MONASH University
Business and Economics

Centre for Health Economics

Research Paper 2009 (38)

Why we should not minimise cost per QALY

Theoretical and empirical evidence

Professor Jeff Richardson
Foundation Director, Centre for Health Economics
Monash University

Dr Kompal Sinha
Research Fellow, Centre for Health Economics
Monash University

Mr Angelo Iezzi
Research Fellow, Centre for Health Economics
Monash University

July 2009

Centre for Health Economics
ISSN 1833-1173
ISBN 1 92118737 9

ACKNOWLEDGEMENTS

The research described in this paper is made possible through the National Health and Medical Research Council (NHRMC) Project Grant ID 491162 and NHMRC Professorial Fellowship ID 284363.

ABSTRACT

In the present framework for economic evaluation benefits and costs are ‘disembodied’ from individuals and their characteristics. The decision criterion of minimum cost per life year or QALY is uninfluenced by issues of fairness which, in the welfare theoretical framework, can be achieved independently. We argue here that the assumptions which allow this conclusion cannot apply in the context of a National Health Scheme. Public preferences are therefore likely to embody both efficiency and fairness considerations in a way that is inseparable. Evidence obtained from a representative cross section of the Australian population suggests that fairness may dominate efficiency in terms of its quantitative significance. This implies that a primary focus upon cost per QALY or the implicit use of a cost per QALY threshold across all patients may be misleading.

TABLE OF CONTENTS

List of Tables

List of Figures

Why we should not minimise cost per QALY: Theoretical and empirical evidence

1 Introduction

Cost Effectiveness and Cost Utility analysis seek to minimise cost per life year or Quality Adjusted Life Year (QALY) and this objective is embodied in policy advice. The theoretical rationale for this is usually based upon Welfare Theory and, in particular, the assumption that equity may be achieved by a redistribution of the efficiently produced output or outcome. This is unproblematical in the case of technical efficiency. However achieving economic efficiency necessarily discriminates between projects or services and in the case of health care it therefore discriminates between patients.

Particularly in the context of a National Health Scheme established for reasons of fairness, the efficient distribution may conflict with social objectives. This is widely acknowledged in the literature in both principle and practice: there is the need for an equity-efficiency trade-off. This has been discussed particularly in the context of the health gradient between different social classes. There is also a growing literature which is focused upon age related social preferences for health and health care and upon disease severity as reasons for departing from maximum efficiency. Other reasons for an equity-efficiency trade-off have also been explored.

The present article is concerned with one of the least discussed, but most problematical examples of the equity-efficiency trade-off, namely the distributional effects of differences in the cost of treatment. Individuals who are afflicted by a high cost disease generally have as little or less responsibility for the disease than other individuals have for being in a particular social class. But economic theory implicitly recommends that such individuals should have lower priority access to health services. In principle, this will allow more health to be produced. Similarly, ignoring equity and focusing solely upon efficiency may allow the greater production of health. But it results in a distribution of health which may or may not correspond with social values. We report below the results of a study carried out to investigate the extent to which the public would be prepared to deviate from health maximisation in order to overcome 'discrimination' caused by the varying costs of health care. In the section below we discuss the theoretical framework and the expected trade-off between cost and equity enhancing effects of discounting cost. In the third section we outline the methods used to test and quantify the trade-off between 'cost and fairness'. Survey and analytical results are presented in Sections 4 and 5 respectively, and the implications for the practice of economic evaluation in Section 6. These include the conclusion that we should not, as a general rule, seek to minimise cost per life year or cost per QALY, which is to say that there is no simple threshold value of life or a life year which can be used in conjunction with league tables to decide whether or not to fund a health service.

2 Theoretical framework

The advice to minimise cost per life year or cost per QALY is based (somewhat uncomfortably)[1] upon Welfare Theory. The logic of this is, in turn, based upon a number of assumptions. Given these, maximising net present value of an intervention allows the possibility of a redistribution of benefits to achieve Pareto efficiency. If the Kaldor-Hicks potential compensation criterion is accepted and Pareto efficiency is deemed to be socially beneficial then maximising net present value is socially beneficial.

There is now a large literature discussing the various assumptions and consequences that follow if they do not apply (see in particular Rice (1998) and Richardson and McKie (2007). For the issue in the present paper two are of particular relevance. First the logic of this argument depends upon redistribution actually occurring or the value judgment that if net benefits are positive and a redistribution is, in principle, possible then the situation *should* be considered to represent social improvement. However, in the health sector it is often impossible, even in principle, to compensate losers. If an individual dies or suffers significant impairment because resources for possible effective treatment have been used for other patients then compensation may be impossible. When less effective therapies are offered which leave a patient worse off no health scheme in the world pays compensation unless this takes the form of a disability pension and even then it is unlikely that this will compensate for a return to full health.

To our knowledge the value judgement that a state *should* be treated as being better because of the *possibility* of compensation even though this *does not actually occur* has never been 'tested' empirically against public opinion. Typically the principle is simply asserted, not defended. Indirect evidence cited below (*cf* the Ultimatum Game) strongly suggests that it will be widely rejected. The population exhibits a very strong preference for people not to be favoured over others and this preference is likely to be intensified by the knowledge that compensation could have been (but was not) granted. Without the Kaldor-Hicks principle, programs which favour some, but not others, cannot be evaluated without an implicit or an explicit value judgement based upon some non technical criterion.

This conclusion is reinforced by the evidence that when an individual's preferences are not simply a function of the material benefit they personally receive but are also a function of other people's behaviours and benefits then the relationship between Pareto efficiency and social improvement becomes problematical. That is, even if the Kaldor-Hicks principle was accepted and Pareto efficiency potentially achievable this potentially efficient state would not necessarily increase social welfare. Sen (1997), for example, has demonstrated in purely logical terms that interdependent preferences may result in a conflict between Pareto efficiency and liberalism: that achieving Pareto efficiency might require coercion. But if adherence to liberal principles in a particular context is deemed more important than achieving Pareto efficiency then this policy recommended by Welfare Theory in this context may be unacceptable or reduce social welfare.

Sen's argument may be generalised. Even when it is not logically inevitable (as in Sen's case) it may be empirically true that some non welfarist principle may be empirically more important (ie to the population) than Pareto improvement. In a social context relevant considerations may include egalitarianism: various forms of duty (to family, country or as taught by a religion) and related to these, various communitarian principles. Envy is a source of potentially strong social preferences.

[1] The seminal reference here is Garber and Phelps (1997). A key question of interest in the theoretical discussion is when, ie

While it is a negative expression of 'equity' there is no technical or logical reason why preferences based upon it should be censored.

Each of these motivations could, in principle, result in a conflict between Pareto efficiency and social welfare: one person's improvement might reduce overall wellbeing.[2] A particularly interesting and well documented example of the conflict arises as a result of yet another related example of socially motivated preferences, namely reciprocity. In the so called 'Ultimatum Game' one individual divides an amount of money between themselves and another person who has the ability to either accept the offered amount or to apply a veto to the choice in which case neither person receives any money. If people accepted the superiority of Pareto efficient states then the second person would accept any positive amount of money offered. They would be better off and the country (of two people) would be better off. However, this does not occur (Guth, Schmittberger et al. 1982). The game which has been widely played for real money in countries as diverse as Indonesia, Russia, France and the USA finds that individuals overwhelmingly reject an ungenerous offer (usually anything less than 30 percent of the total amount), a response which is interpreted as anger and punishment arising from the inequity of the division of money and the lack of reciprocity on the part of the first person (Beinhocker 2007). Interestingly, people with autism who lack the capacity to empathise with others have been observed to play the game like the hypothetical agents in an economic textbook, maximising their own reward and attempting to minimise the amount given to a second player.[3]

The Ultimatum Game is a striking empirical illustration of the fact that, for better or worse, people's preferences are strongly influenced by the state and behaviours of people with whom they interact. The Pareto criterion takes no account of this and the judgement that Pareto improvement represents social improvement must, subject to the caveat below, implicitly assume that such influences *should not* be taken into account in social judgements. This judgement is not justified in Welfare Theory.

The results of the Game also suggest that the Kaldor-Hicks criterion would be widely rejected as an ethical principle and that 'potential' should not be treated as equivalent to 'actual' compensation. The Game indicates a desire to seek punishment when benefits are not actually shared 'fairly'. While the two contexts are not identical – one deals with compensation for loss, one with windfall gain – the strength of the concern over the fairness of the division of benefits revealed by the Game suggests the unlikelihood of widespread acceptance of an uncompensated loss especially in view of the evidence concerning the greater psychological importance of losses as compared with gains.

The Kaldor-Hicks criterion may be reinterpreted as advice that economists should not bear responsibility for the redistribution which makes a potential into an actual improvement: that this is the responsibility of policy makers. While more defensible, it is not an interpretation that has been taken very seriously in the literature. Empirical evidence indicates that very few economic studies investigate the equity effects of recommended policies and provide the information that would be needed for compensation. With this more defensible interpretation of Kaldor-Hicks a

[2] The conflict between envy and the social benefit of what we now call Pareto efficiency is noted in the Bible in the parable of the workers in the vineyard, St Matthew 20:1-16.

[3] One such player is reported as stating 'how can you reject a positive amount of money and prefer it to zero? They (the other player) just did not understand the game', (cited in Lehrer, 2009 p180). Restated, the player was stating that Pareto Efficiency is self evidently sensible if people treat benefits in terms of outcomes.

much greater concern would be expected with respect to the political feasibility of compensation and whether it was technically possible (as it is not, in the case of death).

The caveat referred to earlier is that social concerns including those discussed above might all be defined as sources of 'utility'. Using the revealed preference criterion, fulfilling duty, social altruism and community oriented activities may all be seen as utility maximising: why else would an individual carry out these activities? Similarly, envy and the negative response to other behaviours or situations might be viewed as determinants of preferences which are irrelevant if they do not result in changed behaviour and a manifestation of utility maximisation if they do.

This approach to definition results in a logically consistent framework which appears to overcome many of the problems discussed here. Even Sen's logical demonstration of the impossibility of a Paretian Liberal might be rationalised. If ethical principles are simply another potential source of 'utility' then the liberal facing the possibility of a non-Paretian outcome because of their (latent) preference structure (which includes a (latent) preference for a utility reducing outcome for another person) must compare the potential utility of following their own (latent) preferences with respect to others' outcomes with the potential loss of utility from violating their own liberal principles. Either the individual will act in accordance with their (latent) preferences or their liberal principles thereby indicating which generates the greatest (revealed) 'utility'.

While these definitions preserve the logical integrity of Welfare Theory they are subject to serious shortcoming especially in the context of an NHS. First, they result in a largely sterile framework. The distinctions drawn by psychological and ethical research and the complexity of human beliefs and motivations are lost. They may, of course, be reintroduced by distinguishing 'utility' which is the result of family duty from utility derived from real time fun from the utility of observing fair outcomes. But this approach does little more than add the word 'utility' in front of the previous distinctions and the problems discussed earlier re-emerge with an additional unnecessary word.

This conclusion can be avoided by arguing that in each of the cases where the term 'utility' was applied the same psychological factors were at work in a person, ie that the various distinctions observed by behavioural psychologists might all be mapped into something common. But there is no evidence for this. Rather, recent neurological research suggests the opposite. Different emotions are mediated through different neural networks and the process of reasoning which may or may not override various emotional responses varies in its effectiveness with context and the task and the reasoning processes which lead to the more 'intellectualised' sources of 'utility' are now known, with certainty, to be processed in a separate part of the brain from emotion based sources of utility (Lehrer 2009). (This evidence strongly suggests that rolling together disparate motivations under the heading 'utility' may lose distinctions which are important for understanding social and individual objectives and behaviours.)

An alternative argument is normative: that when people would choose a particular action these actions *should* be treated *as if* they reflect an underlying entity irrespective of the true psychological behaviours. Once again, however, there is no evidence that this libertarian value judgement receives sufficient support to be incorporated into economic theory.

Possibly the most damaging criticism of this concept of utility in the context of a NHS, is that the willingness to pay criterion it implies would be unacceptable as a measure of social benefit. This can be illustrated repeating an argument in Richardson and McKie (2007). As medical care is income elastic utility defined in terms of willingness to pay would rise with income. To achieve an efficient outcome in which marginal cost was equal to marginal benefit defined by this notion of utility, more would need to be spent on wealthy individuals with additional services paid by

taxation. For each additional dollar that the person was willing to pay for their own health, a completely different person – the 'taxpayer' – would have to pay an additional dollar. There appears to be no likelihood that this would correspond with social values.

The conclusion of this discussion is that two of the key assumptions in the usual theoretical justification for the current practice of CEA/CUA cannot be fulfilled in practice. Compensation cannot or will not be paid to losers, when allocative efficiency is achieved, implying distributional consequences which may be unwanted. Even if such compensation was paid and 'Pareto efficiency' achieved in terms of health outcomes or incomes, this would not necessarily fulfil social objectives. These are likely to be concerned with dimensions of fairness and sharing which are largely ignored by the theoretical framework. Decision makers currently take some account of fairness and do not follow a rigid minimum cost/QALY rule (Devlin and Parkin 2004; Harris, Hill et al. 2008). But this occurs on an ad hoc basis unaided by economic theory.

The need for explicit importance weights reflecting social values has been recognised in the literature for some time. In an early study Nord et al. (1995a) surveyed a cross section of the Australian population and found that QALY maximisation was generally rejected in favour of sharing of resources. This implied preferential treatment for selected groups who would receive lower priority if QALYs were maximised. Subsequent studies explicitly calculated age weights based upon various trade-off techniques (Lewis and Charny 1989; Busschbach, Jessing et al. 1993; Cropper, Aydede et al. 1994; Johannesson and Johansson 1997). These are criticised by Richardson (2002) for the internal inconsistency of the weights derived (which sometimes imply negative weights for some age groups) and more recently Maestad and Norheim (2009) have demonstrated that the trade-off methods used result in systematic bias as the result of comparing average weights over two periods of time when the weight's function might be increasing or decreasing at different rates with age.

Other factors influencing social preferences are reviewed by Dolan (2005) and egalitarian motivations discussed by Tsuchiya and Dolan (2009). Since different socioeconomic (SES) groups have different health status an argument may also be mounted for the weighting of programs favouring the least healthy SES groups. In one non-survey based study Williams (1997) assumed a particular willingness to pay for the achievement for health equality and, in combination of data on class differences in mortality and QALY life expectancy, constructed social indifference curves between the life expectancy of different classes. From these, importance weights were derived which could, in principle, be applied to various health programs.

Other studies have found social preferences for programs which do not leave patients in severe health states. This is not simply because greater health benefits may be obtained when a person is in a severe health state. The results suggest that this benefit is additional to the benefit of health improvement as judged by the individual's own assessment of health states before and after treatment. (For example, a health program moving a person from health states 0.1 to 0.2 (on a personal utility scale) would receive greater social support than one moving a person from 0.8 to 0.9). This implies the need for a system of weights which reflects severity independently from health gain.

Empirical studies of the effects of severity have been reviewed by Richardson et al. (2007) where evidence was also presented that the severity effect applies not simply to patients who are in a severe state before treatment but also to patients who are left in a severe state after treatment. Once again, the evidence suggests that the effect is additional to the effect of health improvement which implies the need for the use of a severity weight if these population values are to be translated into policy.

Little attention has been given to the role of costs in economic evaluation. Irrespective of the unit of benefit and how it should be modified, it appears self-evident that this unit should be obtained at the lowest possible cost. However, in 1995 Nord and Richardson found that the public did not appear to agree with this conclusion (Nord, Richardson et al. 1995a). A random sample of 551 Australians was asked whether they agreed or disagreed with a number of statements which included the following:

1. Amongst patients who are equally ill, those who can be helped at low cost should have priority over those who can be helped at high cost because this will allow more people to be helped when money is limited.

2. It is unfair to discriminate against those who happen to have high-cost illness. Priority should therefore not depend on the cost of treatment (except in cases where costs are extremely high).

Only 19 per cent supported the first and 81 per cent supported the second argument. Because of the possibility that people were giving unreflective responses, a sub-group (119) were questioned intensively in the context of an interview. Initially, the two propositions were repeated but emphasising the limited budget available for total treatments. Secondly, interviewers conducted a structured argument which challenged the individual's logic and pointed out that less health would be achieved with the second option. Thirdly, a numerical example was given to illustrate this point and, finally, the individuals were asked to allocate a budget between several options that varied in cost per life saved and clearly showed the mortality consequences of the allocation.

Generally, individuals did not change their views. Only 6 percent allocated resources to the health maximising option in the final stage.

In Spain, Abellan-Perpiñán, and Prades conducted a similar exercise asking individuals to allocate a budget between two options A and B where the cost of achieving outcome B was double that of achieving outcome A (Abellan-Perpiñán and Prades 1999). Respondents did not allocate all of the resources to option A as predicted by economic theory but allocated them in the ratio 1:2 so that the cost disadvantage of option B was completely offset. Similar results were obtained in a different context by Ubel and Lowenstein (1996a), who asked respondents to allocate resources, this time a limited number of organs, between patients with different prognoses for survival. Rather than allocating entirely to the health maximising group with a good prognosis, respondents allocated some organs to the other groups.

3 Methods and data

In the study reported here the hypothesis under investigation was that respondents discount the role of cost in order to share resources. The study was designed, however, to quantify people's deviation from a purely cost based decision algorithm and to quantify the importance of different elements of their choices.

To do this we asked people to imagine that they were allocating resources on behalf of the Australian National Health Scheme – Medicare. Initially respondents were told that they could allocate only $10,000 to one of the four patients shown in Figure 1, which would extend their life by 12, 8, 6 or 4 years respectively (resources could not be shared). Patients were all 25 years old and, without care, faced immediate death. The initial representation of Figure 1 included only the four empty blocks of years shown in bold outline and respondents were asked to click on the

block of the patient to whom they would give the resources. (Respondents were expected to click on the option for Patient 1.) This block then filled with colour to indicate that it had been selected and the subsequent block became available for the same patient (Figure 1a). Respondents were told that there was another $10,000 available and to allocate this amount to one of the available blocks for any of the four patients. The expected choices were now Patient 1 or Patient 2. This procedure continued with new blocks opening further along the screen each time a block of resources was assigned to a patient. The respondents were reminded on several occasions that they should treat each decision as if the current budget was the total budget. The exercise terminated when there was no longer any choice. The order in which blocks were clicked – resources allocated – was recorded for analysis.

Two possible outcomes are illustrated in Figures 2 and 3. The first (Figure 2) is the ordering of someone who maximised the number of life years gained. Patient 1 would always receive priority over Patient 2 over Patient 3 over Patient 4. This is very close to the orthodox economic prediction, with some minor variation caused if time discounting was taken into account. In contrast, the second outcome (Figure 3), is the ordering of an extreme egalitarian who allocated resources to the person with the shortest life expectancy and, in cases where patients had equal life expectancy, to the patient gaining most life years. This is tantamount to Rawls' 'maximin' principle of justice which would require budgetary allocation be made to the most underprivileged patient.

For each 'click' in the exercise, a respondent initially created four observations, three for the patients receiving no resources and one for the patient clicked, ie who was assigned the resources. However, after the maximum life extension of 48 years patients were removed from the choice set and a click would result in 3 then 2 observations. At some point in the exercise respondents would no longer have a choice. For example, if a person followed the efficiency maximising rule in Figure 2 there would be no choice after click 29 as only Patient 4 could receive resources. No further observations were then used in the analysis. Depending upon the order of responses this resulted in a variable number of observations per respondent.

Initially paper versions of the questionnaire were used for piloting of the exercise. These were administered to members of the Monash Centre for Health Economics and to students. Postal versions of the final questionnaire were then sent to targeted individuals including economists, politicians, civil servants and a cross section of the public. Subsequently a web-based version of the questionnaire was developed. This was first used to target selected individuals (economists and a control group).

Subsequently a commercial panel company[4] was employed to obtain volunteer web respondents. This had the advantage that, because of the numbers enrolled with the company, it was possible to purposefully target particular cohorts of the population to obtain representative respondents with respect to chosen population characteristics.

In order to assess people's comprehension of the exercise respondents were asked to write brief comments and particularly upon the reasons for their prioritising decisions. Cases were deleted when these indicated a misunderstanding of the exercise.

[4] Research Now.

Figure 1 Web-based allocation exercise

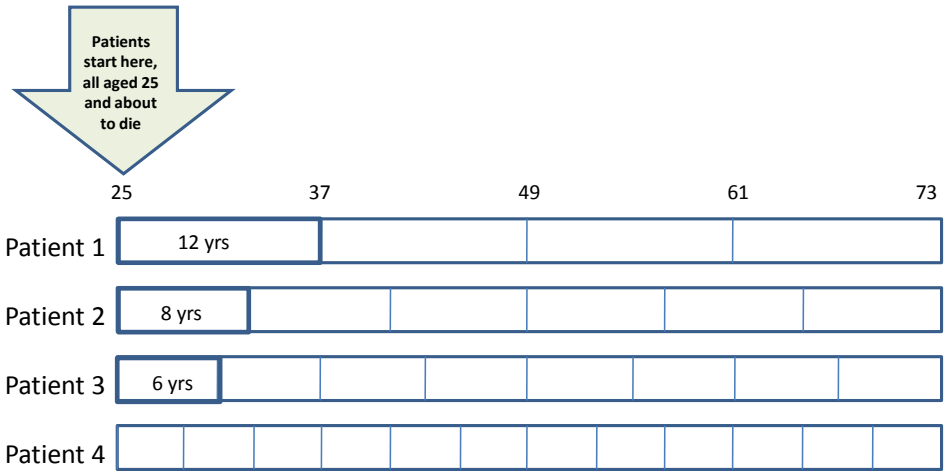


Figure 1a. The diagram represents 4 patients, all aged 25 facing immediate death. Whichever block is selected will extend that patient's life for the number of years indicated in that block.

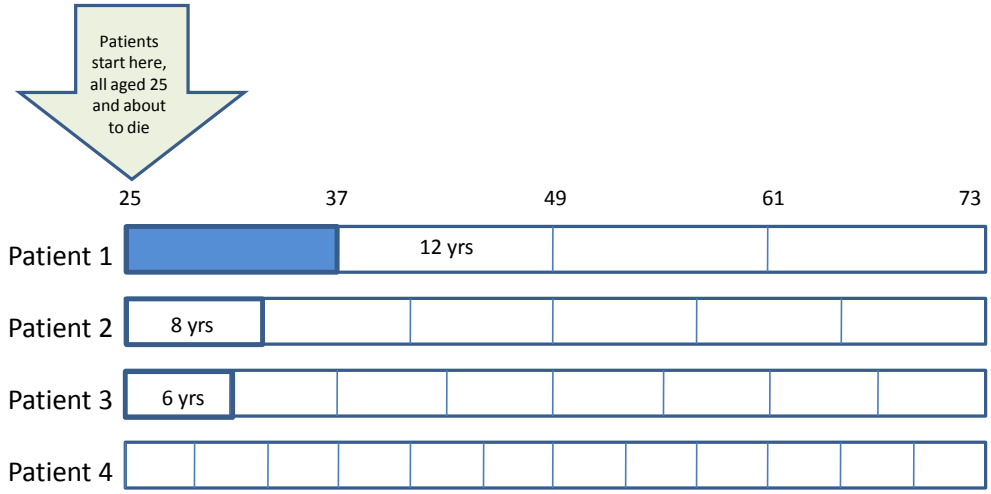


Figure 1b. When a block is clicked, it fills with colour and the next block becomes available. The patient will now live until the end of the filled block.

Figure 2 Orthodox economic ordering

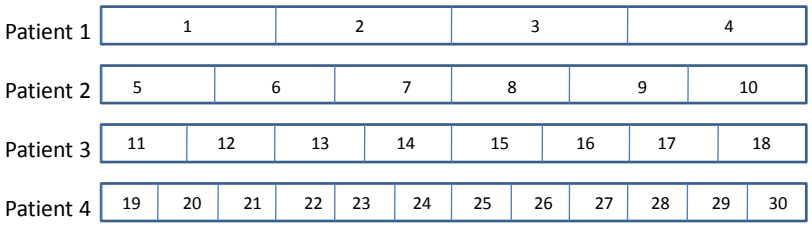
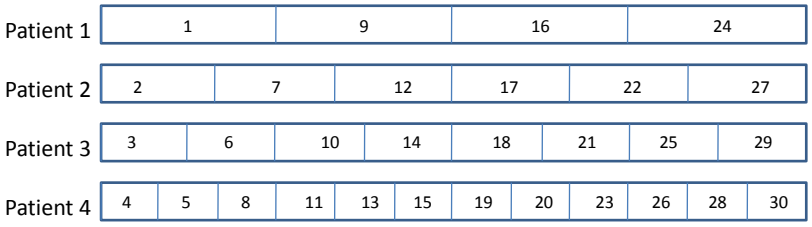


Figure 3 Extreme egalitarian ordering



4 Survey results

The web-based survey resulted in the return of 532 completed consistent sets of results. Two criteria were used for dropping data. The first was a comment indicating respondents misunderstanding of the exercise. The second was a sequence of clicks which did not start with Patient 1. Since the four individuals in the exercise are described as being identical there is a compelling reason for assigning resources to the person who receives greatest benefit. In principle, a respondent could have adopted a 'lottery' criterion and, for ethical reasons, assigned resources at random. However independent survey evidence suggests that an overwhelming majority of the public believe this to be inappropriate and we made the judgment that such results were more likely to reflect misunderstanding than unusual respondent values. Comments made by these respondents largely confirmed this judgement.

The first and most common reason for misunderstanding was that people equated more costly treatments with more serious illnesses which cause greater suffering (despite statements indicating the contrary). This error was the main stated reason for preferring high to low cost illnesses. Second, and less commonly, a small number of respondents assigned different ages to different patients. In both cases results from these respondents were not used.

The final dataset consisted of 129 responses from the initial paper survey and 403 responses from the web. The demographic and educational characteristics of the sample are reported in Table 1 along with the number of respondents which would perfectly reflect the demographic and educational characteristics of the Australian population.

Table 1 Respondents to the Sharing questionnaire and numbers representing the Australian population

Age	15-24	25-44	45-65	65+	Total	Rep*
Male						
High School	13	15	12	15	55	59
Trade	12	28	30	13	83	109
TAFE	3	9	8	3	23	29
Uni/College	18	37	38	10	103	69
Sub Total	46	89	88	41	264	266
Rep* (M)	45	95	81	45		
Female						
High School	10	14	20	25	69	72
Trade	7	19	14	3	43	59
TAFE	2	14	15	5	36	42
Uni/College	31	48	35	6	120	96
Sub Total	50	95	84	39	268	269
Rep* (F)	46	94	84	46	279	
Total Respondents	96	184	172	80	532	
Rep* Total	90	186	165	90	532	

Key:
Rep* = number achieving an exact representation of the 2006 Australian population recorded by ABS census data, Rep* (M), Rep* (F) = number achieving exact representation of males and females respectively.

Sharing indices

In the first analysis indices were constructed using equation 1 which gives the proportion of potential life years people sacrificed when allocating resources. The index was constructed by summing the number of years obtained each time a decision was made and dividing this sum by the sum of the years which would have been obtained if a person had made the most 'efficient' choice and allocated resources to the person whose life would gain most life years.

$$\text{share} = \left(\sum_{i=1}^{n} \text{Years}_{\text{i}} \text{ Allocated} \right) \Big/ \left(\sum_{i=1}^{n} \text{Possible Years}_{i} \text{ Allocated} \right) \qquad \text{Equation 1}$$

n = final click when choice was possible

Thus, for example, if an individual had followed the orthodox economic ordering in Figure 2 for the first four decisions the index numerator would have increased by 12 + 12 + 12 +12. As these were also the maximum years available for the choices the denominator of the index would have increased by 12 + 12 + 12 + 12 producing an overall index of 1.00. In contrast, the first four choices of the extreme egalitarian in Figure 3 would have added 12 + 8 + 6 + 4 (= 30) to the numerator and 12 + 12 + 12 + 12 to the denominator. If the index had been calculated at this point its value would have been 30/48 = 0.625.

By construction, the 'efficient' allocation of Figure 2 results in an index of 1.00. After the maximum possible iterations (ie where choice is still possible) the index for the egalitarian equals 0.58.

Results for different age, sex and education cohorts in Table 2 indicate that the overall average index for the full sample of 511 people is 0.625 which is closer to the score of the egalitarian than to the efficient allocation. Overall individuals make choices which sacrifice 37.5 percent of life years to achieve other goals. Their average share index differed from the egalitarian score by only 10 percent of the difference (0.42) between the egalitarian and efficiency scores. Indices were remarkably stable by demographic cohort with only ED 3 and ED 4 differing significantly. But the lack of a consistent pattern suggests that this is a coincidence. The 33 economists in the sample scored 0.655 – 63 percent greater than the amount by which the average person exceeded the egalitarian score, but not statistically different from the average owing to the large standard error, and still 0.345 away from the efficient score of 1.00.

Table 2 Sharing indices: average values*

A	Age	Mean	se	n	Age	Mean	se	n
	1	.625	.01	86	3	.624	.008	166
	2	.627	.006	179	4	.641	.011	91
		No significant differences between means						

B	Education	Mean	se	n		Mean	se	n
	1	.620	.007	121	Total	.626	.004	511
	2	.624	.009	125	Economists	.655	.019	33
	3	.600	.011	59				
	4	.640	.007	209				
		Ed 4, Ed 3 differ at .008 confidence level						

Key:

Age	1	-	< 25	Education	1	-	High school
	2	-	25 to <45		2	-	Trade/App
	3	-	45 to < 65		3	-	TAFE
	4	-	65 +		4	-	Uni/College

5. Analysis and results

The data collected were used for two analyses. In the first analysis we compared the cumulated life years which were allocated by respondents with the maximum life years which they might have allocated. In the second analysis we sought to predict choices on the basis of three hypotheses: (i)the first was that respondents would seek to maximise efficiency by allocating the budget to the patient where the greatest health gains could be obtained; (ii) that, consistent with Rawls, resources would be allocated to persons with the lowest life expectancy; and (iii) that resources would be shared between patients using some combination of rules; efficiency, Rawlsian maximin and a 'fair share of the total budget' criterion.

Hypotheses testing

For the first analysis the following variables were defined:

Ser	=	1 if a patient received a service
	=	0 otherwise
Income	=	Cumulative No. of clicks (each click represents a monetary allocation of $10,000 to one of the four patients)
Cost/LY	=	1/(life years obtained)
LE	=	life expectancy, ie cumulative life years obtained from received services
Rel LE	=	relative life expectancy: life expectancy divided by the average life expectancy of the patients in the choice
Share	=	the patient's share of the total (dollar) budget

Summary statistics for these variables are reported in Table 2. Variables were used in logit analyses to determine the probability of a person receiving resources. The dependent variable was Ser which indicated whether or not an individual received resources.

Equation 0 in Table 3 is a 'control': the result from a hypothetical individual who maximises efficiency and allocates resources in the order shown in Figure 1. Income is significant but chiefly because it standardises for the effect of reduced choice: with higher income more people reach maximum life expectancy and the likelihood of remaining patients receiving the resources is increased. Of the two key 'behavioural' variables the cost/life year is significant. Life expectancy is insignificant: that is, the results conform with expectation.

When the full dataset is used there are 41,361 observations (initially 4 per 'click', decreasing to 3 then 2 as the choices facing respondents decline). Income, cost/LY and LE were combined with different combinations of Rel LE and Share. To increase flexibility the quadratic form of the last three variables was included in some regressions.

Table 3 Logit results: All valid observations. Dependent variable: received services = 1

Equation / Variable	1		2		3		4		5	
	b	**z**	**b**	**z**	**b**	**z**	**b**	**z**	**b**	**z**
Income/10,000	1.7	11.2	1.32	(36.2)	1.33	(36.4)	1.59	(38.9)	1.50	36.3
Cost/LY	-76	6.7	-1.08	(-4.75)	-0.93	(-4.03)	-1.20	(5.33)	-1.90	-7.92
LE	0.15	0.7	-0.07	(34.8)	-0.05	13.1	-0.10	(-36.3)	-0.03	-8.46
$(LE)^2$					-.0007	-9.7	-0.03	(15.1)	-0.002	-19.6
(Rel LE)									0.01	5.2
$(Rel LE)^2$									-0.003	24.3
Constant	7.0	72	-1.06		-1.25		-0.93		-1.28	
n	52		41,361		41,361		41,361		41,361	
LR chi 2	2.13		1957		2053		2183		2986	
Log like	-37.9196		-23.375		-23.327		-23.262		22.860	
Pseudo R^2	0.02		0.040		0.042		-0.045		0.06	

Results indicate that contrary to the first hypothesis choice is not dominated by the cost per life year. Cost is consistently significant but its significance level is relatively low and the variable has relatively little explanatory power. Nevertheless it is sufficiently important to contradict the hypothesis that cost is unimportant and that people follow a maxi-min principle with respect to life expectancy.

The dominating variable in each of the results is LE, life expectancy. As expected from the fairness/sharing hypothesis an increased life expectancy corresponds with a reduced probability of receiving resources or, conversely, those in greatest need as defined by their life expectancy are most likely to receive resources. The coefficient is stable across regressions. Increasing the flexibility of the functional form by the introduction of the quadratic term $(LE)^2$ – equation 3 – improves the fit of the equation but the effect is marginal.

When life expectancy is omitted relative life expectancy becomes highly significant but this is because the two variables are highly correlated. When it is introduced in combination with life expectancy (equations 4, 5) it increases the explanatory power of the equations, but the sign is perverse and there is an offsetting increase in the importance of LE. This suggests the likelihood that the linear and quadratic models are an imperfect fit and exaggerate the importance of life expectancy somewhat, and that this effect is offset by relative life expectancy. It also suggests, perhaps tentatively, that, consistent with the Rawlsian hypothesis, it is LE per se which is the dominating concern, not relative position per se. The introduction of 'share', the variable measuring the percentage of the budget received by a person did not improve the explanatory power of the regressions. Results reported in Table 3 average observations over a wide range of situations. Initially the life expectancies of patients are the same. Subsequently they may differ very substantially. For this reason the stability of the results over the range of observations was tested by splitting the sample after 10 iterations. The first click in the first sample was deleted as observations had been deleted where person 1 was not selected and inclusion of this first result would cause bias. Summary statistics are shown in Table 4. Equations 2A, 3A, 5A in Table 4 report results from the first set of 'early' observations; equations 2B, 3B, 5B are from the second 'later' set. Because individuals were removed from the database if they did not initially select Patient 1 with their first click, the corresponding observations were also removed from the first analysis as there was no effective choice.

Results in equations 2A, 3A, 4A are broadly consistent with the previous results in terms of the sign and significance of the variables. However the magnitude of the coefficients and significance

levels changes. The coefficient on a cost/LY increases by a factor of 5. The coefficient on life expectancy is stable but its significance level falls in equation 2A.

In contrast in equations 2B, 3B, 5B the sign on the cost variable changes from negative to positive: there is a greater probability of individuals receiving resources when their cost is higher. The result is almost certainly an artefact. Throughout this dataset the difference in life expectancy is significantly greater than in the previous dataset. It has resulted from the preferential treatment of those with a low cost per life year and respondents have reacted to this by favouring those with the higher cost. In principle this effect should have been picked up from the other variables in the regression however they do not appear to have fully captured the significance of the life expectancy. The result, however, indicates that even after partial – incomplete – standardisation for LE, the pure cost effect is offset by the importance of discrepancies in LE which have developed. In contrast with the results from the 'early observations', relative life expectancy (Rel LE) exerts an independent effect. Those with a lower relative LE are significantly more likely to receive additional resources. The likely explanation for the difference is that, once again, the discrepancy in LE between individuals will have become more apparent and important in the later observations.

Comparing the two split samples it is clear that the role of efficiency changes. When life expectancies are similar it plays a role *albeit* not the dominating role in people's choice. When life expectancies are very different efficiency becomes unimportant and the full explanatory power of the analysis depends upon the fairness variables.

The relative magnitude of the effects of cost/LY and life expectancy may be illustrated by setting other variables at their mean and determining the reduction in life expectancy which would offset an increase in cost per life year from its minimum to its maximum value, that is, by a factor of 3. The results for the three datasets are:

All observations	:	reduced life expectancy = 2.5 years
Early observations	:	reduced life expectancy = 9.8 years
Late observations	:	no effect of cost

Table 4 Logit results: split sample results. Dependent variable: received services = 1

Equation / Variable	Choices 2-10						Choices 11-44					
	2A		3A		5A		2B		3B		5B	
	b	z	b	z	b	z	b	z	b	z	b	z
Income/10,000	1.20	(13.3)	1.64	(16.8)	1.63	(8.47)	1.23	(25.3)	1.81	(30.1)	1.98	(32.1)
Cost/LY	-4.71	(-10.7)	-6.30	(-13.5)	-6.35	(-13.5)	3.11	(10.96	2.95	(10.1)	1.17	(3.78)
LE	-0.08	(-19.2)	-0.15		-0.14	(-11.0)	-0.07	(-25.4)	0.06	(9.03)	0.21	(22.1)
$(LE)^2$			0.002	(11.5)	0.002	(6.39)			-0.003	(-20.8)	-0.006	(-33.0)
Rel LE					0.003	ns					-0.02	(6.03)
$(Rel\ LE)^2$					0.001	(2.9)					-0.004	(30.5)
Constant			-0.04		-0.07		1.98		4.0		-5.7	
n	15,160		15,160		15,160		22,864		22,864		22,864	
LR chi 2	438		560		569		1766		2120		3433	
Log like	-8344		-8282		-8278		-13911		-12,792		-12,136	
Pseudo R^2	0.025		0.03		0.033		0.06		0.07		0.12	

6. Discussion

The principle conclusion from the quantitative results above is that elements associated with fairness – a person's life expectancy, relative life expectancy and share of the budget – play a dominating role in the public's preferences for the allocation of resources and that, in comparison, efficiency is of secondary importance. The chief conclusion of the earlier theoretical analysis is that issues of fairness cannot be separated from issues of allocative efficiency as occurs in Welfare Theory. This implies that the overwhelming emphasis upon efficiency in applied evaluations and the neglect of distributional consequences do not reflect population values and that this cannot be excused by appeal to an economic theory which is inappropriate in this context.

The earlier theoretical discussion suggests that there is a simple reason why evaluation has evolved in this way. The assumptions of Welfare Theory allow costs and benefits to be 'disembodied': the effect, particularly of the Kaldor-Hicks criterion, is that total net benefits are an appropriate subject of analysis and a belief that Pareto optimality leads to social improvement supports the view that such analyses are relevant for improving social welfare. Without these assumptions these conclusions cannot be supported. Benefits must be regarded as 'embodied'. Allocation decisions necessarily effect distribution: decisions favour some, disadvantage others and make interpersonal judgements unavoidable in a way the 'New Welfare Economics' sought to avoid.

When equity becomes an inextricable part of the analysis the type of trade-off we find in our empirical results become consistent with broader economic theory and is unsurprising. This is illustrated in Figures 4a-4f.

Figure 4a represents the normal analysis of a demand for two goods, x and y by an individual. A budget constraint, AB, is defined by their relative prices and individual's preferences are subject to diminishing returns with respect to each of the commodities. As a result equilibrium occurs at an internal point M_1 where the volume of x and y which maximises an individual's utility depends upon a person's preferences and the relative prices.

In contrast with this, Figure 4b represents a similar budget constraint, AB. But the 'indifference curves' are no longer subject to diminishing returns. Rather, it is assumed that there is a constant benefit from the product. This is not unrealistic when the benefit is defined entirely in terms of final health outcome and service x and y are demanded by groups with sufficient individuals requiring the product (medical care or a drug) that the same benefit will be obtained over the decision space.

However, as argued in this paper, health related outcomes are not the only consequence of allocative decisions. People have distributive preferences. Figure 4c represents extreme egalitarianism with respect to life years allocated to two groups. In this figure there is no social benefit if life years are not allocated so that the society moves along the expansion path ABCD. This extreme case is improbable although not impossible. If one group was favoured over another the social response – dislike of inequality, etc – might offset the benefit to the group which is advantaged.

The more probable scenario is depicted in Figure 4d. In this there is a trade-off between egalitarianism and efficiency. Commencing at an egalitarian outcome, P, a reduction in the life

years allocated to group 1 by PA would have to be offset by a more than equal increase in the life years allocated to group 2, for example AC. Similarly a reduction to group 1 of PD would need a disproportionate increase in the life years allocated to group 2, DF. With increasing social disutility associated with a departure from egalitarianism the same result is achieved socially as for the individual with diminishing marginal utility. The social indifference curve becomes convex from below and the optimal point for the society becomes M_1, a trade-off which is analogous to the result in Figure 4a.

This last result is a significant departure from orthodox efficiency analysis. Health – life years – have not been maximised as this would be achieved by moving to point A. Rather, because of the fairness effect the population has traded life years for fairness, the extent of the trade-off depending upon the relative importance of these two effects. Figure 4f illustrates the effect of the change in the relative importance of the effects. Again, the result is analogous to normal consumer theory. As the price of life years to the second group declines, the budget line rotates from AB to AC and more services will be allocated to the second group. Efficiency matters but its effects are mitigated and depend upon the respective strength of preference for efficiency and fairness.

These results have significant implications for the practice of economic evaluation. They imply that the population would be strongly opposed to the use of simple QALY league tables for the maximisation of health benefits defined by QALYs or for any other league table constructed on the basis of health outcomes alone. In principle, the results are consistent with the use of formula to determine a cost threshold beyond which services would not be funded. However the threshold would be determined by a formula which consisted of more than a single cost per QALY.

In the simple community of four used in the present study the logit results could be used to construct league tables if decision makers accepted the probability of social choice of a project as the appropriate criterion. In the illustrative table the probability has been set at 50 percent. However with external budgetary constraints this could be set at a lower or higher level.

An interesting application of the analysis is that the result of any such study would depend upon the definition of the groups where a substitution of efficiency for fairness was permissible. That is, while the probability of social acceptance might be considered to be an important part of the decision process, there is no technical reason why this should be the only component. Social values – or ethical judgements generally – cannot be determined by empirical evidence (the naturalistic fallacy). Thus, external constraints might be placed upon the groups where trade-offs were permissible and a separate budget determined for each of these. Alternatively, importance weights might be applied in addition to the results of the formula. In Australia a separate budget or preferential treatment might be given to the Aboriginal and Torres Strait Islander community. More generally weights might be applied to social classes and different budgets to different diagnostic groups if it was considered that the different illnesses represented different social problems (for example cancer versus accidents versus psychological problems).

While this is, of course, presently possible the results here highlight the inextricable nexus between allocation and distributive fairness. The dimensions of distributive fairness have been informed by remarkably little empirical evidence. The desired boundaries cannot be defined and the relationship between the desired boundaries and appropriate policy has also received relatively little discussion in the mainstream literature.

Figure 4a Orthodox demand theory 'Efficiency perspective'

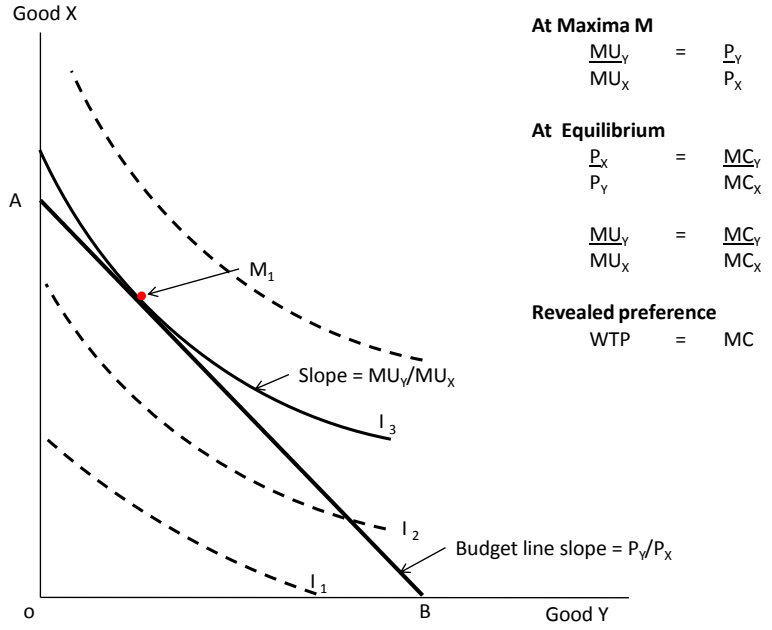


Figure 4b Costs, benefits CEA (efficiency) $MC_Y < MC_X$ → corner solution

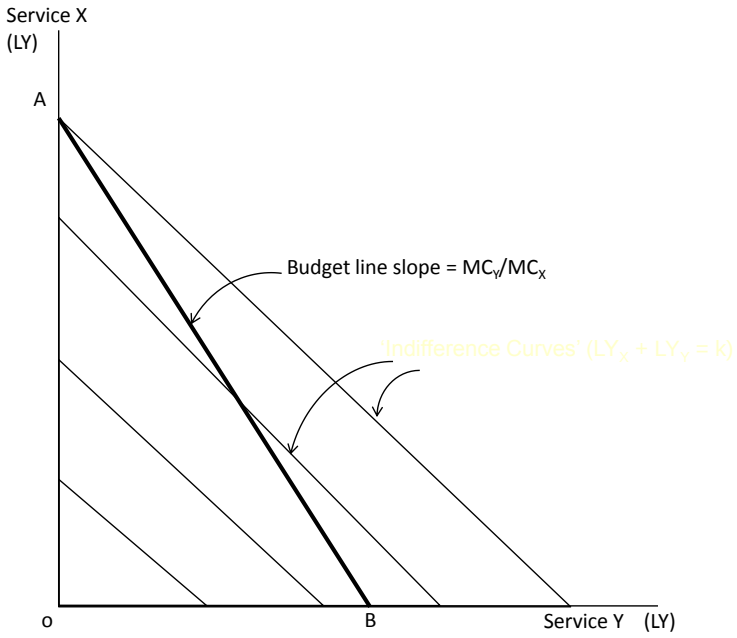


Figure 4c Indifference 'curve' perfect egalitarianism IC(PE) fairness perspective

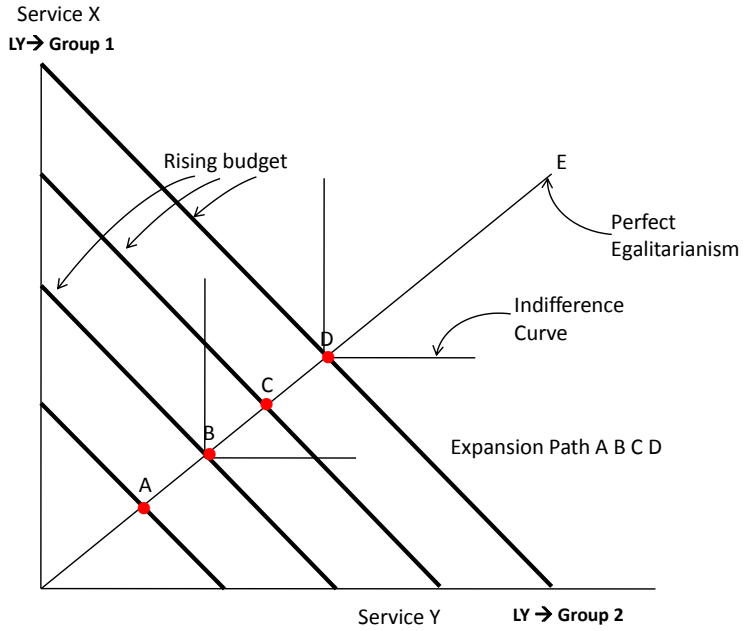


Figure 4d Indifference curve with a fairness-outcome trade-off

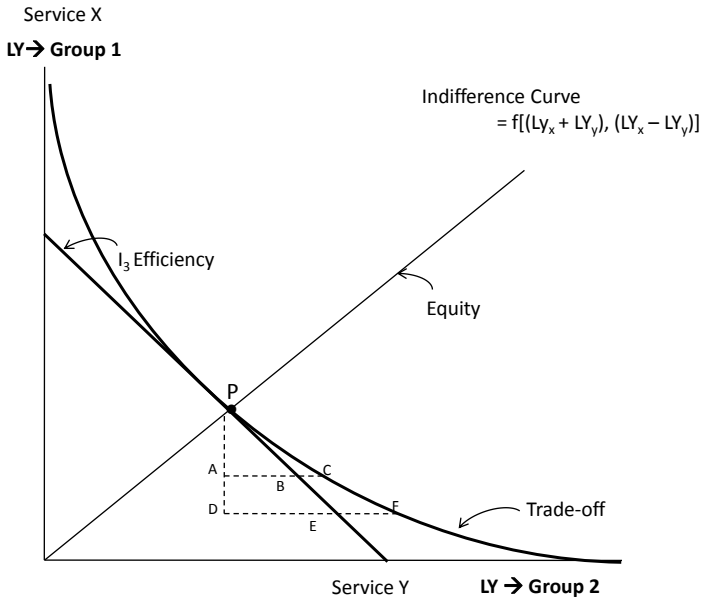


Figure 4e SW = f(fairness efficiency) $MC_Y > MC_X$

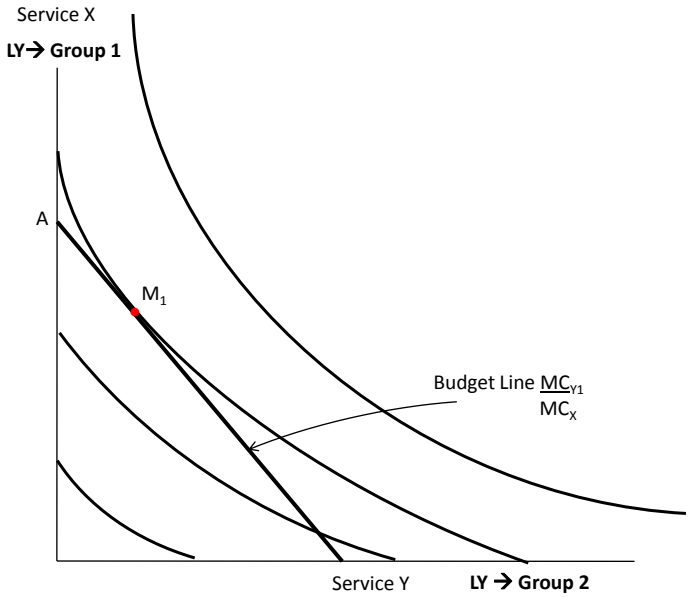
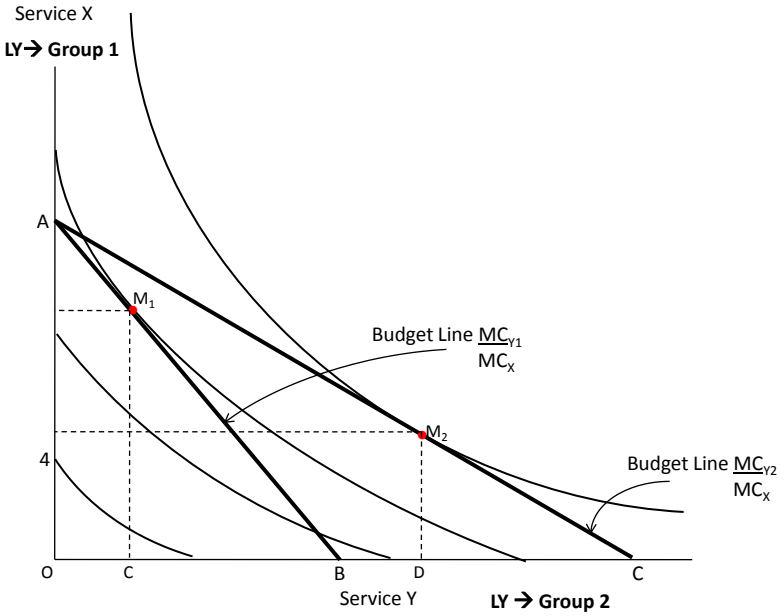


Figure 4f SW=f(fairness efficiency) fixed budget $MC_{Y1} > MC_X > MC_{Y2}$



7. Conclusions

The present study has argued that the efficiency focus of evaluation theory and practice in the health sector cannot be justified. Benefits cannot be 'disembodied' from individuals and criteria which rely upon cost per QALY data do not reflect social values. Rather they have been imposed as a result of Welfare Theory which relies upon key assumptions which cannot apply in the health sector.

Present policy takes account of some of these considerations but in an ad hoc fashion. Evidence suggests that policy makers are more likely to fund high cost per QALY services when the probability of death is higher or there are no substitute services. This is despite an economic orthodoxy which overwhelmingly emphasises pure efficiency defined in terms of cost per QALY and a practice which typically provides little information on the fairness elements of the decision. Our study suggests the need for a reconsideration of the importance of the various dimensions of fairness both in theory and practice and that, especially in the context of a NHS fairness and its measurement should be a central concern of the evaluation framework and of decision making.

References

Abellan-Perpiñán, J.-M. and J.-L. P. Prades (1999). "Health State After Treatment: A Reason for Discrimination?" Health Economics **8**(8): 701-707.

Beinhocker, E. (2007). The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics. Milsons Point, Random House.

Busschbach, J., D. Jessing, et al. (1993). "The utility of health at different stages of life: A quantitative approach." Social Science & Medicine **37**(2): 153-158.

Cropper, M., S. Aydede, et al. (1994). "Preferences for life saving programs: How the public discounts time and age." Journal of Risk and Uncertainty **8**(3): 243-265.

Devlin, N. and D. Parkin (2004). "Does NICE have a cost-effective threshold and what other factors influence its decisions? A binary choice analysis." Health Economics **13**: 437-452.

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

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

Guth, W., Schmittberger, et al. (1982). "An experimental analysis of ultimatum bargaining." Journal of Economic Behavior and Organization **3**(4): 367-388.

Harris, A., S. Hill, et al. (2008). "The role of value for money in public insurance coverage decisions for drugs in Australia: A retrospective analysis 1994-2004." Medical Decision Making **28**: 713-722.

Johannesson, M. and P.-O. Johansson (1997). "Is the valuation of a QALY gained independent of age? Some empirical evidence." Journal of Health Economics **16**: 589-599.

Lehrer, J. (2009). The Decisive Moment: How the Brain makes up its Mind. Edinburgh Canongate Books.

Lewis, P. and M. Charny (1989). "Which of two individuals do you treat when only their ages are diferent and you can't treat them both?" Journal of Medical Ethics **15**(1): 28-32.

Maestad, O. and O. Norheim (2009). "Eliciting people's preferences for the distribution of health: A procedure for a more precise estimation of distributional weights." Journal of Health Economics **28**(3): 570-577.

Nord, E., J. Richardson, et al. (1995a). "Maximising health benefits versus egalitarianism: An Australian survey of health issues." Social Science and Medicine **41**: 1429-1437.

Rice, T. (1998). The Economics of Health Reconsidered. Chicago, Illinois, Health Administration Press.

Richardson, J. (2002). Age weighting and discounting: Technical imperative versus social choice. Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications. C. J. Murray, J. Salomon, C. D. Mathers and A. D. Lopez. Geneva, World Health Organization**:** 663-676.

Richardson, J. and J. McKie (2007). "Economic evaluation of services for a national health scheme: The case for a fairness-based framework." Journal of Health Economics **26**(4): 785-799.

Richardson, J., J. McKie, et al. (2007). Severity as an Independent Determinant of the Social Value of a Health Service, Research Paper 18. Melbourne, Centre for Health Economics, Monash University

Sen, A. (1997). The impossibility of a Paretian Liberal. Choice Welfare and Measurement. A. Sen. Cambridge, Harvard University Press**:** 285-309.

Tsuchiya, A. and P. Dolan (2009). "Equality of what in health? Distinguishing between outcome egalitarianism and gain egalitarianism." Health Economics **18**: 147-159.

Ubel, P. A. and G. Loewenstein (1996a). "Distributing scarce livers: The moral reasoning of the general public." Social Science and Medicine **42**(7): 1049-1055.

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

Appendix A Summary Statistics

Table A1 Summary statistics

Full dataset	n	Mean	sd	Min	Max
Income ($10,000)	41361	13.2	7.7	1	30
Cost/LY	41361	0.162	0.06	0.083	0.25
LE	41361				
Rel LE	41361	1.76	8.1	-29.5	33.0
Share		0.25	0.14	0	1

Observations 2-10

Full dataset	n	Mean	sd	Min	Max
Income ($10,000)	15160	5.980937	2.581114	2	10
Cost/LY	15160	0.1569234	0.0615102	0.08333	0.25
LE	15160	9.898285	8.808445	0	40
Rel LE	15160	-0.051715	7.500027	-30	20.5
Share	15160	0.2474793	0.1586775	0	1

Observations 11-30

Full dataset	n	Mean	sd	Min	Max
Income ($10,000)	22864	19.20307	4.785273	12	30
Cost/LY	22864	0.1659869	0.060518	0.083333	0.25
LE	22864	29.19787	9.635497	0	44
Rel LE	22864	-3.131692	8.581076	-33	29.5
Share	22864	0.254863	0.0876288	0	0.75

Table A2 Summary statistics: correlation table n = 41,361

	Cost/LY	LE	Rel LE	Share
Cost/LY	1.00			
LE	-0.24	1.00		
Rel LE	-0.39	0.63	1.00	
Share	0.15	-0.58	-0.18	1.00

Observations 2-10

	Cost/LY	LE	Rel LE	Share
Cost/LY	1.0000			
LE	-0.6067	1.0000		
Rel LE	0.7063	-0.8495	1.0000	
Share	-0.4005	0.4980	-0.5729	1.0000

Observations 11-30

	Cost/LY	LE	Rel LE	Share
Cost/LY	1.0000			
LE	-0.3291	1.0000		
Rel LE	0.2285	-0.6734	1.0000	
Share	0.6015	0.2499	-0.2583	1.0000