CENTRE FOR HEALTH
PROGRAM EVALUATION

WORKING PAPER 108

The Conceptual Basis for SMPH Qualifying DALYs; Dallying with QALYs: How are we to Evaluate Summary Measures in Population Health?

Professor Jeff Richardson
Director, Health Economics Unit, Centre for Health Program Evaluation
Monash University

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434 **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au

ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR ACKNOWLEDGMENTS

This paper was delivered at the WHO Global Conference on Summary Measures of Population Health, Marrakech, Morocco, December 1999.

Abstract

This paper is concerned with various issues surrounding the validity of summary measures of population health (SMPH). As it argues that economic orthodoxy cannot answer this question the paper considers the prior issue of the criteria with which to evaluate SMPHs. It is argued that as there are multiple objectives or uses for SMPHs then it is likely that more than one such measure will be needed and specifically that the SMPH most appropriate for measuring the burden of disease will not necessarily be the same as the SMPH used for economic evaluation. With respect to the criteria for judging SMPHs, it is argued that the use of the Bale of Ignorance, favoured by Murray et al, 1999, is unsatisfactory. Alternative criteria are proposed and discussed. One criteria is that SMPHs should embody ethical values consistent with stable population values. This raises the prior question of the relationship between observed population values and social objectives. It is suggested that the investigation and analysis of this relationship should be described as 'empirical ethics'. Four examples of this in the context of SMPHs are discussed. First, it is shown that population values may be context specific.

Secondly, it is shown that the measures of utility employed to derive SMPHs are all 'contaminated'; that is, they do not measure, empirically, what it is believed that they measure. Thirdly, evidence is presented that the preference number for utility attached to different health states will systematically differ when it is elicited from patients and the community. Fourth, there is an illustration of the way in which empirically derived utility values may be used to derive implications which may encourage the revision of stated values after deliberation (the 'strong interval property'). It is concluded that there is a need to recognise the importance of 'empirical e the techniques and rules which will govern its use and a need to apply empirical ethics to a potentially large range of social problems.

Table of Contents

List of Tables

The Conceptual Basis for SMPH Qualifying DALYs; Dallying with QALYs

How are we to Evaluate Summary Measures in Population Health?

1 Introduction

This paper is concerned with the social valuation of the life years that are the basis of the current summary measures of population health (SMPH). It focuses particularly upon the estimation of the weights which should be used to convert life years into units of 'social value' where this latter term is defined to mean an index of the fulfillment of health related social objectives.

The paper eventually comments on the following issues:

- who should judge social values: patients or the public;
- which of our scaling instruments is to be preferred; the standard gamble (SG), time trade-off (TTO) or the person trade-off (PTO) and what do they measure;
- measurement from behind the veil of ignorance;
- what distinguishes social value from HRQoL/utility; and
- what are the priorities for instrument development and for instrument related research?

The answer to these questions 'discussed in Section 5' depends upon the *criteria* that are used to evaluate the various options and it is this prior question which is addressed first. Restated, the acceptability of the answers to these questions depends upon the conceptual basis of the evaluative framework and the criteria which are implicit or explicit in the framework.

Before addressing this prior question I discuss two common misconceptions; namely, the view that there is a separation between the 'conceptual basis' of normative economics and the ethical and social values of the population; and, secondly, that there is a single unit of health related utility or social value which is 'correct' in each of the various contexts in which an SMPH may be useful.

In summary the argument below is as follows: There is no purely objective basis for our choice between different SMPH's. Each should be evaluated according to explicit criteria. These in turn must reflect, first, the purpose of the analysis and, secondly, normative – ethical – values. If there is more than one purpose it is possible – probable – that more than one SMPH will be needed.(Section 2). We may investigate values empirically and draw out their consequences. However, the criteria and their implied values must ultimately be 'sold' to decision makers and the community (Sections 2 & 3). Criteria for one objective – the economic evaluation of health programs – are proposed and discussed (Section 4) and applied to the questions above (Section 5).

2 Ethics and the Conceptual Basis

The assertion that one state of the world is better than another is always and unavoidably based upon an ethical theory or belief. The distinction between the achievement of 'economic ethical' or 'equitable' intervention arises not from a fundamental difference in kind between economic 'efficiency' and 'equity' but from the breadth of the acceptance of the ethical – normative – basis of 'economic efficiency' and from linguistic convention. Technical efficiency per sé – the ability to do more with the same resources or to use fewer resources to do the same job – only results in an improved state of the world if 'more' is better or if the resources freed by technical efficiency produce more of something which is valued. The most commonly accepted definition of 'economic efficiency', 'Pareto efficiency' (the assertion that social wellbeing is increased if no one is worse off and someone is 'better off') requires a normative judgement about what constitutes being 'better off'. The practical application of the principle will almost always require the additional ethical judgement that the 'envy' of those who are not better off should not be taken into account and, therefore, that rising inequalities that are consistent with 'Pareto' efficiency should not, normally, be considered.

The conventional solution to this latter problems or to the problem of 'losers' the Kaldor Hicks criterion – asserts that one state of the world is better than another if losers may, potentially, be compensated. The widespread acceptance of this self evidently defective principle is perhaps a testament to the strength of economist's desire to avoid ethical debate. At best, the application of the principle encourages the identification of *potentially* better states of the world. At worst, it redefines the meaning of 'better' to mean what, in common language, is meant by 'potentially better' and in so doing permits policy advice to be based upon the covert and dubious ethical proposition that 'potentially better' *should* be equated with 'better'. A more benign interpretation of Kaldor-Hicks is that economists pass back to political decision makers the issue of whether or not to compensate; and the fact that compensation seldom if ever occurs is outside the control of economists. The common disregard of distributional issues in economic analyses casts some doubt upon this interpretation and this is particularly true in the analysis of health and health programs. When decisions affecting life and death are to be made, compensation is not possible, even in principle. As compensation for health and health services not received has never occurred nor ever been contemplated in any country, the Kaldor Hicks criterion is simply

irrelevant and the concept of 'pure economic efficiency' value free improvement – is misleading. While 'efficiency' may be a useful label to designate changes where there is widespread support for the underlying values, it may also serve to promote the wrong notion that economic advice may be based upon technical and not normative criteria.

In the analysis of health systems and health programs, the term 'efficiency' has come to be associated with 'increasing the quantity of health' as measured, for example, by life years or quality adjusted life years (QALYs). However, as increasing 'efficiency' may always offend particular notions of equity or engender envy, 'increasing health' is not value free. It is possible that there is, happily, a near consensus that a particular change or distribution is desirable. There may be near consensus that envy should be ignored. The point, however, is that the desirability of a change or the implied superiority of a state described by a larger SMPH is value laden and the 'conceptual basis' of these measures is the formalisation of ethical values and the exploration of their logical implications. In sum, there is no clear distinction between the conceptual and ethical bases of SMPH.

3 Social Value and Context

It is trite to argue that value is context specific. The value of water to a person dying of thirst differs from the value to an individual in normal circumstances. As noted by Allais in 1953, an incremental increase in the probability of a favourable outcome is valued differently in the context of near certainty and in the context of a low probability of a favourable outcome. It is also commonly acknowledged that criteria of value may be context specific. While clothing may be judged primarily by elegance in temperate climates, in colder climates the most important criterion may be warmth. Despite this, economists and philosophers commonly seek context free ethical rules: because willingness to pay is an acceptable criterion of value in the supermarket it must also apply in the health sector; if utilitarianism is unacceptable in a particular context (increasing global utility through the torture of one individual) then it is discredited generally.

There has been an increasing recognition that context is of ethical relevance in the allocation of health resources. Ubel, Richardson and Pinto (1999), for example, found that respondents discriminated between the treatment of long term quadriplegics and the treatment of previously healthy patients who would become quadriplegics after treatment. The pre-existing condition was of relevance and the problem of 'double jeopardy' (or double disadvantage see below) only appears paradoxical or inconsistent when context is ignored and when utility gain is equated with social value in every possible context (Nord 1999).

More generally, it has been argued by Nord (1999a) and Nord et al (1999) that utility and social value must be distinguished. While utility is an input into social value, social value may depend upon a variety of additional and, in particular, contextual factors. These may include age, severity of the initial health state, final health potential, the maintenance of hope, the achievement of certainty and the duration of the benefits obtained. (For a discussion of these see Menzel et al 1999). In principle, a society has considerable flexibility in determining what constitutes 'social

value' and, in principle, it may add to or subtract from the elements considered in orthodox economics. Thus, for example, Olsen and Richardson (1999) have introduced the notion of 'socially relevant' and 'socially irrelevant' production gains in the context of the debate over the indirect benefits of a health program and the equity 'problem' which occurs when indirect benefits systematically favour the wealthy. Such benefits, like utility gained from envy and sadism, may be socially irrelevant. Importantly, different societies may have quite different values and include or exclude different items or weight the importance of the same items differently.

The immediate relevance of this discussion is that there may be no single SMPH which is satisfactory in every context. The relationship between SMPH's and context is, however, complex and two separate relationships may be distinguished. The first arises because of the differences in values which are held in different cultures both within and between countries. Thus, for example, multi-attribute utility instruments for the measurement of QALYs or DALYs may require different weights or even different items in different cultures. This is illustrated, dramatically, by the difference in the relative importance of the effect of illness on physical wellbeing and upon social relationships between the European and Aboriginal populations of Australia, the latter group sometimes devaluing physical wellbeing almost totally when it interferes with social obligations.

Nevertheless, if each culture specific instrument is valid in its designated setting, then the adjusted life years predicted by the use of the instrument may have a common meaning; *viz* the number of years of full health which people in each culture consider to be equivalent to the life years actually experienced. For particular purposes, such as the distribution of international aid, the value judgement may be made that these units *should* be treated as equivalent. Similarly, they are a defensible metric for the measurement of the burden of disease as perceived by those who are affected by the disease.

Second, and more controversial, it is likely that the social value of a health program will not be accurately measured by the change in the value of the SMPH used for the calculation of some useful conceptualisation of the Burden of Disease (BoD). The social value of a health program may incorporate some or all of the factors discussed above (age, severity etc) and may depend upon other contextual factors. For example, and drawing upon the argument known as the 'rule of rescue', the use of resources in the context of an emergency may be considered more important than the use of similar resources to gain greater health benefit in a less dramatic context.

Generalising, the *concept* sought in a SMPH for the comparison of population health may differ from the *concept* sought for the evaluation of health programs. For the former, it is likely that the most useful concept will be a notion of suffering or a loss of benefits arising from ill-health. The concept may closely resemble the economist's (most commonly used) notion of utility. Its distinguishing characteristic is that it has a psychological property: it is in some sense 'felt' or 'experienced' by the patient. For the evaluation of health programs this concept may be mutated or complemented by factors associated with age, severity, timing (rate of time preference) etc;

factors that enter social value and not utility in the sense just described. These may include intellectual or ideologically held beliefs which are 'thought' and not, primarily 'felt' in the usual psychological sense[1]. (This distinction is discussed in greater detail in the companion paper Richardson (2000), 'Age weighting and discounting: What are the ethical issues?' Working Paper 109, Centre for Health Program Evaluation, Monash University.)

The present DALY used in the global burden of disease is an amalgam of these two concepts. It is a time based concept of wellbeing modified by age weights and a rate of time preference neither of which are time based or experienced (Murray and Lopez 1996). While the latter – time preference – is incorporated in the economic evaluation of health programs, this does not imply that it must be incorporated in a measure of the burden of disease.

Thus it is possible to conceptualise the BoD as a loss of 'utility' defined in time based, psychological terms – such that the loss is in some sense 'felt' through time and to separate this from the concept of the social value of this utility or wellbeing which may also include more 'cerebral' dimensions arising from an intellectual desire to achieve objectives other than time based utility: dimensions including time preference and notions of social justice.

As a general proposition, concepts are useful when they provide useful information and, in particular information which is necessary to help resolve a problem. Thus, the fact that two distinct SMPH's can, in principle, be distinguished does not, in itself, justify the adoption and quantification of these concepts. It is, therefore, implicit in the discussion here that the suggested concepts will help health planners. Without pursuing this theme, it is suggested that each of four concepts and the corresponding SMPH's will be helpful in an appreciation of the consequences of disease *viz* (1) lives lost; (2) life years lost; (3) suffering in psychological terms – loss of units of psychological wellbeing; and (4) the present value of this loss as judged – not by patients – but by present members of the society employing their present notions of social value.

The task of determining an appropriate SMPH for the documentation and comparison of population health is far less demanding than the task of deriving an SMPH for the evaluation of interventions. In the former case there is no need for a single SMPH. Multiple indices may be used and there is no particular problem with describing a population as having, on the one hand, poorer physical health and, on the other hand, better mental health and (on the remaining hand!) higher mortality. Combing these and other dimensions of health into a single index has both advantages and disadvantages. Importantly, however, if no specific policy is to be based upon the single index number, then the failure of the scaling device to precisely combine the

[1] Interestingly, the notion of social value discussed here corresponds, in many respects, with the notion of 'preference utility' as described by the utilitarian philosopher, Hare. According to his view 'utility' is increased when a person's preferences are fulfilled. The preferences, however, are more cerebral than the utility which is felt for extended periods of time. Preferences are more like an intellectual desire than a time based and on-going feeling. Consequently, utility is achieved – as in the case of social value – when the state of the world corresponds with a person's preferences even though the person may be unaware of this. Thus, a person'' preference for an equitable allocation of resources may be taken into account even though the person may never be aware of this. Hare does not appear to discuss the quantification of this notion.

components into a fully validated representation of the underlying concept is not particularly worrying. Broad patterns will still emerge from the summary measure. With no policy consequence, little harm will arise from imprecision.

In contrast, there is far greater pressure to produce a single, valid and coherent SMPH for evaluative purposes. Programs either will or will not be funded and this discrete choice requires the combination of all dimensions of health outcome and their weighting and adjusting in accordance with social values. Error in measurement or the invalidity of the scaling instrument (it does not produce what we believe it produces) will result in the misallocation of resources. In particular, and as noted below, if the scaling instrument used does not have a 'strong interval property' then the combination of quality and quantity of life the defining characteristic of a QALY/DALY – will be misleading and the summary measure will systematically favour either quality or quantity of life promoting programs.

4 Criteria for Assessing Social Value

Since economists regularly offer advice which purports to assist with the achievement of social objectives, it is surprising that there is so little critical discussion of the *criteria for determining social objectives*. At a high level of abstraction, economic theory acknowledges the existence of a 'social welfare function' (SWF) which, in principle, can incorporate any social arrangement which contributes to (undefined) social wellbeing. The term 'function' is presumably used to add gravitas to an otherwise vacuous acknowledgment that society may have a variety of objectives. The SWF becomes an ethical *theory* with ethical content when it postulates that social welfare is a function of individual utilities: it becomes the orthodox *theory* of welfarism. The nature of this theory is determined by which of several meanings of the term 'utility' is incorporated (see Richardson 1994). The important point, however, is that welfarism is no more than a *theory* about social welfare and its assumption rules out other plausible theories cannot be ruled out. For example, from a welfarist perspective, medical care could not be offered to a particular group simply because it saved life, improved health, was necessary for equity, or because it was required for some other theory of social justice. It could only be supported if utility was directly or indirectly increased. Medical care which reduced health but increased utility (possibly because patients substituted medical services for more efficacious self care) could not, ethically, be opposed by the welfarist.

Clearly this theory is not self evidently true. If utility is not defined tautologically to include all possible choices (which removes all content)[2], then it must be conceded that society may seek

[2] Every 'preference', opinion or vote may be defined as reflecting 'utility' and, for example, a person's decision to endanger their life for another may be defined as utility maximising because the person chose to take that action. This approach to the definition of utility is tautological and unhelpful. Every action must maximise utility because maximum utility is defined by taking the action. It is unhelpful as it confuses and impoverishes otherwise clear concepts such as 'duty', 'malice' and 'altruism' which are clearly understood and correspond with separate, possible behavioural motivations. Defining each as a sub-class of utility simply adds verbiage. Likewise, it is an observation that people give different ethical status to different actions and clearly separate the status of actions with concepts and words such as 'selfish' and 'selfless'. If all actions are defined as reflecting equal status utility, then the preferences of Adolph Hitler and Mother Theresa could not be distinguished (as both simply maximise 'utility') unless we separate utility

other objectives than utility maximisation, at least in some social contexts. This brings to the fore the unanswered question of how such objectives are to be determined and when they are relevant.

Rather than contribute to this important question, economists have commonly avoided it by assuming social objectives. Worse, there has been an appeal to the authority of 'economic theory' to support the contention that social objectives are (should be) those embodied in orthodox theory as defined by the acceptance of conventional axioms. For example, Johannesson (1999) argues that Dolan's 'empirical method … for estimating the shape of the social welfare function … has no theoretical foundation' (p381). By this it is meant that Dolan's theory does not build upon economic orthodoxy, ie except the axioms of orthodox economic theory. Gafni and Birch (1995) go one step further and argue that the validity of a measurement instrument 'stems from the validity of the theory (from) which the in
theory is chosen based on normative criterion alone one does not have to establish validity using classical psychometric methods' and (quoting Torrance's 1986 review on this subject) '… the standard gamble measurement technique is valid by definition because it is based directly on … the axioms (of expected utility theory)' p769. That is, the technique is justified by its consistency with one (albeit the orthodox) economic theory. This approach is unconvincing and hard to defend as the axioms and, therefore, economic theory are empirically wrong in this context an there are no good grounds for adopting the axioms in the normative context of decision making (Richardson 1994). This again raises the question of how to judge the acceptability of a normative theory concerning social objectives.

Murray, Salomon and Mathers (1999) are amongst the few who recognise that different SMPH's may be needed for the description and comparison of population health and for determining the social value of an intervention. They are also amongst the few who propose an explicit procedure to choose between alternative SMPH's specifically. They propose that Rawls' principle of a veil of ignorance be employed to assist with the determination of criteria for evaluating the former type of SMPH, ie a summary measure for the purpose of description but not evaluation. Five such criteria are then derived. While these criteria are shown to eliminate some SMPHs, it is questionable whether they justify the need for a Rawlsian veil. The criteria could almost certainly be justified without a veil; it is hard to imagine that anyone would disagree with the proposition that a useful definition of 'healthiness' would identify populations with lower mortality, age-sex specific incidence and prevalence of poor health, etc.

It may also be doubted that the veil of ignorance device will produce an unambiguous or even acceptable definition or criterion for 'healthiness'. Consider the two populations in Table 1. It is not clear that there is an unambiguous definition of 'the healthier' population. From behind a veil of ignorance the individual's choice will depend very largely upon their risk behaviour. In this example, a risk plunger or someone adopting a maxi-min criterion would prefer membership of

derived from selfish and from selfless actions. The term 'utility' then adds nothing but an additional layer of language; but unless this cumbersome distinction is made, our language and concepts would be impoverished.

population 1. Someone who was risk averse or maximised the expected value of health would select population 2. Normally, however, the concept of 'healthiness' is unrelated to a person's attitudes towards risk and uncertainty and it is unclear what advantage this dependence upon risk attitude achieves. More generally, it is unclear what people will think behind a veil of ignorance. In principle, the choice could be dominated either by issues of self interest or by beliefs concerning social justice. In sum, while the veil of ignorance represents an instrument for encouraging a particular perspective, it is not known what the result of this will be[3].

Table 1 Distribution of Population by QoL

QoL (Utility Score)	Population 1	Population 2
1.0	0.5	
0.8	-	0.9
0.4	0.5	-
0.0 (Dead) 0.1	-	0.1
Total	1.0	1.0
E(U)	0.70	0.72

Ethical and economic theory does not and cannot provide an unambiguous answer to the question 'What criteria should be used to assess an SMPH appropriate for assisting with the evaluation of health interventions?'. Unlike positive theory which may be tested empirically, normative or ethical theory can only be judged using some normative criteria. (For example, we may ask whether principles adopted are those which accord with moral intuition; – which accord with a religious view; – which are logically consistent with other commonly accepted ethical behaviours ... etc.) However to demonstrate that a particular set of criteria is 'correct' in some sense requires the application of 'meta-criteria' which themselves need justification. That is, any attempt at ultimate justification leads to an infinite regress[4]. Eventually, a moral theory or a set of criteria must simply be stated and 'sold' in the' intellectual market' place. This is, essentially, what is done by Murray et al (1999) to establish minimal criteria for evaluating an SMPH appropriate for the burden of disease. Orthodox economists explicitly – or more commonly,

[3] As a second example an individual might vote for state A (99 percent full health; 1 percent paraplegic) because they considered it to maximise their self interest. The person might simultaneously believe, that state A was unfair to the resulting paraplegics who had opted for state A and were 'needlessly unlucky' and that State B (100 percent close to but less than full health) was more just. A second person might, however, vote for state B because they place justice above self interest. That is, from behind the veil any outcome is, in principle, possible. The veil would only be a useful device if either the population accepted it as a means for achieving procedural justice or if they accepted that the state of health it implied was the best option. But with this second justification the veil would be unnecessary and a simple vote would suffice.

[4] An infinite regress is ultimately avoided either by negotiation (and the acceptance of procedural rules) or by the referral of the issue to Parliament. That is, ethical decisions are resolved in practice by the acceptance of certain procedural rules and an instrumentalist view of ethics might be to accept that the final judgement after procedural justice has been fulfilled represents the ethically 'correct' conclusion. Certainly an analyst could accept this instrumentalist interpretation even if, personally, he or she disagreed with the outcome.

implicitly – attempt to 'sell' the context free axioms of the expected utility hypothesis as the basis for their normative theory that individual utility 'should' be measured using the standard gamble. Elsewhere I and others have argued that there are context specific reasons for the systematic and rational violation of the axioms and that, because of these they should not be 'bought' as the basis for theory or policy[5] (Richardson 1994).

In this earlier study, I proposed four criteria for evaluating the units of an SMPH appropriate for evaluation and, more specifically, appropriate for the assessment of the scaling techniques used to convert life years into units of social value (TTO, SG, PTO, RS, ME[6]). I believe these four criteria are still appropriate (and should be sold!) They are that:

1 More units are considered to be of greater social value

2 The units should have a clear unambiguous meaning

3(a) The units should have a 'weak' interval property; viz, incremental units should, in some *easily understood* sense, mean the same irrespective of the number of units already obtained.

3(b) The units should have a 'strong' interval property; viz, that an x percent increase in measured quality of life at any point along the QoL spectrum should have, in an easily understood way, the same value as an x percent increase in the length of life.

4 The scaling techniques should be sensitive to a change in a health state and be reliable and valid.

The reason for criteria 1 and 3 are relatively self evident. Criterion 2 is included because, firstly, the SMPH must be 'sold' to non expert decision-makers. Secondly, units will be 'traded' against other objectives. This cannot be done if the decision-maker cannot easily comprehend the meaning of the SMPH.

I now believe an additional criterion should be added which was only implicit in criterion 1 above; viz,

5 The SMPH should embody ethical values which are consistent with stable population values and which reflect any relevant context dependent ethical values.

[5] For example, there is no good reason why individuals should ignore feelings of hope, anxiety, fear, anticipated regret, etc or any of the other utility relevant factors that occur before the outcome of a risky situation is known. Such 'pre outcome' emotions are ignored by expected utility which is only concerned with the utility of the final outcome states.

[6] TTO = Time Trade-off; SG - Standard Gamble; PTO = Person Trade-off; RS = Rating Scale; ME = Magnitude Estimation.

5 Empirical Ethics

Criterion 5 is proposed explicitly to encourage 'ethics empiricism' or 'empirical ethics'; viz testing the consistency of hypothesised ethical values against population values. The enquiry is, therefore, positive: it seeks to determine population values. As the analysis proceeds, it seeks to distinguish superficial, spontaneous values – opinions or reactions – from deliberative responses; that is, to answer the question 'what are the population's stable ethical values after enforced reflection?' The proposal is explicitly designed to discourage empirical free theorising about ethical views which the population may have (or, as in some analyses, views which it is asserted the population should and therefore does have because individuals are 'rat

In sum, it is suggested that ethical views including those embedded in economic orthodoxy should be elicited in an iterative way. Researchers should postulate population values (ethical principles) and then embark upon a series of empirical studies, both qualitative and quantitative. During these, the implications of population responses should be clarified. (For example, the implications of the 'strong interval' property should be made explicit, viz, that in QoL measurement for the purpose of QALY construction, a ten percent drop in quality is of equal importance-(value)-as a ten percent drop in the quantity of life). Postulated ethical principles should be reformulated in view of population reaction to this information and then 're-submitted' to empirical testing. The process should continue until acceptable, stable (reliable and deliberated) ethical principles are identified, ie principles that withstand both a priori ethical criticism and the test of population support. The information obtained from this procedure should then be 'fed into' the decision making process. Under some circumstances (which may be the subject of empirically informed debate) it may be desirable to adopt the deliberative responses from the population as the appropriate indicator of social value and, therefore, policy should be directly based upon these responses. In other cases it may be more appropriate for decision makers to be 'informed' about population values; that they should exercise discretion and modify the direct policy implications; that is, trade-off population against other views. Finally there are a class of population values – hopefully small – where it would be expected that decision makers would override or dismiss population values (racism, sadism etc).

As discussed and illustrated in a companion article (Richardson 2000) it is possible to empirically examine the 'meta issue' of whether or not a particular moral issue *should* be decided by, or even influenced by, public opinion. There is no inconsistency in the public indicating, on the one hand, that they, personally, would prefer outcome A to outcome B but that, on the other hand, they would prefer the decision to be made by government. Each individual, recognising that they have only one vote (and that even this may be the result of a less than full appreciation of the issues) may have greater confidence in a government based decision process than one based upon the manipulable voting patterns of fellow citizens. Further empirical evidence on the acceptability of 'meta criteria' is probably infeasible[7].

[7] Voting upon the meta issue about the use of empirical evidence is conceptually challenging and it is almost certainly infeasible to move further along the infinite regress to empirically test the acceptability of the meta question. The rules of the debate at this point must be established independently (see footnote 3 earlier).

While numerous studies have been carried out which investigate people's attitudes to ethical issues, there has been little discussion of the status of such evidence and much of it is open to the criticism that the public opinions elicited are the result of 'spontaneous' rather than 'deliberative' answers. That is, there have been limited attempts to encourage deep thought, debate and reflection before responding to ethically difficult issues which may not previously have been considered. This proposed process will be difficult. Evidence suggests that most people do not have clearly articulated and consistent ethical views and, consequently, responses to simple questionnaires may be artifacts of the questionnaire framing. It is also true that, in principle, population values may be unacceptable (racist, sexist and nasty!). This problem has been discussed by Goodin (1986) who argues for the use of 'laundered preferences': preferences that are accepted or rejected after ethical scrutiny. Where and how the process of 'laundering' cuts in is unclear but it is probable that, as an empirical fact, there will be a broad domain of principles and values which may legitimately be accepted or rejected by their correspondence with community values. The alternative to this suggestion is to exclude the population from ethical decision making and return it to 'expert ethicists' who have, to date, failed to provide a consensus answer.

As demonstrated earlier the use of empirical ethics could be criticised because it will not lead to ultimate justification: any such attempt would lead to an infinite regress. To this extent 'empirical ethicism' is no different from other disciplines. Many of the physical sciences, economics and indeed mathematics itself are not based upon universal truths applicable in every possible context. The foundations of physics have undergone a significant evolution. This does not mean that there has been no progress made in these disciplines in the last thousand years or that useful policy conclusions cannot be drawn from what we currently know about these disciplines.

6 Implications and Evidence

Four issues are briefly discussed below. Three arise from the contention that the conceptual basis of an SMPH should take account of the results of empirical ethicism. Each of the issues illustrates a different facet of empirical ethics. The first, the achievement of a strong interval property, illustrates the type of information that may be provided to survey respondents to encourage deliberation and to challenge the validity of a utility score provided before deliberation. The second issue, double jeopardy, illustrates the relevance of context for ethical decision making. The third, choice of scaling instrument, highlights the need for explicit criteria for the evaluation of scaling techniques as distinct from a reliance upon 'orthodox theory'. Finally the choice of perspective is an overtly ethical issue which may have a quantitatively significant impact upon the numerical values of an SMPH.

The Strong Interval Property

When DALYs or other QALY-like SMPH's are used for evaluation, their most fundamental property is that they represent an exchange rate between the quality and quantity of life. A ten

percent increase in either component has the same impact upon the numerical value of the SMPH. Consequently, an error in the utility index could, quite literally, be lethal if the scores were used to allocate resources between programs which enhance the quantity and quality of life. This was highlighted by Nord et al (1993) where it was shown that the utility scores in the QWB instrument implied that curing four and five people with a cough and with pimples respectively were both equivalent to saving a life.

Data reported in Table 3, (explained below) imply that the value of returning the 'average' member of the non-hospitalised community to (self assessed) 'good health' is almost three times as great when the value is measured either by the patient themself or by the AQoL instrument as compared with the value measured by the 15(D) or HUI. There is, therefore, a corresponding threefold difference in the quantity of life which could be justified by this improvement. Despite its quantitative importance virtually no attention has been given to validating this strong interval property.

This issue is almost as significant for the measurement of the BoD using an SMPH. The adoption of a utility score of 0.9 when the true value is 0.95 does not appear too serious. However it will result in double the number of DALYs or units of another SMPH when it is used for measuring the BoD. This point was also made by Nord (1993) in relation to QALYs. Once again, this issue has received almost no attention in the literature although it is fundamental for the plausibility of the DALY/QALY.

The Relevance Of Context (Double Jeopardy)

With a context free assessment, the utility score assigned to a health state such as quadriplegia would not depend upon the background history of the patient. This exposes long-term quadriplegics and other permanently disabled people to a 'double jeopardy'[8]. The acceptability of ignoring the historical context was tested empirically by Ubel et al (1999). Three scenarios were described. In each, one of two health programs had to be selected. The first saved the life of a hundred patients and returned them to full health. The second returned a number of patients to quadriplegia. This second number was determined using the PTO technique. In the three scenarios the quadriplegia was pre-existing (scenario one); a consequence of the disease in a previous healthy patients (scenario two); and avoidable: previously healthy patients would be returned to quadriplegia but at a lower cost than the cost of cure (scenario three). As shown in Table 2 the context had a highly significant impact upon the results. Respondents did not discriminate against pre-existing quadriplegics but treated the value of a cure resulting in quadriplegia as significantly less for the second and third groups.

[8] The term 'double disadvantage' would be more descriptive. The long-term quadriplegic is first disadvantaged by being a quadriplegic. He/she is further disadvantaged because the value of any life, restoring intervention will be less than the value from saving a normal patient if the health state 'quadriplegia' is always treated as having less value than normal health.

Table 2 The Importance of Context in the Evaluation of Paraplegia

Context	Median Number Equivalent to Returning 100 from Death to Normal Health	*n*
Pre-existing paraplegia	100	100
Onset paraplegia	5,000	85
Avoidable paraplegia	500,000	66

Source: Ubel, Richardson, Pinto Prades, 1999.

These results support the earlier argument that a distinction must be made between the value of a program and the utility gained from improvements in QoL. As the contexts are quite different in these three scenarios, it is not possible to argue that respondents were irrational. Their response, summarised in Table 2, is consistent with a broader view of utilitarianism which recognises that there would be a significant loss of utility for paraplegics for the entire duration of their life if they rightly believed that they would receive lower priority in any rationing of medical services. That is, utility maximisation might require the elimination of this source of disutility by positive discrimination in the narrow context of prioritising life saving treatments. Alternatively, the data in Table 2 are consistent with several non-utilitarian theories of justice. The chief significance of these results here is that they suggest that ethical values which are context specific may drive a wedge between narrowly defined QALY maximisation and social objectives.

Choice Of Scaling Instrument

In Richardson (1994) I used the first four criteria summarised above to assess the relative merits of the available scaling instruments; viz, the rating scaling (RS), magnitude estimation (ME) the standard gamble (SG) the time trade-off (TTO); and the person trade-off (PTO) instruments. It was argued that (i) the RS and ME did not have a clear meaning and that there was no possible link between them and the (all important) strong interval discussed earlier; (ii) that, by contrast, the three trade-off instruments all required a comparison of the quantity and quality of life; (iii) that the meaning and the interval property of units derived from the SG were seriously confounded by the (irrelevant) risk context of the instrument; and (iv) that the choice between the PTO and TTO should be determined largely by the perspective – personal or societal – that was to be embodied in the instrument. While still supporting the framework for assessing the instruments (viz the application of explicit criteria) I would now modify this conclusion somewhat.

All three trade-off instruments are confounded and for the same reason. The technique used to elicit a numerical score, viz, varying risk of death, length of life, and distribution of benefits introduces a (normally) extraneous factor into the assessment. The risk embodied in the SG will normally be quite different to the risk facing a patient which may not entail any probability of death. The TTO is confounded by time preference and the different life expectancies in the

contrived TTO question. The PTO introduces a distributional consideration which will normally be quite dissimilar to the distributional implications, if any, of a program under review.

Supporters of each scaling instrument may seek to defend its integrity by arguing that what is called an extraneous element here – risk, time or distribution – is, in fact, a part of what is to be measured. In the case of the standard gamble advocates have gone one step further and argued that the introduction of the extraneous element – risk – is necessary for the measurement of 'utility'. Values calculated in the shadowy world of 'Under-risk' are mutated in some way due to their contact with risk – not the risk associated with the real world context of the decision, but with any risk – and magnitudes emerge as 'utility' which, because of the connotations of the word, is presumed to be an accurate reflection of the intensity of a person's preferences and the appropriate object of measurement. The world of 'Under-risk' is described here as 'shadowy' because the way in which the risk of instant death captures the intensity of a person's preferences with respect to the relief of, for example, pain and the risk that the paracetemol taken will not relieve the pain, has not been explained in the literature and the relationship implied appears more mystical than scientific.

In all three cases the argument that the extraneous factor is justified cannot be supported. The contentious element – risk, time or distribution – is an artifact introduced for the purposes of measurement and not for description. In each case, the instrument varies one dimension of a contrived health state scenario; *viz*, its duration, the risk (usually of death) or the number of people affected. This permits the unique identification of only one – not two – elements of the real world health state and the stated purpose of each instrument is that this element should be the health related quality of life – not the real world risk, duration or distributional effect. In sum, a single instrument cannot (accept by chance[9]) simultaneously measure two unrelated and fixed real world values. The truth of this proposition becomes self evident when the relationship between targets and instruments is formalised[10]. The impossibility of the task is only made more evident by the fact that the second element – the real risk, time or distribution associated with the health state is not normally included in the health state description presented to patients.

Other sources of bias in these instruments may exist. For example, it has been noted in the literature on adaptation that the utility score assigned to poor health states by those who have experienced them for a significant time it generally greater than the score assigned to the same

[9] It is possible that a particular procedure might expose a person to a real risk of death of, for example, 0.4 and that the utility value of the health state before the procedure was 0.6. Then and only then would it be possible to argue that the standard gamble captures the relevant attitude towards risk. An analogous argument applies with respect to the TTO and duration; the PTO and distribution.

[10] In the abstract theory of policy it is known that, except by coincidence, the simultaneous achievement of n objectives requires n policy instruments. More formally, a system of n simultaneous equations – one determining the value of each policy objective – can only be solved if there are at least n independent variables whose values may be adjusted as policy instruments. In the present case, n = 2: we wish to set both QoL and the other variable – risk, time or distribution – at their real world values and 'solve' for the numeraire or instrument variable. But with only one instrument we can only 'solve' for (determine the equivalent value of) one not two – of the real world variables and the existence of two variables confounds the relationship – there is no unique solution. The solution obtained will depend upon the risk, time or distributional element described in the scenario and also embodied in the instrument and each of the two other variables.

state by a person contemplating the possibility of entering that state. Evidence suggests that a contributory factor to this discrepancy is that in the second case those contemplating the state will typically be influenced by the *sadness of entering the state*. To the extent that this is generally true and to the extent that we wish to measure the utility of a state per se then the person trade-off instrument is subject to a second confounding effect (Kahneman 1999) *viz* a confusion of the health state disutility and the sadness – disutility – of the process of entering the state.

The importance of these confounding factors depends upon their magnitude. My earlier article implicitly assumed that the extent of the bias introduced by artificial risk was greater than the bias introduced by the TTO and PTO. While this may be true it is an empirical issue and not one that maybe resolved a priori. Dolan and Gudex (1995) find a zero discount rate implied by their TTO results which suggests that there is no time distortion associated with the TTO. But this is a single study in need of replication.

A second and similarly empirical issue is the extent to which survey respondents understand and respond accurately to the three types of questions. Economics, as a discipline, has a tradition of assuming rationality and good information and this is reflected in the implicit assumption that the answer received from respondents will literately and precisely reflect the considered opinion of the respondent. Increasing evidence suggests that this is not so and that the validation of all instruments requires a careful consideration of the meaning of the information obtained from respondents.

Closely related to this issue is the common assumption that the (almost) spontaneous answers received from respondents are valid. Studies which have encouraged deliberation have not found empirical support for this assumption (Murray and Lopez 1996; Dolan et al 1999). However, there is no good evidence about the techniques which best promote deliberation. The methods in the studies cited above may or may not improve deliberation. A recent search by the Monash Health Economics Unit failed to identify a literature discussing and validating alternative procedures for encouraging deliberation. Perversely, some doubt was cast upon the intuitively plausible use of focus groups.

In sum, for the reasons given by Richardson (1994) and by Dolan and Gudex (1996) the TTO may well be the most appropriate scaling instrument for estimating utility scores. If the desired concept of social value was simply individual utility then this would imply that the TTO might also be the scaling instrument of choice. However if social value is conceptualised (if we find it most useful to conceptualise) in terms of population values, then the PTO would probably become the instrument of choice. As the other elements of social value – age weights etc – are not experienced by patients then the use of the PTO and the values of the general population are to be preferred. However these issues depend, to a significant extent, upon the empirical questions associated with instrument bias, respondent comprehension and deliberation and upon the perspective which society or its representatives wish to have embodied in health state valuations. None of these are ‘technical issues’ to be resolved by a priori ‘economic theory’.

Patient Versus Public Perspective and Values

There is a significant literature on whose values should be incorporated into an SMPH. It is reviewed in both Brazier (1999) and Richardson et al (1998). While there is a broad consensus that a patient perspective should be adopted, it is commonly argued that the values incorporated should be those of a cross section of the population as it is the community, as taxpayers, who pay for health programs. This argument is neither necessary nor compelling. Taxpayers do not specify how their funds should be spent in other contexts. We do not vote on the composition of the armed forces, the location of roads etc. and we do not specify how recipients of social service payments should spend their money. It is perfectly reasonable for taxpayers to fund health services which are prioritised by others.

In a recent contribution, Nord et al (1999) combined elements from both sides of this argument to produce a proposal for a two stage procedure in which the patients' perspective and values are used to produce utility scores and societal representatives used to convert these into units of social value. It should be recognised, however, that these decisions are ethical and economists have no particular expertise in this area. At best, empirical ethicism may be employed to gain insights.

Results of one such study are reported below which was designed to observe the relationship between patient values and those of the general community. The underlying hypothesis was that, in addition to the types of ethical issue discussed above, the general population may have difficulty in comprehending, – 'appreciating' – and evaluating health states that are significantly different from those that they have experienced and that, consequently, some of the solutions discussed on the assumption of full information and comprehension are infeasible or subject to significant distortion. This is, however, one of the ethical issues where economists have no particular expertise and empirical ethicism is relevant.

Table 3 is extracted from a larger validation study of three well known MAU instruments and the (Monash Assessment of Quality of Life) AQoL instrument. 129 members of the community or hospital inpatients were asked to rate their own health state using the TTO. More specifically, each person was asked to nominate what proportion of their life they would sacrifice to move from their current health state to a 'good' health state (as defined for them). Each respondent was personally interviewed and asked to complete the four MAU instruments and the SF36. As the 'best health state' defined by these instruments differed, standardisation was achieved by re-scaling. For each instrument, an average score was calculated for respondents who self rated their health as 'very good' or 'better' on the SF36 instrument. Each of the instrument scores were then rescaled by multiplying them by the reciprocal of this average score. For each instrument this resulted in a scale in which those with this average score now obtained a score of 1.0. Respondents with higher scores were also set equal to 1.0. Consequently, for each instrument the end points were defined by death and good health which assumed the values of 0.00 and 1.0 respectively. The procedure results in a scale with identical end points for each of the instruments which is a prerequisite for valid comparison.

Table 3 Personal *vs* Population Values: Evidence from 4 MAU Instruments

	Rescaled so SF36 Self Rated 'very good health' = 1					*n*
	TTO	AQoL	EQ5D	HUI2	15D	
Community	0.86	0.87	0.90	0.94	0.95	39
Inpatient	0.76	0.52	0.62	0.68	0.82	90
Adjusted		0.52	0.59	0.62	0.74	

Key: Rescaling is described in the text.

The most striking result from this comparison is that while the average self-rated TTO for community respondents was similar or lower than the self TTO scores, the average utility for hospital inpatients was significantly lower on each of the MAU instruments, except for the 15D which consistently gives higher utility scores than other instruments. As all the instruments except the AQoL predicted higher community values than the community members themselves, a further adjustment is reported in row three in which the EQ5D, HUI2 and 15D are further adjusted so that the average MAU score for community respondents equals the average self reported TTO score. The result is that for the first three instruments, predicted utilities drop 24, 17, and 14 percentage points below the self rated TTO score and even the adjusted 15D gives a lower predicted utility.

The most probable interpretation of these results is that the community values incorporated in these MAU instruments are not too dissimilar to the directly elicited community values. However they significantly undervalue the utility scores of those in more serious health states. This suggests a failure to fully comprehend and appreciate the significance for QoL of these health states. An alternative explanation is that patients adapt to health states and the community respondents, while knowing this, choose to adopt pre-adaptation values for ethical reasons. This alternative hypothesis seems implausible. Hospital patients have recently entered the health state and would not, generally, have adapted. It is unlikely that community respondents had the sophistication of thought to draw the hypothesised conclusions during the space of a relatively brief interview. The conclusion supported by these results, therefore, is that unless the judgement is made that assessed utility scores are appropriate then use of community values is questionable.

Table 4 presents more direct evidence on this issue. As part of a larger, pilot population mail questionnaire respondents were asked their opinion about the use of patient versus public values[11]. The question did not emphasise the likelihood of patients' preferences being misinterpreted by the general public as suggested above. The wording of the respective

[11] The survey is discussed in somewhat greater detail in the companion paper.

questions emphasised the argument for community decision making. Despite this, the results reported indicate a rejection of the common approach and a preference for the use of patient values.

Table 4 Patient *vs* Public Values: What the Public Thinks
n = 67

Government decisions about health spending take account of many things, **including how unpleasant it is to have different illnesses**. When deciding how unpleasant illnesses are:	Percent Response
the Government should listen mostly to patients, because they know best how unpleasant their illness is;	58
the Government should listen mostly to a well-informed sample of the public, who have been told about the different illnesses, because they represent the community who must pay for the health care.	42
TOTAL	**100**

7 Conclusions

The following conclusions are supported by this paper

1 'The conceptual basis' of an SMPH should be based upon ethical considerations. When answering the normative question 'what should we do or what policies should be implement', the distinction between 'efficiency' and 'equity/ethics' is largely artefactual. Improvement in social welfare cannot be value free, and this applies to the benefits of improved 'efficiency'. As a minimum, a judgement must be made concerning the meaning of 'improved'. If this term implies an increase in something which is valued, then there is the unavoidable ethical question of what it is that is valued. In practice, policies designed to increase 'efficiency' normally involve a re-distribution of benefits and, therefore, impinge upon ethical issues and this cannot be avoided by use of the concept of 'potential Pareto efficiency'. The appeal to 'economic theory' to justify policies which have ethical content is misleading. It implies a wrong perception of the role of economic theory.

2 There has been a surprising neglect of the criteria which should be used for evaluating policies or measurements which have implicit or explicit ethical implications. Criteria exists which should attract widespread support.

3 There is a need for the increased conduct of 'empirical ethicism', viz, studies which elicit and validate population values and expose them to ethical challenge. The information

obtained from this process should be provided to decision makers. How this information ought, ideally, to be used should be the subject of further debate.

4 The evaluation of SMPHs has almost totally ignored the strong interval property – the equivalence of quality and quantity effects. This is despite the strong interval property being, arguably, the most important property for the validity of an SMPH in the context of either evaluation or the measurement of the burden of disease.

5 Context free ethical analysis is likely to be invalid. Evidence suggests that context may often have ethical significance for the population.

6 The choice of scaling instruments should be guided by empirical evidence on the extent to which the three trade-off instruments are subject to distortion and embody the desired perspective.

7 The determination of techniques which promote deliberation and the eliciting of valid population responses should be very high on the research agenda and many of our accepted conclusions should be revalidated using such procedures.

8 Public valuation of the utility of severe health states probably exaggerates the QoL loss and exaggerates the benefits of cure. With fixed expenditure this could result in programs favouring quality enhancement over life saving. Limited evidence suggests that the population does not support the common argument for the use of public values for the scaling of SMPHs.

References

Brazier J, Deveril MM, et al 1999, 'A review of the use of health status measures and economic *Health Technology Assessment*, vol 3, no 9, pp 1–161.

Dolan P and Gudex C 1995, 'Time preference, duration and health state valuations', *Health Economics*, vol 4, pp 289-299.

Dolan P and Gudex C 1996, 'Valuing health states: A comparison of methods', *Journal of Health Economics*, vol 15, pp 209-231. **(CHECK THIS REFERENCE)**

Dolan P, Cookson R and Ferguson B 1999, 'Affect of discussion and deliberation on the public's views of priority setting in health care: focus group study', *British Medical Journal*, vol 318, no 3 April, pp 916–919.

Gafni A and Birch S 1995, 'Preferences for outcomes in economic evaluation: An economic approach to addressing economic problems', *Social Science and Medicine*, vol 40, no 6, pp 767-776.

Goodin RE 1986, 'Laundering preferences' in *Foundations of Social Choice Theory,* J Elster and A Hyllind (eds), Cambridge University Press.

Johannessen M 1999, 'On aggregating QALYs: A comment on Dolan', *Journal of Health Economics*, vol 18, no 3, pp 381–386.

Kahneman D 1999, 'How can we know who is happy: Conceptual and methodological issues' in *Wellbeing: The Foundations of Hedonic Psychology*, D Kahneman, E Diener & Schwarz (eds), Russell Sage Foundation, New York.

Menzel P, Gold M, Nord E, Pinto Prades J, Richardson J and Ubel P 1999, 'Towards a broader view of values in cost effectiveness analysis of health', *The Hastings Centre Report*, vol 29, no 3, pp 7–15.

Murray C and Lopez A 1996, *The Global Burden of Disease*, WHO/Harvard School of Public Health/World Bank.

Murray J, Salomon J and Mathers C 1999, 'A critical examination of summary measures of population health'. Paper presented to the WHO Conference on Summary Measures of Population Health, Marrakech, 6-9 December.

Nord E, Richardson J and Macarounas-Kirchmann K 1993, 'Social evaluation of health care versus personal evaluation of health states: evidence on the validity of four health state scaling instruments using Norwegian and Australian survey data', *International Journal of Health Technology Assessment*, vol 9, no 4, pp 463-478.

Nord E 1993, 'Unjustified use of the quality of wellbeing scale in Oregon', *Health Policy*, vol 24, pp 45-53.

Nord E 1999, *Cost Value Analysis In Health Care*, Cambridge University Press.

Nord E, Pinto Prades, J, Richardson J, Menzel P, and Ubel P 1999, 'Incorporating societal concerns for fairness in numerical evaluations of health program' *Health Economics*, vol 8, pp 25–39.

Olsen J and Richardson J 1999, 'Production gains from health care: What should be included in *Social Science & Medicine*, vol 49, pp 17–26

Olsen J, Richardson J and Mortimer D 1998, 'Priority setting in the public health service: Results of an Australian survey', Technical Report 9, CHPE.

Richardson J 1994, 'Cost utility analysis: what should be measured', *Social Science & Medicine*, vol 39, no 1, pp 7–21.

Richardson J 2000, 'Age weighting and discounting: What are the ethical issues?' Working Paper 109, CHPE.

Richardson J, Olsen JA, et al 1998, 'The measurement and evaluation of utility based quality of life: An introduction and overview of issues and options', Working Paper 97, CHPE.

Tsevat J, Dawson N, et al 1998, 'Health values of hospitalised patients eighty years or older', *Journal of American Medical Association*, February 4, vol 279, no 5, pp 371–375.

Ubel P, Richardson J and Pinto Prades J 1999, 'Lifesaving treatments in disabilities: Are all *International Journal of Technology Assessment In Health Care*, vol 15, no 4, pp 738-748.