al., 1980, Garber and Phelps, 1997, Miyamoto et al., 1998, Miyamoto, 1999), and an increasing utilisation of this measure in cost effectiveness analyses used by decision makers. Current practice is to assign equal social value to a unit of health improvement no matter who the recipient, how the benefit is generated, or the reason an intervention is required in the first place ("a QALY is a QALY is a QALY"). However, there is now a large body of literature which suggests that members of the public do not hold this view (see Schwappach, 2002, Dolan et al., 2005 for reviews). Against this background, it has become increasingly clear that decision makers need to take account of equity issues. Indeed, decision makers that make explicit use of cost effectiveness analysis frequently make an explicit commitment to considering the equity implications of their choices, but lack a formal framework in which to consider it.

Williams and Cookson (2000) have proposed a four stage model of the development of decision making (see Box 1), ranging from the most intuitive stage in the absence of robust evidence to the most formalised stage where all relevant parameters are quantified and known, and it is merely a matter of combining these quantities to reach the decision. It would be fair to say that almost all current decision making processes are at best in Stage 2, which builds on case law and precedents. The National Institute for Health and Clinical Excellence (NICE) has recently published a set of principles regarding Social Value Judgements (NICE, 2005) which move it towards Stage 3, governed by explicitly stated but not quantified principles. The methods guide of the NICE Appraisal programme (NICE, 2004) suggests that NICE is committed to moving on to the final Stage 4 as and when knowledge of society's preferences allow. In contrast, many local decision making processes rely upon intuition and case law.

Researchers have proposed amendments to the QALY model which would allow the incorporation of explicit differential societal preferences for health gain to different groups within society – so called equity weights (Wagstaff, 1994, Williams, 1997, Nord et al., 1999, Dolan and Tsuchiya, 2006), or "super QALYs". The adoption of these equity weights and super QALYs would move a decision making process such as the NICE Appraisal Programme from Stage 2 to Stage 4 in the Williams and Cookson model above.

To date, research on equity weights has focussed on candidate criteria for equity weights and methods for estimating such weights. It has implicitly assumed that should legitimate, valid and reliable equity weights become available, they would simply be incorporated in to cost effectiveness analysis (CEA). Two simplistic approaches are possible. In the first approach, the cost effectiveness threshold will be varied according to the equity characteristics of the patients affected. In the second approach, the calculation of incremental cost effectiveness ratios will incorporate equity weights in a three stage process. In the first stage, the current methods of CEA would be used to construct estimates of the costs and utilities for alternative interventions. The second stage would apply the equity weights to the utility estimates. The third stage would estimate incremental cost effectiveness ratios using the costs from stage 1 and the adjusted utilities from stage 2.

However, these approaches are likely to be appropriate in only a handful of cases and, if adopted, would often lead to a misrepresentation of society's preferences over health and equity that the QALY and associated equity weights are intended to reflect. In addition to imposing an additional level of technical complexity to the production of CEAs, the incorporation of equity weights will require the reappraisal of a number of other issues.

In the first part of this paper we explain why such an approach is likely to be appropriate in only a handful of cases. The second part of the paper examines the hidden equity weights in current practice, and highlights some implications adopting an explicit 'disaggregated equity' strategy in cost effectiveness analyses.

While the paper concentrates on the discussion of equity-based weights, a large part (but not necessarily all) of the discussion also applies to efficiency-based weights such as giving larger weight to parents relative to non-parents. In particular, the disaggregating equity-based age weights and efficiency-based age weights will be an interesting task, but is not addressed in this paper.

Trying to incorporate Equity Weights into CEA

Equity weights may be attached to more than one characteristic and there is no reason to believe that patient groups will be homogenous with respect to these characteristics. Thus, within a patient group, some patients‘health gains will be valued more highly than others' for equity reasons, even when the size of the health gain is the same. Further, an individual will have certain characteristics that may attract an equity weight at some points in their life and not at other times; e.g. smoking status. Decision models are generally concerned with long term costs and effects and therefore patients, and groups of patients, will attract different weights at different points in the model.

The simple method of incorporating equity weights by adjusting the utility estimates in the ICER will only be valid where three conditions hold. Firstly, patients for the treatment in question are homogenous with regard to equity characteristics. For example, treatment for conditions that typically hit infants will have a more homogeneous patient base than treatment for conditions that can affect anybody. Secondly, equity characteristics of patients remain constant over the time horizon of the analysis. For example, ceteris paribus equity characteristics of patients with acute conditions are more likely to remain constant than those with chronic conditions. Thirdly, equity characteristics are defined broadly relative to the time horizon of the analysis so that patients do not change categories. For example, having just two age categories ("under 65" and "65+") is more promising than having a larger number of categories.

The extent to which these conditions will hold in practice depends, in part, on how sensitive our equity weights are. For example, if different equity weights are applied according to age, but these weights are for very broad bands, then one could imagine there would be situations where the same weight would apply throughout e.g. flu treatment for otherwise healthy adults. If the equity weights categories are not sufficiently broad, then to some extent, patient subgroup analysis could still address this; e.g. in the flu example if there were different weights for men and women as well as age then one would get a different ICER for men aged 20 than for women aged 50. This is for those situations where the patient characteristics themselves do not change throughout the time horizon of the model. It is worth noting at this point that there may be a tension between analyses for clinically defined sub-groups and equity sub-groups. We consider the potential for unexpected consequences of equity weights in more detail in section two.

Where patient equity characteristics change over the time course of the analysis, because the weight is not constant and/or because the categories are not broad enough, then the simple approach of adjusting the end ICER outlined above is not appropriate. Under such circumstances the simple approach would produce results that can be over- or under-estimates depending on the nature of the (combination of) equity characteristics. Ceteris paribus, the error would be more important the smaller the difference in effectiveness between the comparators.

The appropriate method for incorporating equity weights will depend upon a number of factors, including:

- degree of patient heterogeneity with regard to the equity characteristic;
- whether there is one or multiple equity weights relevant to the patient group;
- whether progression through the model is dependent upon the equity characteristic; and
- whether the equity weight for a specific characteristic is constant throughout the time horizon of the analysis.

Where there is a single equity characteristic of relevance to the decision and patients are homogenous with regard to that characteristic over the duration of the analysis, it should be possible to create a payoff matrix that is a function of the equity characteristic. For example in a lifetime cohort Markov model, where we wish to weight QALYs according to the patient age then it would be simple to incorporate the equity weights by making the reward (QALY) in each stage a function of time.

Where patients are not homogenous (even by subgroup) it may be possible to still adjust the rewards at each stage of a cohort model but applying a weighted average of the equity weight in each stage. For example, if we are only interested in age weights then we can either assume a mean aged patient, or estimate the proportion of patients in each state at each stage. However, it may be impossible to do this when the characteristics that determine the scale of the equity weight also will determine the path through the model.

Cohort models will not be suitable frameworks for direct estimation of equity weighted ICERs, when the incorporation of equity weights will generate a large number of distinct states. Analyses which will generate a large number of distinct states include those where:

- there are multiple equity characteristics;
- patients are heterogeneous with regard to the equity characteristics;
- the path through the model is dependent on the equity characteristics; and
- the equity weight itself is a function of the length of time spent in a health state (e.g. weights for severity, or the length of time with the condition).

Under these circumstances states can no longer be accurately defined in terms of the actual health state with an external payoff. Rather, states are defined as the interaction between the health characteristics, the characteristics of the individuals with those health characteristics and society's equity preferences for a person with those characteristics in that health state. It is clear that even for quite simple models, the number of states may easily become unmanageable. Thus, the incorporation of weights for QALYs may necessitate a greater use of individual sampling models. The computational issues associated with the use of individual sampling models for incremental cost effectiveness analysis have been rehearsed elsewhere. It is sufficient to note that substantial progress, along the lines indicated by O'Hagan and colleagues, will be required if incorporating equity weights into cost effectiveness analysis is not going to significantly impact upon assessment of decision uncertainty that is increasingly seen as standard.

Modelling Equity Weights for CEA

In the same way that health is multi-dimensional, people have equity preferences over more than one domain – e.g. age, gender, severity of initial health, and accountability for ill health. It is arbitrary to consider only one domain of equity. If equity weights are to be routinely incorporated into cost effectiveness analysis then we should consider all equity domains for which there are a prima facie case to assume that they are relevant and legitimate. An individual's equity weight will therefore be a function of their status in each equity characteristic, and we would need to understand the form of the equity weight function in order to incorporate it into CEA. Do equity weights combine in a simple additive manner, or is the function multiplicative? The present evidence base on the relative social value of health does not provide us with strong evidence on the important equity domains and there has been no attempt to test alternative functional forms for a multi-attribute equity preference function. Until such research has been done, incorporating equity weights into cost effectiveness analyses will be as likely to produce biased as accurate ICERs and the results will be no more defensible than the current unweighted, or rather implicitly uniformly weighted, ICERs.

Disaggregate equity weights

Current best practice in cost effectiveness analysis is not free of equity considerations - or in other words, it does not reflect efficiency alone. It is merely that the equity weights are implicit and hidden away. For example, cost effectiveness analyses will normally only provide gender specific estimates of cost effectiveness when there is evidence of a difference in the natural history of a condition, or the effectiveness of the interventions between genders. However, the life expectancy for females is systematically higher than the life expectancy for males, thus combined analyses of treatments with life-long effects overstate the cost effectiveness of treatment for males and understate its cost effectiveness in females. Similarly, analyses are not produced for different socio-economic groups, even though the difference in life expectancy varies substantially between the wealthiest and poorest socioeconomic groups. The result is that the health gains of a therapy in poorer socio economic groups are systematically overstated. At the margin these weights might lead to therapies being provided to populations in whom they are not cost effective.

This therefore leads to the importance of 'disaggregate' analyses. Implied equity weights should be removed by adopting subgroup specific measures of health, and only then, explicit equity weights should be introduced. Simply adding explicit equity weights to current practice will lead to double counting. For example, if we had an equity weight for health gain to people in the lowest socio-economic group, it would be important that the life expectancy used in the analysis was socio-economic group specific. Equally, if there was evidence that compliance varied by socio-economic group, it would be necessary to model this variation explicitly. Otherwise, the total equity weight (explicit plus implicit) for health gain to different socio-economic groups would not reflect the measured equity preferences. This principle would apply to any equity weighted characteristic where there was evidence that other parameters in the analysis varied systematically by that characteristic.

Imagine the case where larger equity weights are assigned to those whose expected lifetime QALYs (ELQ) are low. Suppose further that this subpopulation's lower ELQ is reflected in smaller net benefit from treatment for this subpopulation. The exercise to determine the level of the equity weight will be based on the actual magnitude of the health inequality between this population subgroup and the rest of the population (or some other reference population). In other words, the obtained weights are relevant to the actual degree by which this subpopulation has lower ELQ. Therefore, as noted above, if weights are to be incorporated into economic evaluation, they should be applied to subgroup-specific measures of net benefit. It is inappropriate to simply apply the weights to measures of benefit derived for the overall population.

Let us assume that the baseline "current practice" is to conduct economic evaluation based on whole population variables, with uniform weights. Let us further assume that the subgroup specific measure of net benefit is 50% of the net benefit to the population overall (i.e. 1.0 QALY gain for the overall population, and 0.5 QALYs gain for the low ELQ population); and that the equity weight given to this low ELQ population subgroup relative to the overall population is 1.5. If equity weights are applied to population subgroup specific net benefits, then the weighted subgroup net benefit will be 0.75 QALYs (the product of the benefit and the weight), whereas the net benefit for the overall population will be 1 QALY. Then, costs being the same, unless the subgroup specific benefit is larger than 50% of the overall population ***or*** the weight given to them is larger than 2, those with lower ELQ will be disadvantaged by the introduction of equity weights, and will be better off under current practice with no equity weights.

One may think that such an outcome is unacceptable, and that the introduction of equity weights should always advantage those with lower ELQ (or any other vulnerable groups). The question however is advantage relative to what. The above example clearly disadvantages those with lower ELQ compared to current practice. But this is because current practice already inflates the benefits of those with low ELQ in a way that is not necessarily formally accounted for or justified.

Does the above imply that current practice can substitute for the explicit incorporation of equity weights into economic evaluation? Given the complexity of incorporating multiple weights in CEA, as explored in the first half of this paper, this might be an attractive approach. If the only reason for justifying weights is based on variation in the size of net benefits, then current practice already implicitly incorporates weights in the right direction, because those with smaller net benefits receive the inflation, or implicit positive weighting. Depending on the combination of net benefits and the size of equity weights, this may overstate more precise subgroup weights (as in the example above), but it may also understate. Therefore, the additional benefit of precision associated with subgroup level (or even individual level) analyses in the context of an economic evaluation as a whole needs to be weighed against the additional costs of carrying out such analyses. Given the level of uncertainty associated with all the other variables in the economic evaluation process, it may be concluded that the additional benefit of precision in equity weighted net benefits is not worthwhile, or in other words, that current practice is reasonably robust, if the only reason for justifying weights is based on variation in the size of net benefits.

Unfortunately, this logic does not apply to all cases of equity weights. If the main reason for equity weights is variation in ELQ, then this may or may not be directly proportional to the size of net benefits. If smaller net benefits are associated with higher ELQ (because for instance those with higher ELQ due to better baseline health have less scope to improve their health further), then the direction of implicit weights embedded in current practice goes in the "wrong" direction, because it inflates the benefits of those who have better ELQ. Moreover, if equity weights are not related to measures of health benefit but based on the extent to which individuals are responsible for their own health problem, then the implicit weights built in current practice cannot be justified.

Summary

In this paper we have attempted to describe why the estimation of equity weighted ICERs will often be more complicated than practitioners, including ourselves, have presumed. Over the past decade CEA models have developed from primarily simple decision trees, subjected to one or two-way sensitivity analysis to state transition models with probabilistic sensitivity analysis. The incorporation of equity weights will often require the utilisation of more complicated, patient level models, with an associated increase in the complexity of the methods for analysing uncertainty. The state of knowledge may not yet be sufficient to support the routine provision of probabilistic sensitivity analysis for models that include equity weights.

Such developments assume that the evidence base for society's preferences over health and equity is sufficiently developed to allow us to specify the important domains of equity (in health) and how these are valued relative to each other. This is not currently the case. As a result, any equity weighted cost effectiveness analysis will be applying an essentially arbitrary and indefensible factor to the health gain for a sub-group of the population, producing results that would not be a suitable foundation for resource allocation decisions.

Should the state of knowledge of society's preferences over equity, and the technical methods of cost effectiveness analysis, develop to the point where it is feasible to produce evidence based equity weighted ICERs, with an appropriate assessment of decision uncertainty, there may still be reason to pause before proceeding. If we wish to incorporate equity weights for domains that are correlated with other parameters in the cost effectiveness model, it will be necessary to estimate disaggregate sub-groups specific parameter values in order to provide unbiased estimates of the equity weighted ICER. This may have some unexpected implications, including making an intervention less cost effective in sub-groups whose health gain is valued most highly.

This is not to argue that we should avoid taking explicit account of society's preferences over equity. We broadly agree with Cookson and Williams arguments for transparency. However, given the challenges outlined above, it seems that there is an entirely underdeveloped are of research to be explored before CEAs can incorporate equity weights as a routine matter.

References

Dolan, P., Shaw, R., Tsuchiya, A. and Williams, A. (2005) QALY maximisation and people's preferences: A methodological review of the literature ***Health Economics,*** 14, 197-208.

Dolan, P. and Tsuchiya, A. (2006). The elicitation of distributional judgements in the context of economic evaluation, In Jones, A. ed. ***Companion to Health Economics,*** Elgar.

Garber, A. M. and Phelps, C. E. (1997) Economic foundations of cost-effectiveness analysis ***Journal of Health Economics,*** 16, 1-31.

Miyamoto, J. (1999) Quality-adjusted life years (QALY) utility models under expected utility and rank dependent utility assumptions ***Journal of Mathematical Psychology,*** 43, 201-237.

Miyamoto, J. M., Wakker, P. P., Bleichrodt, H. and Peters, H. J. M. (1998) The zero-condition: A simplifying assumption in QALY measurement and multiattribute utility ***Management Science,*** 44, 839-849.

NICE (2004). ***Guide to the Methods of Technology Appraisal*** (reference N0515)

NICE (2005). ***Social Value Judgements: Principles for the development of NICE guidance*** http://www.nice.org.uk/page.aspx?o=283494 (accessed 22 Aug 06).

Nord, E., Pinto, J. L., Richardson, J., Menzel, P. and Ubel, P. (1999) Incorporating societal concerns for fairness in numerical valuations of health programmes ***Health Economics***, 25-39.

Pliskin, J. S., Shepard, D. S. and Weinstein, M. C. (1980) Utility functions for life years and health status ***Operations Research,*** 28, 206-224.

Schwappach, D. (2002) Resrouce allocatoin, social values and the QALY: A review of the debate and empirical evidence ***Health Expectations,*** 5, 210-222.

Wagstaff, A. (1994) QALYs and the equity-efficiency tradeoff. In Layard, A. and Glaister, S. eds. ***Cost-Benefit Analysis***, Cambridge University Press (reprinted from ***Journal of Health Economics***, 10:21-41, 1991).

Williams, A. (1997) Intergenerational equity: An exploration of the 'fair innings' argument ***Health Economics,*** 6, 117-132.

Williams, A. and Cookson, R. (2000) Equity in health. In Culyer, A. J. and Newhouse, J. P. eds. ***Handbook of Health Economics***, Vol. 1B, Elsevier Science.

Box 1: Framework for development of resource allocation processes.

Stage 1	Decisions are made by the designated decision maker based upon intuition.
Stage 2	Decisions are made with reference to case law. Case law is limited to principles derived from precedents set by previous decisions.
Stage 3	Principles are codified in order to allow their generalisation; but without specifying relative weights to be used when these principles conflict with one another.
Stage 4	Relative weights are specified and quantitative trade-offs are incorporated into a formula. Future argument then centres on the terms of the formula.

Taken from Williams and Cookson (2000)