Some methodological issues in economic evaluation in health care

by

Jan Abel Olsen

Acknowledgements

Content and list of papers

Introduction

Paper 1: Olsen, J.A. *Verdsetting av menneskeliv - eller hvordan vi kan sette pris på hverandre*, Norsk Økonomisk Tidsskrift 1990 (1): 1-27.

Paper 2: Mooney, G. and J.A. Olsen, *QALYs: Where Next?*, in McGuire, A., P. Fenn and K. Mayhew, Providing Health Care: The Economics of Alternative Systems of Finance and Delivery, Oxford University Press, Oxford 1991.

Paper 3: Olsen, J.A., *Production gains: Should they count in health care evaluations?*, Scottish Journal of Political Economy, 1994 (1): 69-84.

Paper 4: Olsen, J.A., *On what basis should health be discounted?*, Journal of Health Economics, 1993: 39-53.

Paper 5: Olsen, J.A., *Time Preferences for Health Gains: An Empirical Investigation*, Health Economics, 1993: 257-265.

Paper 6: Olsen, J.A., *Persons vs years: Two ways of eliciting implicit weights*, Health Economics, 1994: 39-46.

Paper 7: Olsen, J.A. and C. Donaldson, *Willingness to pay for public sector health care programmes*, submitted to a journal.

Paper 8: Olsen, J.A., *But* health *can still be a necessity ...* Journal of Health Economics 1993: 187-191.

Appendices:

Appendix 1: An English translation of Paper 1: *Valuation of human life.*

Appendix 2: The questionnaire used for Paper 5.

Appendix 3: The questionnaire used for Paper 6.

Appendix 4: The questionnaire used for Paper 7.

Introduction

The aim of this thesis is to contribute to the improvement in the methodologies of economic evaluation in health care.

The "dismal science" of economics has always stressed that resources are scarce. In that respect, the health sector is not exceptional. Although health care expenditures have increased more rapidly than GDP in most OECD countries (OECD 1990), the awareness of scarce resources depends not only on the absolute level of available resources, but also on the extent to which known, medically effective treatments can*not* be realised. Hence, *the wider the gap* between available health care resources and the available treatment opportunities, the stronger becomes the feeling of scarcity. In this environment, attention turns to economic evaluation. Not surprisingly for a sector with strong emphasis on humanitarian concerns and ethics, the economic concept of efficiency is often met with suspicion. Apparently, for health economists to argue for our noble intentions is therefore a topic of itself. In his very eloquent style, Alan Williams (1987) suggests that "health economics is *The Cheerful Face of the Dismal Science*", because it is "concerned to improve the quality of people's lives to the maximum feasible extent".

The term *economic evaluation* is now fairly well established in the health economics literature (see e.g. Drummond et al 1987, Drummond et al 1993). It includes essentially any type of analysis which compares costs and consequences. Hence, cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility anlysis (CUA) would all come under the umbrella term of economic evaluation. *Economic appraisal* appears to be a somewhat synonymous term (see e.g. Torrance 1986, Blades et al 1987).

An increasing number of economic evaluations have been undertaken. In his first volume of *Studies in economic appraisal in health care*, Drummond reviewed 101 studies published before 1980 (Drummond 1981). The second volume covering the next five year period contained a review of 100 published studies. This period had seen the quantity of

published work increase "dramatically" (Drummond et al 1986). Gerard (1991) surveyed the CUA studies published in the English language over the period 1980-90 and counted 51 studies. She estimated that at present CUA studies would constitute less than ten per cent of all published economic evaluations in health care. Bergmo (1993) surveyed published economic evaluations by Norwegian authors since 1980. Of the 42 publications which purported to be economic evaluations, only 17 appear to satisfy the requirements needed for being labelled CBA, CEA or CUA.

In his introduction to the selected readings in *The Economics of Health*, Culyer (1991) held that in the field of economic evaluations "[t]he variance in quality is *extremely* high". Certainly, this characteristic also seems to be appropriate for those published evaluations with a Norwegian origin (Bergmo 1993). In their recent paper *Standardising methodologies for economic evaluation in health care*, some of the worlds' leading authors in the field (Drummond et al 1993) suggested that while the quality of economic evaluations has improved, there are still unresolved methodological issues. Six issues were listed on which they held that there is still low level of agreement about what is the correct procedure to follow. Four of these are, to varying extents, dealt with in this thesis. They are:

1. "Inclusion of indirect costs and benefits."
2. "Choice of discount rate for health benefits."
3. "Method of measuring the utilities of health states."
4. "Incorporation of considerations of equity in economic evaluations."

The other two issues which this thesis does *not* deal with are:
"Inclusion of health care costs in added years of life."
"Inclusion of intersectoral consequences of health care programmes."

In addition to the issues listed by Drummond et al (1993), I would add one where disagreements among health economists have been disclosed. That is: 5. *Accounting for the nature of the health care programme which produces the health outcome*, which includes the topic of 'process utility', and the idea of programme specific weighting of QALYs. This is the fifth issue covered by this thesis.

In this introductory chapter, I shall attempt to place the contributions of this thesis in relation to these unresolved methodological problems. The five methodological issues will be discussed consecutively, each dicussion focusing on the main findings in the thesis and on their specific relevance to health policy making.

Which paper deals with which issues
Issue 1 is the topic of Paper 3 but it is also raised in Paper 1, which is a review of the value of life literature. *Issue 2* is the one which is dealt with most extensively and is covered in Papers 4, 5 and 6. There is an immense literature on *Issue 3*, reflecting the wide research interest not only among health economists but also in other disciplines. The contributions to the issue in this thesis (see Paper 2) are admittedly small. *Issue 4* is also a big area. While none of the papers is exclusively devoted to this issue, different aspects of equity are discussed in Papers 2, 5, 6 and 7. *Issue 5* is dealt with in Papers 2 and 7.

Paper 8 does not focus on a methodological issue in health care evaluation, but on the labelling of health care as "a luxury". Paper 8 may appear as something of a detour in this thesis. It has a link, though, through the connotation attached to the wider preferences for health and health care, which is a general topic covered in some of the papers, particularly Paper 7. Furthermore it links to the ethical view underlying the objective of allocating health care according to need.

1: Methodological issues

1.1 Inclusion of indirect costs and benefits

Paper 1 highlights some of the ethical implications of a form of priority setting based on the inclusion of indirect costs, i.e. one that would favour the treatment of those patient groups with the highest expected future earnings, at the expense of elderly and patient groups outside the workforce. The paper prompted a response from Magnussen (1990), which initiated further work on this issue (Olsen 1990, Nygaard and Olsen 1992, Paper 3).

4

Quite a number of economic evaluations, which pretend to be CBAs, use the present value of future earnings on the benefit side, although such a measure of benefit is not consistent with the welfare economics theory on which the CBA-approach is based. Analysts may, however, have two good reasons for using this benefit measure: first, it represents the easiest and cheapest way of getting a monetary estimate of benefits, and second, they may have vested interests in boosting the benefit side in order to conclude that the particular intervention would yield net economic benefits to society, thereby enabling the authors to recommend the implementation of the particular programme.[1]

In the context of CEA/CUA-studies which do not have a monetary benefit side but which normally use health outcome as the denominator, the issue is whether "indirect benefits" should be subtracted from the direct health care costs to estimate net social costs, before the cost per unit of outcome is calculated. On this issue, I would first propose that economists should stop using the terminology of "direct" and "indirect" costs and benefits in this context. This is because such terminology bears perhaps too strong a resemblance to the distinction between variable and fixed costs, and to the distinction between 'internal costs' and 'external costs' (i.e. costs generated by own activity but imposed on others). Further, it implicitly suggests an all-inclusive split: what is not direct must be indirect. Rather we should use "health care costs" when referring to costs arising from the use of health care resources, and "production gains" when referring to the increased production that would follow from allowing sick people to return to work.

Drummond et al (1993) observe that the views which analysts have on inclusion of production gains depend on the prevailing values in society. In some countries they are excluded because of the bias in favour of the economically productive groups, while in other countries, this may not be such a concern. Williams (1992) maintains that British health economists have got themselves out of the dilemma through the use of a plain economic argument: due to the high unemployment, sick employees can be replaced by fit unemployed people, leaving the production level "virtually unchanged". Gerard and Mooney (1993), in the context of QALY-league tables, appear to take the same view as in

---

[1] Professional politeness prohibits me from giving references.

the seminal paper by Weinstein and Stason (1977), that production gains should not be included. The only costs to be included are "the components that add to or subtract from the resources available for health care" (Weinstein and Stason 1977). In other words, production gains do not represent opportunity costs for the health service and should consequently be left out of the analysis. The model presented in Paper 3 attempts to bring this issue some steps further forward.

1.1.1 The main findings on Issue 1

The model makes a split between the production relationships and the objectives, which enables value judgements to be analysed separately from the issue of increased production. When health is to be maximized, which is the case in CEA/CUA, the model shows that *not* all production gains are welfare improving. It is the increased health care output which might arise from allowing productive people to return to work which is to be subtracted from the health care costs of curing productive people. In the context of Pareto-improvements, the condition for welfare gains is that the increased health care which arises, outweighs the health care resources needed to cure them, i.e. as long as productive people "pay their way" in terms of health care. However, when welfare is a function not only of health but of consumption goods as well, then in the context of Pareto-improvements, the increased production of consumption goods can be used to compensate the non-productive for letting the productive be prioritised in terms of health care.

1.1.2 The policy relevance of Issue 1

The view that production gains should count to reflect the 'economic' gains to society from health care, is frequently expressed, particularly by non-economists. It should be noted that while this argument is put forward in a macro-context, the implication of it at the hospital admission level is rarely used in practice. That implication would be always to prioritise those patients who will return to work. However, the waiting list criteria proposed by the Norwegian Surgeons' Association would prioritise "Illnesses which would make one person who would otherwise be capable of work, incapable of doing so" (TNL 1991). Is it for surgeons to promulgate such non-medical recommendations?

6

It appears to be a rather common belief that health care personnel should be treated first. A Norwegian psychiatrist, in a newspaper interview, indicated that he gave priority to people who "were to help others - doctors, nurses, social workers and professors" (Nordlys 2.11.1990). Intuitively health care personnel might be most capable of "paying their way" in terms of health care. However, if the amount of health care is constrained by resources transferred to the health sector as a fraction of total production (as suggested by the model in Paper 3), then it is the value of the production rather than what is produced that matters. In other words, when the constraint lies in the level of funds available to employ those who are to provide cure, the argument of first curing health care personnel breaks down.

A choice which seems rational to a decision maker who ignores intersectoral externalities will seldom lead to social efficiency. From the perspective of the "public purse", when the costs of cure are less than the sickness benefits paid during waiting, it may therefore appear to be a step in the right direction if the "Sickness Fund" purchase treatment for those who will return to work. However, if this means that a temporarily employed "substitute" is made unemployed again, the Sickness Fund might be charged a corresponding amount by the institution which pays unemployment benefits. The crucial matter is thus *not* whether sickness benefits are being paid or not, but rather if, and by how much, the production level changes. When discussing this argument, it is crucial to distinguish between real resource costs and transfer payments.

1.2 Choice of discount rate for health benefits

The issue of the discount rate has two separate aspects; first, whether health benefits should be discounted at a different rate from costs; and second, whether to apply a different discount rate in health care programmes from that used in other public sectors. Paper 4 analyses these two issues. It holds that, when the matter at hand is the ranking of health care programmes within a constrained current budget, one could discount health at a different rate from that used for costs. However, it also shows that the lessons from the general economics literature on discounting suggest that one should apply the same rate on both sides and that this rate should be based on individuals' intertemporal preferences for

health. If the intertemporal preferences with respect to health are different from those for consumption goods, it follows that the discount rate to be applied in health care programmes would differ from the rate used elsewhere.

The sovereign consumer argument suggests that these comparisons should reflect citizens' preferences. There are at least five different ways of eliciting individuals' implied time preference rates for health. The first is as in the research agenda set out in Paper 4 and later used in the empirical study referred to in Paper 5. This involves asking people to make intertemporal *social* choices between health care programmes whose only differences are the number of units of health benefit and the timing of their occurrence. The second is based on private intertemporal choices between the delay of onset of a temporary inferior health state and its duration. Own life stage considerations however may 'contaminate' the implied time preference rates.

A third approach, which is adopted in Paper 6, is to elicit the rate from choices between hypothetical *streams* of health. From the observation that people do not trade off life years and life quality at a constant proportion independent of length of time in a health state, the time-trade off approach of Paper 6 suggests how private time preference rates for own health can be estimated. While this approach involves choices between certain alternatives, the fourth approach is to use the standard gamble technique and frame choices between different probabilities and durations. However, this approach may reflect risk attitudes rather than time preferences, or perhaps both.

A fifth approach, which to my knowledge has not yet been tried empirically, is to use willingness to pay (WTP). As with intertemporal social choice referred to under the first approach above, it would be possible to describe two health care programmes both of which are to be funded in the same time period and which are identical in all respects except for the point in time at which the health gains occur. Assuming constant marginal utility of income in the actual payment interval, the implicit time preference rate can be estimated from the two stated WTP figures. I would hope to see this approach being tested.

8

1.2.1 The main findings on Issue 2

The empirical studies reported in Paper 5 support previous findings that the annual discount rate declines with length of time. It is therefore necessary to be cautious in using a rate elicited from a short time horizon for projects with a long time horizon (and vice versa). New findings are that time preferences are stronger with regard to health improvements than for life savings, and that health planners have lower time preferences than have a random sample of the population.

According to Drummond et al (1993): "The choice of discount rate is a societal value judgment which is likely to vary from country to country". This seems very plausible and supports the need for setting up a "standardised methodology" in different countries so that international comparisons can be made. In the meantime, however, I would conjecture that there is evidence from Scotland, the US and Norway (Cairns 1992, Cropper et al 1992, Paper 5) that using 10% is more in line with individuals' intertemporal preferences for health than is 5%, which appears to be the current international reference rate.

1.2.2 The policy relevance of Issue 2

The Norwegian National Health Plan (St.meld. nr. 41, 1987-88) suggested that more resources should be spent on prevention and argued that:

"The marginal effects of further increases of resources allocated to the curative sector of the health service are probably small in terms of improved health. On the other hand, it is very likely that efficient preventive programmes will have considerable impact on the population's health status, and hence on the necessary resources for the health sector".

There are three important points to be taken from this quotation. First, it seems to reflect a belief that prevention reduces total health care expenditures. There is, however, no evidence of any kind that increased prevention does reduce the total health care costs (Evans 1993). The merit of prevention is therefore to be judged by the health gains produced. Second, in order to substantiate the above suggestion, one has to be able to make comparisons between "improved health" from cure and "considerable impact on the population's health status" from prevention.

Third, when making comparisons between programmes which yield benefits at different points in time, such comparisons are by definition intertemporal which calls for the use of a discount rate to adjust for the different timing. Paper 5 shows that the median respondent in a random sample of the population discounted health improvements occurring in 20 years time at a rate of 10.2%, which was implied from the valuation of present health gains as seven times higher than the gains occurring in 20 years. Returning to the above quotation from the National Health Plan, this means that if prevention is to be considered more efficient, ceteris paribus, the health gains produced in 20 years will have to be at least seven times as large as the health gains produced now.

1.3 Method of measuring the utilities of health states

There are essentially two types of health gains: *life quantity* in terms of extending lives and *life quality* in terms of improving health states. Multiplying the health status increment [0-1] by its duration in years, gives the number of Quality Adjusted Life Years (QALYs) (see e.g. Torrance 1986, Williams 1985). Although there is wide agreement in the health economics literature on the need for a unidimensional yardstick, there appears to be less agreement on its *name*.[2]

Utility, value or simply health?
Cost-utility analysis (CUA) implicitly assumes that the QALY denominator reflects a measure of *utility*. However, it has been questioned whether the term utility is the most appropriate one to use. First, individual differences in relative intensities of preferences are ignored because of the assignment of the same finite end points, 0 and 1, on the cardinal health scale. Paper 2 consequently holds that QALYs are at best a 'quasi-utilitarian' welfare measure. Second, among the measurement techniques that are currently being used, it is only the standard gamble that is explicitly connected to the notion of utility, and then only with utility as an *index of choice* rather than as *welfare* (Culyer 1989).

---

[2] "Kjært barn har mange navn".

There are various interpretations of the entity which measures health on the scale [0-1]: utility, value, health index, quality of life index or quality of well-being. Some authors' disquiet over the term 'utility' seem to reflect an aversion to the concept of cardinal utility or at least to its measurement. However, in a seminal paper on this issue, Torrance (1986) takes a pragmatic view: "Utilities are cardinal values that are assigned to each health state on a scale [0-1] that is established ...", and "The utility values reflect the quality of the health states ...". Further; "A cardinal preference is sometimes referred to as 'value' and often referred to as 'utility'. ... at this stage of the development of applications in health the distinction has not proven to be very important." Hence, which of the terms - utility, value or health status index - one prefers to use may be basically semantic.[3]

Alan Williams and the EuroQol® group (The EuroQol Group 1990) use the term 'health state value' rather than utility. Tony Culyer refers to QALYs as a "pseudo-quantitative measure" of health gain, while Wagstaff (1991) takes a similar view: "The policy objective underlying the QALY literature is the maximization of the community's health. An individual's health is measured in terms of QALYs and the community's health is measured as the sum of QALYs". Gafni and Birch (1993) seem to distinguish between what they imply from Wagstaff as "health *per se*", as opposed to "*value* of health" from the EuroQol® group. I have problems in seeing that such a distinction can be made, simply because I cannot see how the different dimensions of a health state can be reduced to a uni-dimensional "objective" health ("per se") index which has no relation whatsoever to individuals' preferences for, or valuation of, the particular health state.

From my experience in teaching, I find that people often have much less aversion to the idea of quantifying health than they have to quantifying "life-quality". People seem to be more willing to accept the term "Health Adjusted Life Years" (HALYs) (Nord 1989) than QALYs. This suggests a shift in focus from the unquantifiable notion of life quality and the complications surrounding interpersonal comparisons of utility, to a focus on health as "characteristics of people" (see Culyer 1990).

---

[3] Certainly, not only economists may have an ambiguous interpretation of the concept of utility: "'Nytte meg her og nytte meg der, noen *er* - og noen er *ikke*', sa Klatremus, hva han nå kunne mene med det" (Egner 1988, a classic Norwegian childrens' book).

The importance of the choice context

Moving away from semantics, I believe that the important issue here is not the name of the scale but *the focus of the choice context* from which the indices are being inferred. I would therefore conjecture that Nord's (1992) distinction between *individual value* vs *social value* represents a fruitful baseline. If the choice context is presented as one of "imagine *yourself* in that health state" (which is the case for category rating (CR), standard gamble (SG) and time-trade-off (TTO)), the inferred value is an individual one. However, if the choice context is one of comparing different patient groups (which is the case for person-trade-off (PTO)), the inferred value is a social one. The transformation of an individual value scale to a social value scale (Nord 1993) visualises the differences. Providing a certain health improvement, of the same *individual* value, at the lower end of the scale has a higher social value than at the upper end of the scale. In other words, there appears to be a diminishing marginal social value of individually valued health increments.

Given that the health state values differ depending on the choice context, the question immediately arises as to whether a private or a social choice context is the appropriate one to use. These choice contexts place the respondent in two different roles. In TTO the respondent is put in the hypothetical role of the patient choosing between two streams of health, from which an index is inferred. PTO, on the other hand, places the respondent in the hypothetical role of a social planner choosing between two patient groups, to neither of which she herself belongs. While the former focuses on private preferences, the latter focuses on social preferences.

However, on its own, neither of these two choice contexts is sufficient. Given that an individual has both selfish preferences for own health, and social preferences for her fellow citizens' health (see the extended utility function in Evans and Wolfson 1980, or Evans 1984), the two choice contexts will focus on only one of these two arguments. In a paper written with Derek Clark (Clark and Olsen 1994), we have argued for a split preferences model which reflects idividuals' *ex ante* preferences for the allocation of health care resources. I believe that an adaptation of this model would suggest that the correct index is the one which reflects ex ante preferences, which again would be some weighted average of the individual value and the social value.

12

1.3.1 The main findings on Issue 3

In the context of interpreting QALYs as a utility measure in the denominator of cost-utility analyses, Paper 6 makes the distinction between "unweighted" and "weighted" QALYs. It is the health states without, and with, treatment which have undergone utility elicitation. Preferences related to the duration of the health improvement have rarely been elicited. Paper 6 focuses on three separate concerns which tend to change the value of additional years with improved health: time preferences; diminishing marginal utility of increased years; and, in a social value context, equity preferences. Further, two ways of weighting QALYs so that they correspond better with individuals' valuations are presented. Attention is drawn to the distinction between private vs social framing of the choice contexts.

A brief discussion of the different techniques for measuring health states is provided in Paper 2, which suggests that among the alternative methods, time-trade-off (TTO) poses the right QALY question. Further, the paper draws attention to the fact that QALYs ignore individual differences in relative intensities of utility from health. This is partly because of the problem of making interpersonal comparisons of utility, but also the equity aspect of having each person's QALY of equal value.

However, normalising the health value scale to [0-1] also implies that programme specific variations in the value of QALYs are being ignored. In other words, QALYs cannot take account of the evidence which suggests that health outcome produced by rescue-type programmes is valued higher than the same outcome produced by other programmes. This issue is elaborated on in section 1.5 below.

1.3.2 The policy relevance of Issue 3

In most public health care systems the overall objectives deal with efficiency and equity. Efficiency is often defined as some form of health maximization. The principal objective of the Norwegian health service is "to combat illness and promote health" (NOU 1987:23), or "to improve the health status of the population" (St.meld. nr. 41 1987-88). These

formulations bear a strong resemblance to the economists' notion of health maximization. In order to assess whether a life extending treatment yields more health gains than a health improving alternative, the two sorts of gains *must* be made commensurable. As such, QALYs represent the unit for measuring health, and hence are an aid to decision makers in their attempts to maximize health. However, the Norwegian Commission on Priority Setting (NOU 1987:23) revealed a rather vague attitude to QALYs.

In addition to *efficiency* and *equity*, the stated aims of the Norwegian health service includes *patient autonomy and influence* (Olsen 1993). Paper 2 focuses on the importance of involving patients in decision making, and argues that this aspect is not fully incorporated in QALYs. However, I am not certain as to whether allowing for more patient autonomy is something which requires additional resources, or a shift in mentality and attitude among the medical profession as providers of care. If it is the latter, the policy implication becomes one of changing the doctors. If it is the former, the benefit from patient autonomy is to be traded off against QALYs foregone.

While section 1.5 below suggests that one should be wary of focussing exclusively on QALYs as a guide for priority setting, I still believe, however, that QALYs are potentially the best "exchange rate" for making comparisons across programmes which yield different types of health outcomes. As such a QALY could be called the "health economists' *ecu*".

1.4 Incorporation of considerations of equity in economic evaluations

There is some disagreement among leading health economists (Mooney et al 1991, Culyer et al 1992), as to whether equity should relate to equal *utilisation* or equal *access* for equal need. While this difference is not a topic of this thesis, I would draw attention to the wide agreement with respect to the concept of need. The health economics notion of need refers to the *potential for improved health*. According to Williams' (1978) seminal definition: "'need' is a quasi-supply concept: it means that a 'need' exists so long as the marginal productivity is positive". In other words, those who are in a severe health state, but for which no effective health care exists, will accordingly not be defined as being in

need of care, i.e. they are sick but not in need. Hence, *need* is a measure of potential health benefits.

However, quantifying need in terms of potential QALYs will of course not simultaneously solve the equity issue. As soon as need differs between patients, the question is how these differences are valued by society. The background for Paper 6 was that if a QALY has the same constant social value independent of how many QALYs a person gets, then no considerations of equity are incorporated. One interpretation of an equity concern is to attach *diminishing* social value to increasing numbers of QALYs received by any individual.

Paper 1 notes that if there are regional differences in costs per life saved, and resources are allocated in accordance with equal marginal costs per life saved, the result will be that individuals will face unequal risk exposure regionally. However, if, for reasons of equity, cost variations are disregarded, the opportunity cost of saving one life in an "expensive region" is more than one life foregone in the "cheaper region". Hence, when considerations of equity are to be incorporated in such choices, one would make trade-offs between efficiency in terms of the number of lives saved (or QALYs) and equity.

1.4.1 The main findings on Issue 4

Paper 6 presents a method for eliciting the implicit rate at which successive QALYs to one individual could be weighted. The suggested method is a form of person-trade-off which yields social weights for years with improved health when these are distributed between two different groups of people. The two corner options reflect the extreme views that only the number of years with improved health matters (efficiency) or only the number of persons benefiting matters (equity), while the inner options reflect trade-offs between efficiency and equity.

Paper 6 suggests that the more equitable preferences one has, the higher is the rate at which one would devalue an individual's additional year of improved health. Hence, a 'discount' rate is used as a mechanism for dealing with interpersonal equity. Further, attention is drawn to the adverse equity implication of the chosen discount rate. While

preferences for *interpersonal equity* imply a high rate, preferences for *intertemporal equity* (across generations) imply a low rate.

The study reported in Paper 7 attempted to elicit preferences for regional equity in terms of willingness to pay for a helicopter ambulance service. Although the reason "more equal access to health care regionally" was stated to be of high importance for WTP, it was *not* a significant determinant of WTP. On the other hand, WTP in the urban areas, where own probability of use is relatively low, was not significantly lower than WTP in the rural areas, where own probability of use is higher.

1.4.2 The policy relevance of Issue 4

A review of recent governmental commission reports and white papers on the objectives of the Norwegian health service (Olsen 1993) suggests that the most precise formulation of the equity objective is "equal access for equal need", but beyond that, the issue of equity is treated rather vaguely. Further, there is no indication of how considerations of equity could be incorporated. The impression is sometimes given that need is interpreted as synonymous with severity level rather than with potential health benefit. From this review it becomes evident that health programme analysts should focus more on the distributional consequences of a programme and preferably develop *methods* which incorporate equity concerns. Paper 6 has shown one method for incorporating equity concerns, while Paper 7 used the WTP approach as an attempt to elicit an implicit preference for regional equity.

1.5 Accounting for the nature of the health care programme which produces the health outcome

In Paper 7 comparisons were made between the mean WTP per QALY gained from three different programmes. Assuming that the health outcome is captured in terms of QALYs gained, the implication is that the higher the WTP/QALY in one programme compared with another, the more this programme is valued for reasons other than the QALY-outcome. A high WTP/QALY indicates the existence of valuable additional attributes

inherent in the particular programme, which are not incorporated in QALYs, *and/or* that the health outcome in terms of the QALYs produced by the programme is valued more highly. This distinction between benefits *beyond* outcome vs different values of the same QALY-outcome across programmes is important and should be treated as analytically separate issues.

1.5.1 Benefits beyond health outcome

Paper 2 suggests that there are more utility bearing attributes of health care than are incorporated in health outcome as measured by QALYs, e.g. information. There are at least two aspects here about which there is currently a low level of agreement. First, there is the question of the extent to which health care use may yield utility beyond that which is derived from improved health. Secondly, if there are non-health utility bearing attributes, should these be included in health economic evaluations?

There appear to be three different schools of thought on this issue. Gavin Mooney holds that, beyond QALYs, health care may 'produce' non-health benefits such as information, caring, anxiety reduction, communication. Further, he argues that there is utility from the process by which care is provided (e.g. degree of patient autonomy), which he calls 'process utility' (Mooney 1991, Gerard and Mooney 1993, see also Donaldson 1993, Ryan 1993, Ryan and Shackley 1993). However, in the literature these two types of benefits are often referred to as 'process utility'. The second school of thought holds that 'process utility' is already picked up by QALYs (the view held by Alan Williams), i.e. that there is no such thing as 'process utility' as separate from 'health outcome utility'. Lastly, Jack Dowie has recently argued that to the extent that there exists additional utility bearing attributes of health care, these *should* be included in the outcome measure (Dowie 1993).

A simple formalisation of this issue may help. Ignoring externalities, such as caring and contagion, the utility gained from own health care use can be expressed by a utility function based on Evans and Wolfson (1980). A particular individual gains utility, U, from health care use primarily through the positive impact of health care, HC, on health

status, HS . Health care *per se* is in general a dis-good, i.e. the actual consumption of health care yields disutility.[4] The +/− indicates the sign of the effects.

$$+ \quad + \quad -$$
$$(1) \quad U = U \,[HS \,(HC), \, HC]$$

Using this framework, 'process utiliy' cannot pass through the first argument. Hence, it involves the reduction of the disutility expressed by the second argument. Alternatively, a third argument is required to denote the attributes of health care which have *no* impact on health outcome. They are referred to as A for amenities (an elaboration of this approach is given in Clark and Olsen 1994).

$$+ \quad + \quad - \quad +$$
$$(2) \quad U = U \,[HS \,(HC), \, HC, \, A]$$

The extent to which there is such a thing as process utility as separate from health outcome utility will depend on which dimensions of health are accounted for in the health status measurement. If 'information', 'caring', etc. are interventions which have an effect on any of the health dimensions, they are classified as HC ; if not, they are classified as A . Which of the schools of thought that is 'right' as to the mere existence of the A-argument can be tested empirically, e.g. by the use of WTP. Different people can be presented with two types of treatments for the same diagnosis, one treatment gives health outcome only, while the other has *the same description of outcome*, but additionally some process utility yielding attributes are included. The relative difference in the stated WTP-figures would indicate the value of non-health benefits compared to health outcome. An alternative approach could be conjoint analysis (Ryan and Shackley 1993).

---

[4] Dental care is a good example. While most people experience disutility from undergoing dental treatment, certainly, it is the health outcome which makes us visit the dentist. In the classic Norwegian childrens' book 'Karius og Baktus': "Jens gikk hjem fra tannlegen. Han var lykkelig *fordi han ikke hadde huller og tannpine lenger*" (Egner 1990, emphasis added).

18

As for the second aspect, the extent to which benefits beyond health outcome are to be included in health economic evaluations crucially depends on where one stands as to whose preferences are to count. Is it *the sovereign consumer preferences of the patient* or *the paternalistic preferences of contributors* (i.e. those who subsidise the health service)?

Given that HC and A compete for the same resources [HC + A = H (= the total health care budget], there is an opportunity cost of providing non-health benefits in terms of health outcome foregone. I would conjecture that the value of A relative to the value of HC depends on who it is who is asked. The awareness of the importance of process utility is based on studies of *patients'* preferences (see e.g. Mooney and Lange 1993). However, asking those who subsidise the health service may lead to a different value of A. It is for *health* that there is a caring externality (Culyer 1980). The need for health care in an obligational sense relates to the actual *health* produced by health care. Evans (1984) puts it this way: "it is only effective, needed care that is the merit good. The social interest is in [the health status of one's fellow citizen], and in her health care use *only insofar as it contributes to that*" (emphasis added). Hence, if the health service aims to maximize *health*, then the 'process utility' from health care becomes irrelevant. If the health service aims to maximize *patients utility* from health care, then 'process utility' from health care must be accounted for.

Clark and Olsen (1994) argue that, on its own, neither of these two roles (patient and contributor) is correct in a universal, comprehensive health service. Rather account should be taken of the *ex ante* preferences of citizens who, for two different reasons, contribute to the funding of the health service. They have a subsidy motive for effective health care for *other people* in need, and an insurance motive for health care (both effective and process utility yielding) should they *themselves* require it.

Paper 7 uses willingness to pay (WTP) as a method for eliciting wider preferences for health care programmes. The focus of the study was that of valuing public sector health care programmes through the eyes of the citizens, in which we deliberately framed the questions as "willing to *contribute*" rather than "willing to *pay*". The respondents may have a positive probability of own benefit from the programmes and/or a subsidy motive

for providing the services to fellow citizens in need. The reasons given for the stated WTP revealed that respondents had both 'selfish' and 'altruistic' concerns.

1.5.2 The nature of the health care programme

Paper 7 does not provide insights as to the utility gained from the actual treatment *process* within the three health care programmes (helicopter ambulance, bypass operations and hip replacements) simply because the descriptions did not give information on this aspect. Rather, the descriptions focused on: health states before and after treatment; duration of benefit; and characteristics of those benefiting from the actual health care programmes to which resources could be allocated. Thus, in the context of eq. (2), the higher value of QALYs, in terms of WTP, produced by the helicopter ambulance is not explained by a valuable A , but rather that the particular HS (health outcome per se) is more highly valued. The high WTP/QALY gained from the helicopter ambulance compared with that for hip replacement seems to correspond to the "Rule of Rescue", which suggests that life saving QALYs are more highly valued than are health improving QALYs, i.e. different values of the same QALY-outcome.

According to Eddy (1991), one reason why Oregon's initial method failed was exactly this: "By focusing on the outcomes of the services instead of the services themselves, some important information was lost about the *nature* of the services that might affect their desirability. (...) one problem was that the use of [Quality of Well Being] states filtered out important information about features of particular services" (Eddy 1991). Hence, in the context of Paper 7, it is 'the *nature* of these services', e.g. bypass operations vs hip replacements, rather than the process by which the treatments are provided, that explains the different WTP/QALY gained.

In other words, one problem with focusing on QALYs is that people assign different values to them depending upon which programme produces them; analogous to the observation of different implicit values of life across programmes (Paper 1). The implication of this is that one should use programme specific weighting of QALYs. [This is not to be confused with programme-specific QALY-measurement (see Donaldson et al

1988, Weinstein 1988, Williams 1989]. Certainly, more research is required in this area. While acknowledging the measurement problems, Paper 7 draws attention to WTP as a methodological approach for eliciting the wider community preferences for health and health care beyond those which are included in the QALY algorithm. Another approach is the use of the person-trade-off type 'allocation questions', as suggested by Eddy (1991).

1.5.3 The policy relevance of Issue 5

The conventional wisdom in the elicitation of health state values is to describe the health states according to three dimensions; physical, psychological and social (Torrance 1986). In other words, the diagnosis is not described, nor the actual health care programmes that might do good. In a cost/QALY context, this implicitly denies the possibility that individuals' utility levels for health care resource allocation are affected by any characteristic other than the change in health outcome as measured within these three dimensions.

Preferences for priority setting in health care are certainly more diverse than those taken into account in cost-utility analysis. The Oregon experience is supportive for making such an assertion. If one aims at coming up with a priority list which corresponds as well as possible to the preferences which citizens have for the resource allocation of their health service, I find it hard *not* to take account of citizens' preferences beyond those which are encompassed in pure health state valuations.

2: Necessities or luxuries?

It does sound somewhat strange to distribute 'luxuries' according to need. It is this counter-intuituive labelling of health care as a luxury that appears to have evoked concern in the literature on cross-national comparisons of health care expenditures. Paper 8 shows that *health* might still be a necessity even if health *care* behaves as a luxury.

Although cross-national comparisons suggest that health care is a luxury, recent *cross-sectional* comparisons of household spending in the US (Rasell et al 1993) shows that the

budget share spent on health care *falls* with increased income level, implying that health care would be labelled a 'necessity'. While the low income families on average spend 20.5% of their income on health care, the richest group spend 10.2%. However, the absolute spending on health care increases from an average of $ 1,756 in the low income group to $ 13,234 in the richest group (Rasell et al 1993). Assuming that the relative increase in health care spending is not the same for every type of health care, there will be some types of health care which behave as luxuries.

The luxury vs necessity distinction is a purely technical one. I am somewhat puzzled that it could be surprising that health care is income elastic. Food is also "needed" but some types of food are evidently income elastic. Certainly, the commodity health care is very heterogeneous with respect to both effectiveness and willingness to pay. Some types of care are obviously less "needed" than are others. For some health care commodities consumed by the rich, I suspect that the luxury-label would be an appropriate description - beyond its technical meaning.

3: Concluding remarks

The health policy purpose of economic evaluations is to aid decision makers to allocate scarce health care resources in order to achieve as much as possible of a specified aim. This aim may of course vary, but the prinicipal objectives of most public health care systems relate to efficiency and equity. A prerequisite for sensible priority setting in the health sector is basically to compare costs and consequences. In 'demystifying' the CEA-approach, I have much sympathy with Eddy's (1991) "two general facts about cost-effectiveness analysis". I would therefore quote him at some length:

"One [fact] is that cost-effectiveness analysis is not a particular method, any more than mathematics is a particular formula or chemotherapy is a particular drug regimen. Rather, cost-effectiveness analysis is a collection of methods bound together by a set of principles. It is applicable to any problem in which there are multiple activities, the activities produce different amounts of 'goodness' and require different amounts of

resources, the total amount of resources is limited, and the objective is to maximize the total amount of goodness that can be produced by the available-resources. (...) The second general fact about cost-effectiveness analysis is that it does not assume or promote any particular philosophy about how resources should be allocated. (...) It is up to the people who assign the values to different services to decide whether they want to assign a higher value to preventing visible deaths than statistical deaths" (pp. 2140-1).

To the last sentence, I would add: and to assign the distributive weights which best reflect society's equity preferences.

The simple rationale for the use of economic evaluation in health care arises from a banality, namely that health care resources are scarce. The challenge then is to allocate these scarce resources in accordance with distributive rules which best correspond to the preferences of society. If these preferences relate to the maximization of health, and health can be captured in terms of QALYs, then the most efficient allocation of resources would be the one which maximizes QALYs, in which any QALY counts the same across programmes; i.e. "a QALY is a QALY is a QALY". However, in some of the enclosed papers I have argued that distributive preferences are more complex than that, implying that QALYs do *not* count the same; i.e. "a QALY is *not* a QALY is *not* a QALY". Paper 5 suggests that QALYs are weighted depending on *when* they occur. Intertemporal preferences weight present QALYs seven times higher than those arising in twenty years. Paper 6 suggests that QALYs are weighted depending on *how many* QALYs a patient acquires. In the pilot study, the twentieth year with a given health improvement was valued of less than one tenth of the first year of the same improvement to a different person. Paper 7 suggests that QALYs are weighted depending on *the nature of the programme* which produces it. There are additionally other weighting issues which require empirical research.

The principle of consumer sovereignty, to which most economists subscribe, requires that the value basis of economic evaluations corresponds to individuals' preferences. To varying extents Papers 2, 4, 5, 6, 7, and 8 deal with preferences for health and health care. Papers 5, 6, and 7 have been concerned with developing methodologies for eliciting preferences, and questionnaires have been designed for each of these papers (see the

appendices). The questionnaires on which Papers 5 and 6 are based use a person-trade-off approach, which focuses on social preferences, while Paper 7 uses a WTP-format which aims at capturing both the social and the 'selfish' preferences.

When deciding upon which approach to use for eliciting individuals' preferences for health care resource allocation, it is crucial to be aware of the choice context in which respondents are placed. There are essentially three different contexts, or decision roles. The first is *the patient*, and most often the hypothetical patient, who is asked to imagine herself in a particular health state from which a health status value is inferred, or who is asked her willingness to pay for the treatment of a particular disease. This context is most widely used in health status measurements, like CR, SG and TTO, and also in WTP studies. The second role is that of *the planners* or *pure contributors*. They are asked to make social comparisons between other patient groups (the PTO-approach), or asked how much they could be willing to subsidise a health care programme for which they themselves are not a part, e.g. charitable giving to health care for the poor. The third role might be termed *the ex ante fit citizen*. This role is the one which most fully captures the dual preferences, i.e. the selfish concern should I become a patient, and the social concern for others who already are, or will become, patients.

This third role is the one which is most relevant in societies with a comprehensive, publicly financed health service. One must not forget that this institution has two principal missions to fulfill. It is both an insurance agency and a redistributive agency. As such, it has to respond to its citizens' expected future use of care for themselves, if and when that is needed, and its citizens' altruism/subsidy motive for the use by other people who already are in need of care. I would acknowledge that the framing of choice contexts which places the respondent in this ex ante role seems to be the most problematical of the three. As such "the perfect could become the enemy of the merely good" (Drummond et al 1993). However, repeating a quotation from Paper 1: "The claim that something is inherently complicated does not excuse us from explicitly stating the principles which should be brought to bear upon the decision" (Marin 1983).
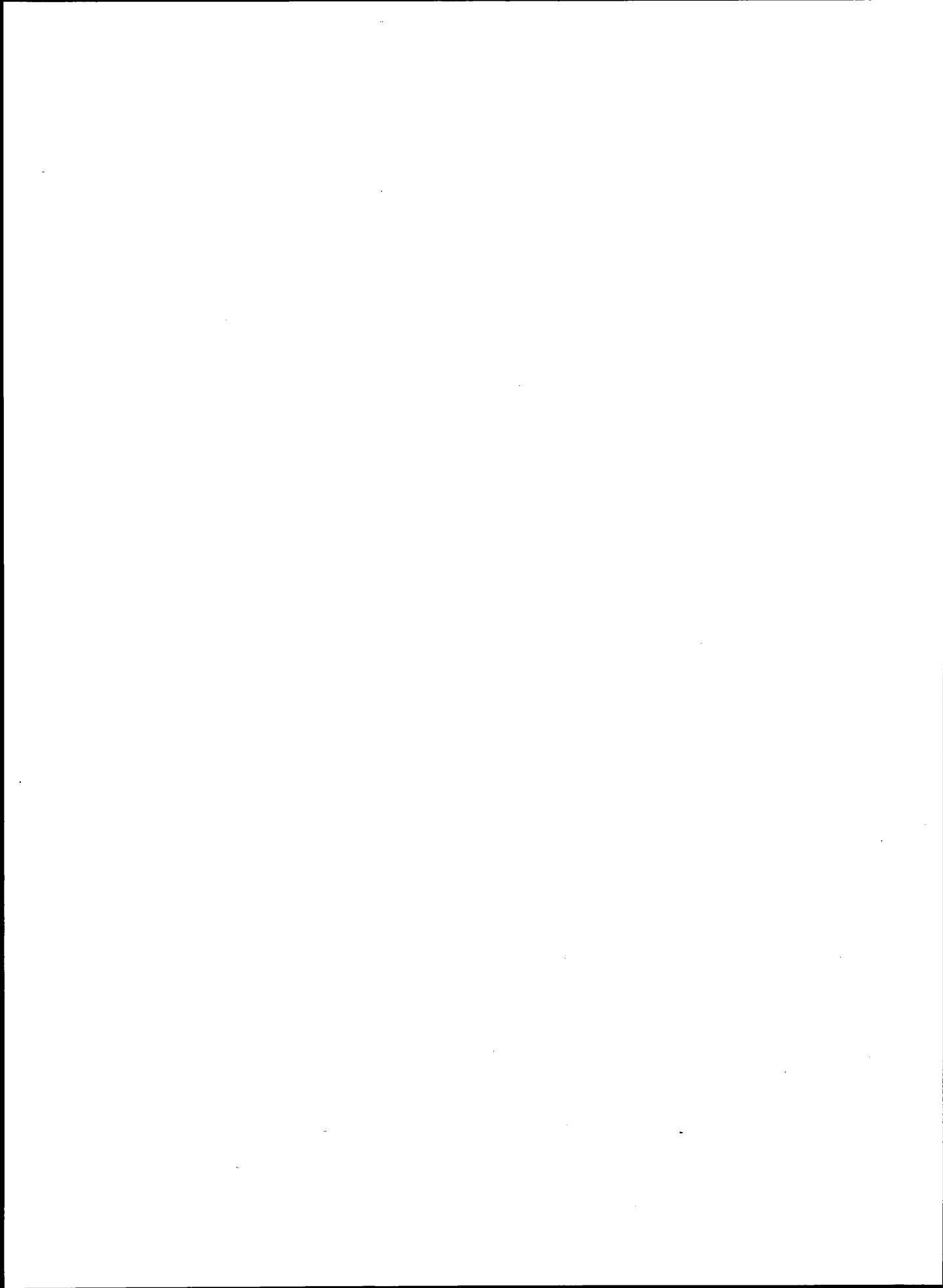
24

References:

Bergmo, T. (1993): Preliminary data from forthcoming Master thesis in economics, Department of Economics, NFH, University of Tromsø.

Blades, C.A., A.J. Culyer, and A. Walker (1987): Health service efficiency: Appraising the appraisers - A critical review of economic appraisal in practice, Social Science and Medicine, 25(5): 461-472.

Cairns, J. (1992): Discounting health benefits. Paper for the 2nd European conference in Health Economics. Paris: December, 1992.

Clark, D. and J.A. Olsen (1994): Agency in health care with an endogenous budget constraint, forthcoming in Journal of Health Economics.

Cropper, M., S. Aydede and P. Portney (1992): Rates of time preference for saving lives. American Economic Review, 82(2): 469-472.

Culyer, A.J. (1980): The Political Economy of Social Policy, Martin Robertson, London.

Culyer, A.J. (1989): The normative economics of health care finance and provision, Oxford Review of Economic Policy, 5(1): 34-58.

Culyer, A.J. (1990): Commodities, characteristics of commodities, characteristics of people, utilities, and the quality of life, in: S. Baldwin, C. Godfrey and C. Propper: Quality of Life: Perspectives and Policies, Routledge, London.

Culyer, A.J. (1991): The Economics of Health. Volume 1 and 2, Edward Elgar, Hants.

Culyer, A.J., E. van Dorslaar and A. Wagstaff (1992): Utilisation as a measure of equity by Mooney, Hall, Donaldson and Gerard, Journal of Health Economics, 11(1): 93-98.

Donaldson, C., A. Atkinson, J. Bond and K. Wright (1988): Should QALYs be programme-specific?, Journal of Health Economics, 7: 239-257.

Donaldson, C. (1993): Theory and practice of willingness to pay for health care. Discussion Paper 01/93, Health Economics Research Unit, University of Aberdeen.

Dowie, J. (1993): Process utility can seriously damage your health service evaluation but the generic measure of benefit should include service outcomes, Presented at the UK Health Economists' Study Group meeting, University of Strathclyde, Glasgow June 30 - July 2.

Drummond, M.F. (1981): Studies in economic appraisal in health care, Oxford University Press, Oxford.

Drummond, M.F., Ludbrook, A., Lowson, K. and Steele, A. (1986): Studies in economic appraisal in health care: Volume 2, Oxford University Press, Oxford.

Drummond, M.F., Stoddart, G.L. and Torrance, G.W. (1987): Methods for the economic evaluation of health care programmes, Oxford University Press, Oxford.

Drummond, M.F., A. Brandt. B. Luce and J. Rovira (1993): Standardizing methodologies for economic evalution in health care: Practice, problems and potential, International Journal of Technology Assessment in Health Care, 9(1): 26-36.

Eddy, D. (1991): Oregon's Methods. Did Cost-effectiveness Analysis Fail?, Journal of the American Medical Association, 266(15): 2135-2141.

Egner, T. (1988): Klatremus og de andre dyrene i Hakkebakkeskogen, Grøndahl, Oslo.

Egner, T. (1990): Karius og Baktus, Cappelen, Oslo.

The EuroQol Group (1990): EuroQol* - a new facility for the measurement of health-related quality of life, Health Policy, 16: 199-208.

Evans, R.G. (1984): Strained Mercy: The Economics of Canadian Health Care, Butterworths, Toronto.

Evans, R.G. (1993): Unpublished keynote speech given to the 14th Nordic HESG meeting, University of Tromsø, 18-20 August.

Evans, R.G. and A.D. Wolfson (1980): Faith, hope and charity: Health care in the utility function, Discussion Paper, Department of Economics, University of British Columbia, Vancouver.

Gerard, K. (1991): A Review of Cost Utility Studies: Assessing Their Policy-Making Relevance, Discussion paper 11/91, Health Economics Research Unit, University of Aberdeen.

Gerard, K. and G. Mooney (1993): QALY league tables: handle with care, Health Economics, 2: 59-64.

Gafni, A. and S. Birch (1993): Searching for a common currency: Critical appraisal of the scientific basis underlying European harmonization of Health Related Quality of Life (EuroQol), Health Policy, 23: 219-28.

Magnussen, J. (1990): Verdsetting av menneskeliv - en kommentar, Norsk Økonomisk Tidsskrift, 104 (3): 247-250.

Marin, A. (1983): Your money or your life. The three Banks review, 20-37.

Mooney, G. (1991): Agency in health care: What's the maximand? Discussing Paper 10/91, Health Economics Research Unit, University of Aberdeen.

26

Mooney, G., J. Hall, C. Donaldson and K. Gerard (1991): Utilisation as a measure of equity: weighting heat?, Journal of Health Economics, 10(4): 475-480.

Mooney, G. and M. Lange (1993): Ante-natal screening: What contributes to benefit?, Social Science and Medicine, 37: 873-878.

Nord, E. (1989): The significance of contextual factors in valuing health states, Health Policy, 13: 189-198.

Nord, E. (1992): An alternative to QALYs: The saved young life equivalent (SAVE), British Medical Journal, 305: 875-877.

Nord, E. (1993): Unjustified use of the Quality of Well-Being Scale in priority setting in Oregon, Health Policy, 24: 45-53.

NOU 1987:23 Retningslinjer for prioriteringer innen norsk helsetjeneste, Universitetsforlaget, Oslo.

Nygaard, E. and J.A. Olsen (1992): A model for discussing the consequences of including production gains in economic evaluation of health care programmes, Working paper 1/1992, Folkehelsa, Oslo.

OECD (1990): Health care systems in transition, OECD Social Policy Studies, no.7.

Olsen, J.A. (1990): Bør vi sette størst pris på de yrkesaktive? - et svar til Jon Magnussen's kommentar, Norsk Økonomisk Tidsskrift, 104 (3): 251-253.

Olsen, J.A. (1993): Aiming at what? The objectives of the Norwegian health service, forthcoming in Christiansen, T. (ed.): The Nordic Health Care System, Odense University Press, Odense.

Rasell, E., J. Bernstein, and K. Tang (1993): The Impact of Health Care Financing on Family Budgets, Economic Policy Institute Briefing paper, Washington DC.

Ryan, M. (1993): What do patients want from their Health Service? A case study of in-vitro fertilisation, Paper to UK HESG/FPM conference, University of York, January.

Ryan, M. and P. Shackley (1993): Economic evalution in health care: getting beyond health outcome, HERU Discussion Paper 06/93, Department of Public Health and Economics, University of Aberdeen.

St. meld. nr. 41, 1987-88 Helsepolitikken mot år 2000, Nasjonal helseplan.

Torrance, G.W. (1986): Measurement of health state utilities for economic appraisal, Journal of Health Economics, 5: 1-30.

TNL 1991, Ventelisteforskriftene, Tidsskrift for Den norske Lægeforening, nr. 13: 1672

Wagstaff, A. (1991): QALYs and the equity-efficiency trade-off, Journal of Health Economics, 10: 21-41.

Weinstein, M.C. (1988): A QALY is a QALY is a QALY - or is it?, Journal of Health Economics, 7: 289-290.

Weinstein, M.C. and W.B. Stason (1977): Foundations of cost-effectiveness analysis for health and medical practice, New England Journal of Medicine 296 (31): 297-317.

Williams, A. (1978): "Need" - an Economic Exegisis, in Culyer, A.J. and K. Wright (eds.), Economic aspects of health services, Martin Robertson, London.

Williams, A. (1985): Economics of coronary artery bypass grafting, British Medical Journal, 291: 326-329.

Williams, A. (1987): Health Economics: The Cheerful Face of the Dismal Science?, in Williams, A. (ed.), Health and Economics, Macmillan Press, London

Williams, A. (1989): 'Should QALYs be programme-specific?' by Donaldson, Atkinson, Bond and Wright, Journal of Health Economics, 8(4): 485-487.

Williams, A. (1992): Cost-effectiveness analysis: is it ethical?, Journal of Medical Ethics, 18 (1): 7-11.

Paper 1

VERDSETTING AV MENNESKELIV – ELLER HVORDAN VI KAN SETTE PRIS PÅ HVERANDRE

Av Jan Abel Olsen*

Artikkelen drøfter ulike metoder for hvordan menneskeliv kan verdsettes; human capital, betalingsvilje for risikoreduksjon og implisitt verdsetting. Vi er opptatt av hvordan metodene samsvarer med tre alternative kriterier for allokering av livreddingsressurser, nemlig likhet, maksimal nytte og konsumentsuverenitet. Etter en kritisk gjennomgang av litteraturen pekes det på noen mulige årsaker til de observerte forskjeller i verdien av menneskeliv på ulike områder. Artikkelen diskuterer hvilke allokeringskriterier som bør følges og dermed hvilke(n) metode(r) som er mest relevant å anvende for offentlige beslutningstakere.

1. Innledning

Når økonomer berører temaet økonomisk verdsetting av menneskeliv, vekker det – ikke overraskende – umiddelbart negative reaksjoner. Selv om man setter pris på et menneske, trenger man vel ikke *prissette* det!? Våre edle hensikter med å sysle med temaet må derfor klargjøres. Vi argumenterer for at et tema som umiddelbart kan synes perverst tvert imot dreier seg om å allokere ressurser slik at ens dødsrisiko reduseres. Det er gjennom eksplisitte prioriteringer at dette kan skje. Nettopp derfor er temaet livsviktig, og det inviterer økonomer til en sjeldent human geskjeft.

Når vi her snakker om menneskeliv, dreier det seg om statistiske liv i motsetning til bestemte navngitte personer. Det dreier seg også om å redusere sannsynligheten for en *for tidlig død*. Det er nemlig ikke bare

J.M. Keynes som kom til den erkjennelse at «in the long run we are all dead.»

I forlengelsen av menneskeliv som nyttemål er det blitt en økende interesse for bruk av leveår og «kvalitetsjusterte leveår» (se bl.a. Willams 1985). Vi skal ikke gå inn på dette, heller ikke pretenderer vi å utlede noen endelig verdi på menneskeliv. I stedet diskuteres temaet prinsippielt.

To spørsmål utgjør selve kjernen i problemstillingen; hvor mye ressurser bør brukes .på livredding og etter hvilke kriterier bør ressursene allokeres? Spørsmålet om hvor stort budsjettet bør være avhenger av hvor mye vi faktisk verdsetter menneskeliv. Dersom samfunnets overordnede mål var å redde flest mulig, ville alle ressurser gå til livredding og livsforlenging. Når en slik ressursallokering er fremmed, er det et uttrykk for at innbyggerne foretrekker et kortere liv med utdanning, kultur og materielle goder framfor et langt liv på «eksistensminimum». Størrelsen på «livreddingsbudsjettet» avspeiler således en *avveining* mellom to typer goder; lav dødsrisiko og andre goder. I nyttefunksjonen avveies materielle goder som røyk og motorsykler mot livsforlengende eller dødsrisikoreduserende goder som grønnsaker og joggesko. Hvor mye ressurser som bør brukes på livredding avhenger følgelig av slike avveininger. Naturligvis er det derfor ikke noe endelig svar på hva som er optimalt budsjett.

I prinsippet er det tre kriterier for hvordan ressurser til livredding bør allokeres. For det *første* at tilgang på livreddingstiltak skal være *uavhengig av inntekt*, «likhetskriteriet». Dette kriteriet samsvarer med den grunnleggende egalitære ideologi bak offentlig helsevesen og også en akseptert etisk norm om at ikke inntekten skal avgjøre ens dødsrisiko. For det *andre* at *flest mulig* menneskeliv skal reddes, «maksimal nytte kriteriet». Dette er et klassisk økonomisk maksimeringsproblem, nemlig størst mulig måloppnåelse til gitt ressursinnsats. Det følger også et utilitaristisk syn om «mest mulig til flest mulig».

For det *tredje* at folks *preferanser* for risikoreduksjon i de ulike sektorer skal bestemme allokeringen mellom sektorer, «konsumentsuverenitet kriteriet». Dette kriteriet bygger på en normativ grunnregel i økonomisk velferdsteori, nemlig at konsumentene alltid vet best. Ut fra dette bør følgelig hensyn tas til hvor sterke preferanser konsumentene uttrykker for risikoreduksjon i de ulike sektorer. Dersom preferansene måles gjennom betalingsvilje, vil dette komme i konflikt med det første kriteriet. Dersom

3

konsumentene er desinformert om de objektive dødsrisiki og/eller har preferanser for *måter* de ønsker å unngå å dø på, vil det tredje kriteriet være i konflikt med det andre.

Konsumentsuverenitetskriteriet har til hensikt å realisere mulige Pareto-forbedringer. Det baserer seg på folks egne avveininger i mikro gitt den eksisterende inntektsfordeling. Dermed blir det klart samsvar mellom de avveininger hver enkelt innbygger gjør og det prinsipp samfunnet bekjenner seg til. Likhets- og maksimal nytte kriteriet tar også stilling til *fordelingsspørsmålet*. De er i tråd med et fordelingsprinsipp som sier at allokeringen skal skje *uavhengig* av betalingsevne.

Formålet med denne artikkelen er å drøfte ulike metoder for verdsetting av menneskeliv med ekplisitt referanse til de tre nevnte kriterier. Etter en kritisk gjennomgang av litteraturen diskuteres kort noen årsaker til de observerte forskjeller i verdsetting av menneskeliv på ulike områder. Deretter setter vi de tre allokeringskriteriene opp mot hverandre og drøfter forholdet mellom konsumentsuverenitet og hhv. likhet og maksimal nytte.

2. Metoder for verdsetting av menneskeliv

Metodene er av tre hovedtyper; human capital, betalingsvilje og implisitt verdsetting. Human capital metoden verdsetter mennesker ut fra deres verdiskapning. Betalingsvilje for risikoreduksjon måler hvordan individer verdsetter redusert sannsynlighet for å miste livet. Implisitt verdsetting måler verdien av menneskeliv ut fra hvor mye ressurser samfunnet faktisk allokerer på tiltak som har til hensikt å redde liv; «livreddingstiltak».

2.1 Human capital

Human capital refererer til menneskenes produktive kapasitet som er bestemt av utdanning, evner og ferdigheter. Den produktive kapasitet, og dermed bidraget til nasjonaløkonomien, varierer følgelig mellom mennesker.

Fram til begynnelsen av 70 tallet var de fleste forsøk på verdsetting av menneskeliv basert på human capital tilnærminga (Dublin og Lotka 1946, Reynolds 1956, Weisbrod 1961, Dawson 1967, Rice og Cooper 1967, Hanlon 1969). Ut fra forutsetningen om at *bruttolønna* viser verdien av

menneskets grenseprodukt, måles samfunnets økonomiske verdi av mennesket til nåverdien av framtidige forventede inntekter justert for forventet produktivitetsvekst. Da såvel lønna som antall gjenværende år i arbeidslivet varierer, følger det at unge med høy lønn har høyere human capital enn eldre med lav lønn.

I en mye referert studie av Rice og Cooper (1967) ble den amerikanske befolkning delt inn i sosiodemografiske grupper etter alder, utdanning, kjønn og rase. Høyest human capital hadde hvit mann i alderen 30–34 år med minimum 16 års utdanning; $ 223.471. Lavest human capital ble beregnet for farget mann over 85 år; $ 396. Kjønnsforskjellene fortjener en kommentar. Ifølge deres tabeller hadde en nyfødt gutt nesten dobbelt så høy human capital som ei nyfødt jente. Kvinner hadde lavere human capital enn menn helt fram til de nærmet seg pensjonsalderen. Men siden husmødres marginalprodukt ble beregnet tilsvarende gjennomsnittslønn for en «husmorvikar» (domestic worker), fikk kvinner en høyere human capital enn menn da de passerte 60 år pga. den servicen de yter sine pensjonerte menn(!):

«As women age and leave the labor force, they are still credited with a dollar value if they stay home and keep house, which prevents their earnings from falling as rapidly as those men who have left the labor force.» (Dorothy Rice og Barbara Cooper 1967)

Human capital tilnærmingen har en klar fordel; den er enkel. Tilnærmingen har imidlertid en enda klarere svakhet; den måler ikke hva den skal måle, nemlig hvor mye vi verdsetter *et liv*, men derimot verdien ved å være *i live*. Derfor har ifølge Rhoads (1980) nesten alle akademiske økonomer tatt avstand fra tilnærminga som metode for verdsetting av menneskeliv. Men metoden lever videre. De seneste bidragsytere synes likevel å ha en mer forbeholden omgang med den og legitimerer den ved å legge et annet innhold i resultatene; i stedet for å måle verdien på menneskeliv måler human capital samfunnsøkonomiske *kostnader ved sykdom* (Hodgson og Meiners 1982, Scitovskty 1982).

Denne eksplisitte dreining mot kostnader ved sykdom samsvarer med den opprinnelige intensjonen med metoden. Ifølge Robinson (1986) var nemlig bakgrunnen for økonomers bruk av human capital metoden å vise at det var samfunnsøkonomisk lønnsomt å allokere ressurser til helsesektoren. Ikke overraskende derfor at human capital baserte nytte kostnads kalkyler er populære blant helsepersonell for å «bevise» overfor politi-

kerne hvor lønnsomt det er for samfunnet å allokere penger til nettopp deres felt. Flere bidrag i denne tradisjonen kommer da også fra forfattere med tilknytning til det amerikanske helsedepartement (Brody 1975, Cooper og Rice 1976, Hodgson og Meiners 1979, 1982).

Intensjonen er hederlig nok. Det er heller ikke uvanlig å prøve å snu noe normativt; «vi mener det *bør* allokeres mer penger på helsevesenet», til noe positivt; «human capital kalkyler viser at det *er* lønnsomt.» Følgende sitat kan umiddelbart tolkes dithen at human capital metoden er en verdinøytral økonomisk kalkyle:

«If one wants to know what the economic burden of illness was last year, what resources will be saved by preventive measures that reduce the incidence of disease, or what the economic impact of improved survival rates will be, the human capital method provides an appropriate, although partial measure.» (Hodgson og Meiners 1982)

Med partiell menes at human capital metoden ikke pretenderer å måle annet enn strengt økonomiske sider. Den måler den *samfunnsøkonomiske nytten* ved å sette folk i stand til å gå tilbake til produktiv innsats – tilsvarende en kalkyle over reparasjon/vedlikehold av en produksjonsfaktor.

Policy-implikasjonen er at *dersom nytten*, målt ved nåverdien av framtidige inntekter, *er høyere enn kostnadene* ved å redde eller behandle en gruppe mennesker, er det samfunnsøkonomisk lønnsomt å bruke ressurser på å redde mennesker i denne gruppen.

2.1.1 Noen reservasjoner

Reservasjonene mot human capital metoden ligger på to plan; metodikk og etikk. Metodisk dreier det seg om bruken av *bruttolønn som mål på ens human capital* – en kapital som har samfunnsøkonomisk verdi i kraft av dens produktive kapasitet. Det karakteristiske ved den «humane kapitalen» er at den er iboende i mennesket. Den er bestemt av ens evner og ferdigheter, og ikke minst er den ervervet gjennom formell utdanning.

Høy lønn *kan* derfor reflektere høy human capital, men det er ikke ensbetydende. Lønna avhenger av ens *relative forhandlingsstyrke* på arbeidsmarkedet; sterke fagforeninger oppnår høyere lønn enn svake. Lønna avspeiler også *diskriminering* som ikke henger sammen med pro-

duktivitetsforskjeller; kvinner og menn med samme human capital har ofte forskjellig lønn.

Et annet poeng er at lønna ikke nødvendigvis er et bilde av hvor mye human kapital en *selv* besitter, men derimot hvor mye fast kapital en har bak seg i produksjonen. I en kapitalintensiv sektor vil arbeidets grense-produkt være høyt, noe som gjerne gir seg utslag i høyere lønn enn i arbeidsintensive bransjer. Det gir derfor ikke mening å bruke slike lønns-forskjeller som indikasjon på ulikheter i menneskers produktive kapasi-tet. (Men i den grad høy lønn skyldes at antall personer som kan utføre en bestemt jobb er *begrenset*, vil naturligvis lønna reflektere en høy *human kapital* hos de som besitter de etterspurte ferdigheter.)

De alvorligste reservasjoner dreier seg om *etikk*. For hva er det et tabelloppsett av Rice og Cooper's type *egentlig* forteller? Det gir den samfunnsøkonomiske nytten av å behandle/redde personer i ulike sosio-demografiske grupper. Den økonomiske fortolkning som følger er at til identisk kostnad pr. reddet menneskeliv vil det være samfunnsøkonomisk mest lønnsomt å redde de med høyest human capital. *Steget er kort fra en intendert verdinøytral kalkyle til en fortolkning av denne som et prioriteringskriterium*, der folk sorteres etter nåverdien av sine framtidige inntek-ter. Bakerst i køa kommer kvinner, folk med lav utdanning og eldre. Jack Wiseman var tidlig ute med å påpeke:

«The young (....), the basically fit, and those with the highest expected earnings, would provide the highest rate of return and would therefore be given access to medical resources on the most favorable terms. The old, in contrast, constitute a liability... Indeed, if growth is the sole aim of our policymaker, there might be a strong case for providing only one medical service for those who can no longer work: euthanasia.» (Wiseman 1963)

Prioriteringer i henhold til pasientenes «samfunnsøkonomiske verdi» kan observeres i vårt hjemlige helsevesen. Innføring av bedriftshelsetje-neste må kunne sies å være begrunnet ut fra et ønske om at de ansatte skal holde seg friske og produktive heller enn at det er ment å være et frynsegode. Ideen om å bruke trygdemidler til økt behandlingskapasitet følger også et human capital resonnement; dersom verdien av produk-sjonsbortfallet i den tida pasienten går sykemeldt og venter på behandling er større enn kostnadene ved behandling, vil det være samfunnsøkono-misk lønnsomt å øke behandlingskapasiteten.

2.2 Betalingsvilje for risikoreduksjon

De første spede forsøk i denne tradisjon søkte å finne fram til *folks verdsetting av sitt eget liv*, V, med utgangspunkt i følgende enkle formel:

$$V = \frac{dv}{dr} \qquad (2.1)$$

der dv er den minimale kompensasjon som individet forlanger for å påta seg en risikoendring dr. Med andre ord, ved å dividere den minimale kompensasjonen med risikoendring, framkommer en endelig pris på eget liv. Dersom en f.eks. aksepterer 20.000 kr ekstra for å påta seg en jobb som innebærer en økt sannsynlighet for å dø på 0,01, vil implisitt verdi på eget liv være 20.000/0,01 = 2 mill. kr.

En åpenbar svakhet ved metoden var forutsetningen om *linearitet mellom sannsynlighet og betalingsvilje* over hele området. Altså, dersom et individ står overfor «valget» mellom sikker død og fortsatt liv, forutsettes at vi ut fra hennes tidligere betalingsvilje for redusert risiko kan utlede hennes endelig verdi på sitt liv! Men, en endelig pris eksisterer ikke. I en slik hypotetisk situasjon vil personen kreve uendelig stor kompensasjon for å bli fratatt retten til livreddende behandling. Det er dermed ikke mulig å realisere Pareto-forbedringer.

I Thomas Schelling's banebrytende artikkel fra 1968 ble fokus dreid fra ens verdsetting av eget liv til ens verdsetting av redusert sannsynlighet for død:

«It is not the worth of human life that I shall discuss, but of 'life saving', of preventing death. And it is not a particular death, but a statistical death.»

Videre var han opptatt av individers preferanser:

«It is worth while to remind ourselves that the people whose lives may be saved should have something to say about the enterprise».

Dette perspektivet tas opp av Mishan (1971) og Jones-Lee (1974, 1976). Mishan's utgangspunkt var at prosjekter med stor sosial nytte vil kunne kreve noen liv. Hvordan skal disse verdsettes i en nytte kostnads analyse? Hans poeng var at det ikke er kjente liv som «ofres» men at alle involverte «ofrer» med økt dødsrisiko. Følgelig er det den økte dødsrisiko som skal kompenseres:

«the relevant sum to be subtracted from the benefit side are no longer those which compensate each persons for their certain death but are those

sums which compensate each person in the community for *the additional risk* to which he is to be exposed.» (Mishan 1971)

La oss tenke oss at et tiltak som gjennomføres vil innebære at ett menneske innenfor en bestemt populasjon vil dø. *På forhånd vet ingen hvem dette blir.* For å akseptere denne økte risiko, forutsettes at befolkningen krever kompensasjon. Når man summerer den uttrykte kompensasjonen som hvert medlem av populasjonen forlanger – som naturligvis vil variere, får vi fram *verdien på et menneskeliv.* Dette kan uttrykkes som:

$$V = \sum_{i=1}^{N} dv_i \qquad (2.2)$$

der V er verdien på et menneskeliv og dv_i er kompensasjonen som forlanges av individ i. Da vi forutsatte at det er ett individ som ofres, vil risikoøkningen pr. definisjon være gitt ved $dr = 1/N$. Legg merke til at denne verdien, V, ikke er framkommet ut fra en beregning av ett bestemt individs betalingsvilje for å unngå en sikker død, som i (2.1). Derimot reflekterer den kompensasjonen en bestemt populasjon forlanger for å ofre ett av sine medlemmer.

Mens Mishan er opptatt av den minimale kompensasjon som er nødvendig for at individer vil godta økt dødsrisiko, skriver Jones-Lee om betalingsvilje for risikoreduksjon. Selv om risikoendringen går i motsatt retning, er tilnærmingene i prinsippet like. De tar begge utgangspunkt i en nyttefunksjon med to goder; velstand (materielle goder) og sikkerhet (lav dødsrisiko). Konsumenten står overfor en avveining mellom velstand og sikkerhet, der økt sikkerhet innebærer redusert velstand. I følge Jones-Lee er den forventede nytten til et individ bestemt ved:

$$E(U) = (1 - p^*)L(W^*) + p^*D(W^*) \qquad (2.3)$$

der p^* er en gitt sannsynlighet for død, W^* er en gitt velstand (inntekt), $L(W^*)$ er nytten dersom en lever, mens $D(W^*)$ er nytten ved å være død (!). (Dardis (1980) opererer med en forenklet utgave av denne nyttefunksjonen, ved å forutsette at tilfellet død gir null nytte i beslutningstakerens ex ante vurderinger, mao. at en ikke tar hensyn til nytten av å gi arv til de etterlatte.)

Dersom individet kan kjøpe seg fri fra en viss dødsrisiko til p < p*, vil hun måtte avstå en del velstand. Det en maksimalt er villig til å betale, V, vil pr. definisjon være det beløp som gir samme forventede nytte som initialt:

$$(1\text{-}p)L(W^* - V) + pD(W^* - V) = (1 - p^*)L(W^*) + p^*D(W^*) \qquad (2.4)$$

Det er to sentrale sammenhenger mellom p og V: (1) over en viss maksimal risikogrense, p', er det ikke noe beløp man vil godta for å øke risikoen, og (2) det maksimale en er villig til å betale for en gitt risikoreduksjon vil avta med avtakende risikonivå, dvs. at ens betalingsvilje for en risikoreduksjon fra 0,2 til 0,1 er mindre enn fra 0,5 til 0,4. Intuitivt virker det rimelig å anta at dess høyere risikonivå en befinner seg på, dess mer vil en være villig til å betale for en gitt risikoreduksjon. Disse egenskaper kan illustreres ved en konveks indifferenskurve i et diagram med to goder; sannsynlighet for å overleve, (1 − p), og velstand, V.
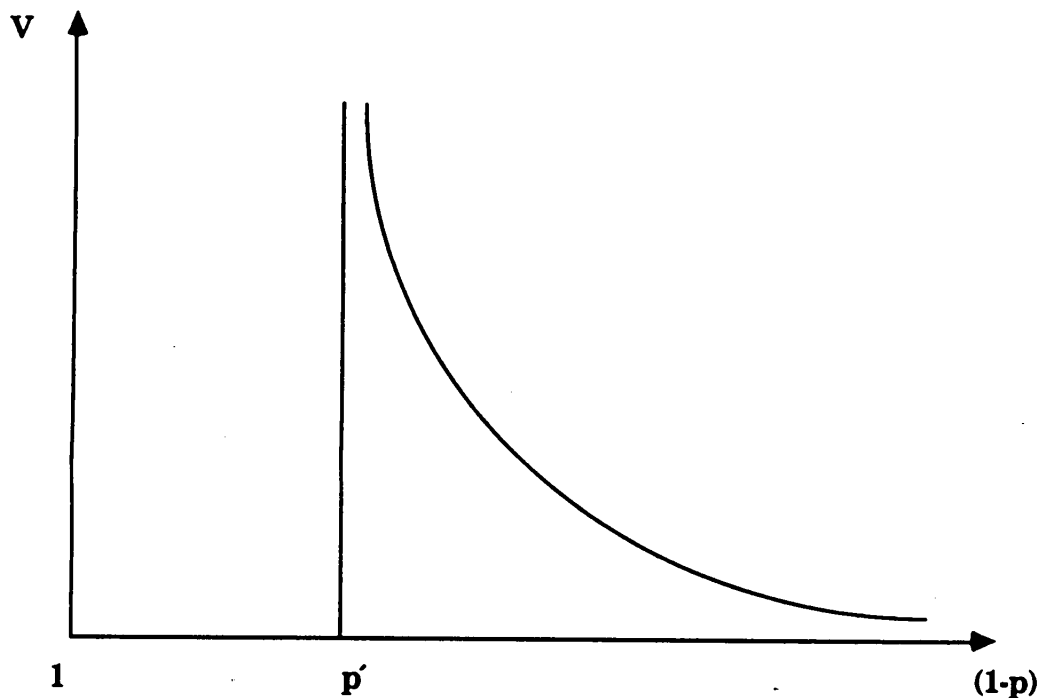
Ut fra en gitt risikoendring vil verdien på et menneskeliv være gitt ved *gjennomsnittet* av konsumentenes marginale substitusjonsrater mellom velstand og sannsynlighet for å overleve, tilsvarende V i (2.2). For en nærmere presisering, se Bergstrom (1982) og Dehez og Dreze (1982).

Felles for kjernebidragene fra Schelling, Mishan og Jones-Lee er at de alle behandler verdsetting av menneskeliv som en ex ante risikoreduksjon, noe som innebærer at *det dreier seg om statistiske i motsetning til kjente liv*. Broome prøvde å imøtegå denne «conventional wisdom» i noe som ble en mye omtalt artikkel (Broome 1978). Den er verd å nevne, da den avstedkom mange svar som inneholdt nødvendige presiseringer.

Essensen i Broome's kritikk er at selv om *individer* er villige til å pådra seg risiko, kan ikke en *offentlig beslutningsmyndighet* verdsette et tapt statistisk liv i en nytte-kostnads kalkyle fordi dette menneske, uansett hvor statistisk det måtte være, vil ha en uendelig pris på livet sitt, noe som innebærer at en nytte-kostnads analyse over et prosjekt som krever minst et liv alltid vil være negativ. Broome hevder at årsaken til at prosjektet lar seg gjennomføre er at *velgerne ikke er fullt informert* — de vet ikke hvem som vil dø. \

Poenget om at tapte statistiske liv til syvende og sist blir tapte kjente liv som ikke kan kompenseres, forkludrer forholdet mellom ex ante og ex post. Som påpekt av Jones-Lee (1979) burde verdivalgene om en samfunnsmessig beslutning reflektere individenes preferanser slik disse er *når beslutningen fattes*. Ut fra den informasjon som da finnes *er* velgerne fullt informert.

Broome ble også kritisert for at han verken leverte alternative metoder eller forslag til forbedringer av eksisterende måter å verdsette menneskeliv på. Mishan omtalte han slik:

«He is a knight errant in search of a holy grail. I wish him luck in his wanderings — better luck, at any rate, than he had in his 1978 essay.» (Mishan 1981)

McGuire og Mooney (1985) hevder at Broome for det første misforstår det økonomiske *verdibegrepet*. Livet har høy bruksverdi men ingen bytteverdi. Risikoen for å miste det kan derimot byttes med materiell velstand — en velstand man naturligvis kun får glede av dersom en overlever. For det andre innebærer Broomes forsøk på å så tvil om ex ante vurderinger at han glemmer at økonomi dreier seg om *valg* mellom alternativ med ulike *forventede* utfall.

Broome synes å ha en heller konspiratorisk oppfatning av nytte-kostnads analyser som et beslutningsverktøy for en stat som søker en legitimering for å ofre noen innbyggeres liv for at prosjekter skal kunne gjennomføres. En annen oppfatning − «nøytral» eller naiv − er at den implisitte verdsetting av menneskeliv som framkommer gjennom summen av den minimale kompensasjon er et bilde av innbyggernes faktiske tilpasning − en genuin respekt for deres risikopreferanser.

2.2.1 Empiriske studier

Ut fra observert atferd vet vi at individer er villige til å betale for å redusere dødsrisiko, samtidig som kompensasjon normalt forlanges for å akseptere økt dødsrisiko. En rekke empiriske studier er gjort for å beregne hvordan folk verdsetter ulike typer av slike risikoendringer. Disse studier er av tre slag; «sikkerhetsrelatert konsum», spørreskjema og kompensasjon for risikable jobber.

Den første metoden er opptatt av å måle hvordan folk gjennom sin *markedsatferd* forholder seg til risiko. Melinek (1974) studerte hvordan fotgjengere i London som løper over veier istedet for å bruke underganger avveier tidsbesparelser med økt risiko. De hadde en verdi på £ 87.000 i 1972. Blomquist (1979) så på bruk av sikkerhetsbelter i bil. «Kostnaden» var at det tar åtte sekunder å ta på og av sikkerhetsbeltet. Med en tidskostnad lik halvparten av bruttolønn ble så verdien av en amerikansk sjåfør beregnet til $ 370.000 i 1978.

Spørreskjema ble brukt av Jones-Lee (1976) for å undersøke folks betalingsvilje for økt flysikkerhet, som forøvrig viste seg å være hele £ 2,5 millioner. Dardis (1980) spurte folk om betalingsviljen for røykvarslere og beregnet amerikanernes verdsetting av livet til $ 101.000−$ 137.000 i 1979. Den første større spørreskjemaundersøkelse, basert på et representativt nasjonalt utvalg, ble utført av Jones-Lee et al (1985) som spurte om folks betalingsvilje for transportsikkerhet. Ut fra gjennomsnittlig betalingsvilje ble et statistisk liv verdsatt til £ 1,5 mill.

At folk må kompenseres for å påta seg risikable jobber ble påpekt av Adam Smith, som hevdet at under ellers like forhold mellom to jobber må den med høyest risiko tilby høyere lønn for å tiltrekke arbeidere. Empiriske studier fra USA av Thaler og Rosen (1976) og Viscusi (1978) støtter denne hypotesen. Imidlertid vil det være slik at arbeidere med lav risiko-

12

aversjon vil være overrepresentert i farlige jobber, nettopp fordi de er villige til å akseptere slike jobber til lavere kompensasjon enn andre. En fersk studie av Robinson (1986b) bekrefter at sammenliknbare arbeidere i farlige jobber tjener bedre enn i sikre. Han la imidlertid vekt på at farlige jobber også var dårligere jobber enn de sikre mht. arbeidsforhold, avansementsmuligheter og status. Det er B-laget på arbeidsmarkedet – med få alternativer – som henvises til de farlige jobber, jobber det er vanskelig å «avansere opp fra».

Ikke overraskende er risikable «hvitsnippjobber» betydelig høyere kompensert. Marin og Psacharopoulos (1982) sin undersøkelse fra England viser en implisitt verdi på «nonmanual workers» som er hele 3,5 ganger større enn «manual workers».

Betalingsvilje for livsforsikringer har vært diskutert som en metode for beregning av ens egen verdsetting av livet sitt, men den uttrykker ikke ens holdning til risikoen for å miste sitt eget liv men ens holdning til risikoen for det økonomiske tap som ens etterlatte vil lide. (Imidlertid viser flere kriminalsaker at det *kan* være sammenheng mellom livsforsikringens størrelse og risikoen for egen død. En høy livsforsikring øker nemlig insitamentene for en person til å avlive sin kjedelige forsørger. En optimal livsforsikring må følgelig baseres på en avveining mellom ønsket om at en etterlatt skal gis økonomisk trygghet og muligheten for at denne personen foretrekker en høy pensjon framfor en levende forsørger.)

2.2.2 Diskusjon

Selv om betalingsviljemetoden for et tiår siden fikk status som «the conventional methodology» blant akademiske økonomer (Jones-Lee 1982), er det flere problematiske sider ved den; persepsjonsproblemer, måleproblemer og fordelingsaspektet.

For konsumgoder som endrer dødsrisiko er forutsetningen om at konsumentene alltid vet sitt eget beste problematisk. Normalt kjøper konsumentene varer med garanterte egenskaper. Et garantert resultat kan ikke tilbys ved betaling av redusert *sannsynlighet* for å dø. Videre kan konsumenter bare gjøre meningsfulle valg ut fra erfaring. Det er ytterst sjelden at en erfarer at «livreddingsgoder» (røykvarsler, sikkerhetsbelter etc.) faktisk redder ens liv.

Men det grunnleggende problem ved betalingsviljemetoden er det faktum at folk har store *problemer med å skjønne innholdet i og forholde seg konsistent til små risiki.* (se bl. a. Hammerton 1982, Thaler 1982). Dette problemet kommer til syne i flere empiriske studier. En interessant observasjon kan gjøres i Blomquist (1982) som ut fra sammenlikning av en rekke empiriske studier viser hvordan verdien på menneskeliv øker med avtakende risikoreduksjon. En av de refererte studier var en spørreskjemaundersøkelse om folks betalingsvilje for redusert risiko for kjernekraftulykker der verdien på menneskeliv ble beregnet til $ 62.000 ved en risikoreduksjon på 10^{-3}, $ 428.000 ved en risikoreduksjon på 10^{-4} og $ 3.576.000 ved en risikoreduksjon på 10^{-5}.

Disse beregningene er meget interessante, fordi de indikerer at respondentene ikke har skjønt innholdet i de ulike risikoreduksjoner, *men oppfattet risikoendringene som relativt like tall.* For dersom de to siste verdiene hadde blitt beregnet ut fra samme risikoreduksjon som det første, nemlig 10^{-3}, ville verdien på menneskeliv blitt hhv. $ 62.000, $ 42.800 og $ 35.760.

Tilsvarende persepsjonsproblem gjorde seg gjeldene blant respondentene til Jones-Lee et al (1985), der nesten halvparten uttrykte samme betalingsvilje for forskjellige risikoreduksjoner. Det er ikke bare risiko*reduksjoner* men også risiko*nivå* folk har problemer med å forholde seg konsistent til. Smith og Desvousges (1987) varierte utgangsrisikonivå («baseline risk») mellom ulike respondentgrupper og fant at verdsetting av risikoreduksjon var nærmest konstant over store forskjeller i risikonivå. Dette rimer dårlig med forutsetningen om at betalingsviljen avtar med risikonivå.

Persepsjonsproblemet berører forholdet mellom objektiv og subjektiv risiko. Objektiv risiko er den faktiske eller statistisk korrekte, mens subjektiv risiko er den konsumentene opplever og forholder seg til. Den subjektive kan enten være større enn den objektive og inneholder da et «angstelement», eller den kan være mindre enn den objektive fordi konsumenten tror hun er mindre eksponert for den objektive risiko enn gjennomsnittet.

For Mishan synes persepsjonsproblemer og begrepet objektiv risiko uinteressante. For siden konsumenter pr. definisjon forholder seg til en subjektiv risiko, er det deres uttrykte betalingsvilje for å redusere denne som er avgjørende i nytte-kostnads analyser. Det spiller mao. ingen rolle

om de betaler for en illusjon eller en realitet, så lenge de faktisk uttrykker vilje til å betale!

Persepsjonsproblemet kan neppe avvises ut fra en slik streng velferdsteoretisk fortolkning. Som Mooney (1978) med rette påpeker avhenger nemlig bruken av den subjektive risikopersepsjon av for det første at konsumenten *er* best og for det andre *tror* hun er best til å foreta de nødvendige vurderinger. Videre at hun *kan* og ikke minst *ønsker* å gjøre dette. Dersom disse fire krav ikke er tilfredsstilt, blir forutsetningen om at konsumentene er de rette til å vurdere risiko tvilsom.

Et annet metodisk problem ved hypotetiske spørsmål er at måten spørsmål stilles på («the framing of questions») påvirker respondenten (Tversky og Kahneman 1981, Loomes 1988). Rhoads (1980) refererer til en studie av Jan Acton om betalingsviljen for en gitt sannsynlighetsreduksjon for å dø av hjertesvikt, der to ulike spørsmål ble stilt. Spørsmålet som vektla nytten for den enkelte respondent ga nesten dobbel så høy betalingsvilje som det spørsmål som vektla nytten for samfunnet!

Betalingsviljen for risikoreduserende tiltak har i prinsippet to *fordelingsaspekter* avhengig av om tiltaket kommer et tverrsnitt av befolkningen til gode eller en bestemt inntektsgruppe. Dersom det kommer et tverrsnitt til gode, vil den samlede betalingsviljen være større dess større inntektsulikhet, fordi risikoreduksjon normalt antas å være et inntektselastisk gode med Engel-elastisitet større enn 1. Et samfunn med skjeiv inntektsfordeling vil således etterspørre mer risikoreduksjon enn et med lik inntektsfordeling. Dersom tiltaket kommer en bestemt inntektsgruppe til gode, vil betalingsviljen avhenge av deres betalings*evne*.

2.3 Implisitt verdsetting

Den implisitte metoden tar utgangspunkt i det offentliges ressursallokering og beregner samfunnets verdsetting av menneskeliv i ulike sektorer ut fra hvor mye ressurser som faktisk blir allokert til prosjekt som utelukkende har til hensikt å redde menneskeliv. Divideres ressursinnsatsen med antall menneskeliv som er reddet, framkommer den minimale implisitte pris på menneskeliv. Dersom f. eks. det offentlige velger å bruke 10 mill. kr på et prosjekt som forventes å ville redde 5 mennesker, er den implisitte verdsetting av menneskeliv *minst* 2 mill. kr. Dersom prosjektet derimot ikke gjennomføres «fordi det er for dyrt», er menneskeliv impli-

sitt verdsatt lavere enn 2 mill. kr. Vi kan altså ikke utlede noen kritisk verdi; kun *minimal verdi* dersom prosjektet gjennomføres og *maksimal verdi* dersom det ikke gjennomføres.

Det er viktig her å foreta et skille mellom kjente og statistiske liv. Når det gjelder kjente, navngitte i fare, brukes sjelden *økonomiske* argumenter for å avblåse livreddingsaksjoner. En annen årsak til at store ressurser settes inn på å redde konkrete liv er at *uten* denne innsatsen går vedkommende mot en *sikker* død. Andre livreddingstiltak har til hensikt å redusere *sannsynligheten* for å dø.

Samfunnets verdsetting av statistiske menneskeliv kan beregnes på ulike måter, men hovedtilnærmingen er studier av det offentliges ressursallokering og lovgivning. Vi har en rekke lover og påbud, spesielt innenfor trafikksektoren, som har til hensikt å redusere ulykker. Svært ofte innebærer dette kostnader både til selve sikringstiltaket og til kontroll med at loven følges. I tillegg kommer de direkte kostnader betalt av konsumentene, f. eks. sikkerhetsbelter. Dividerer vi de totale kostnader ved tiltaket på antall menneskeliv som er beregnet reddet, får vi fram implisitt pris på menneskliv.

Et mye sitert eksempel er innføringen av et påbud om overbygde traktorer i England (Mooney 1977). Total økning i de engelske bønders traktorkostnader ble £ 20 mill., men dette ville redde 200 menneskeliv. Implisitt verdi på en bonde ble dermed £ 100.000. Graham og Vaupel (1981) sin oversikt over 57 studier av implisitt verdsetting av menneskeliv i USA viser enorme forskjeller både innenfor og mellom sektorer.

En alternativ tilnærming er å se på domsavsigelser vedrørende ulykker. Samfunnets verdsetting av menneskeliv oppfattes her å gjenspeiles i kompensasjonen til de etterlatte. I Storbritannia gjelder prinsippet om full kompensasjon ex post, dvs. hva de etterlatte hadde hatt om forsørgeren hadde levd resten av sitt yrkesaktive liv. I USA er det en tradisjon om at den ansvarlige skal straffes, dvs. at de etterlatte skal ha mer enn full kompensasjon.

Det er ikke bare livet, men også våre kroppsdeler som har implisitt verdi i penger. Rikstrygdeverkets invaliditetstabell er interessant lesing i så henseende. Her gis invaliditetsgrad på sykdomsarter og tap av lemmer. F.eks. innebærer total skalpering en invaliditetsgrad på 15−25 %, mens tap av begge ytterører tilsvarer 25 % invaliditet. (Van Gogh ble 15 % invalid da han skar av seg sitt ene ytreøret).

2.3.1 Diskusjon

Human capital og betalingsviljemetodene er i prinsippet nyttekostnads analyser med ulik måte å verdsette menneskeliv på. En human capital basert kalkyle legger menneskers produksjonsverdi på nyttesiden, mens en betalingsviljekalkyle legger menneskenes betalingsvilje for å redusere risikoen for å være blant de uheldige på nyttesiden. Den implisitte metoden derimot har ikke til hensikt å konstruere noen nytteside; den er en kostnads-effektivitetsanalyse.

Metoden har praktiske fordeler i det den synliggjør konsekvensene av beslutninger som er fattet/ikke fattet. Den har således viktige policy implikasjoner. Metoden beregner samfunnets verdsetting av menneskeliv ut fra hvor mye ressurser som faktisk brukes til ulike risikoreduserende tiltak. Det viktigste bidraget fra den implisitte metoden er at den viser i hvilke sektorer samfunnet bør prioritere ressurser til livreddingstiltak og eventuelt i hvilke sektorer livreddingstiltak bør reduseres (vel og merke under forutsetning av at maksimal nytte kriteriet gjelder). Uavhengig av hvor store de totale ressurser er som allokeres til livreddingstiltak, skal tildelingen skje slik at grensekostnadene ved å redde et menneskeliv er like store på alle områder.

2.4 Oppsummering

Human capital metoden samsvarer ikke med noen av de tre allokerings-kriterier. Verdien av et menneskeliv avhenger bl. a. av dets inntekt og er følgelig ikke i samsvar med likhetskriteriet. Metoden pretenderer ikke å redde flest mulig menneskeliv innenfor et gitt budsjett, men derimot å beregne hvorvidt et menneske med gitte karakterisika er lønnsomt å redde. Metoden er heller ikke opptatt av individers preferanser.

Betalingsviljemetoden samsvarer ikke med likhetskriteriet fordi ens uttrykte betalingsvilje avhenger av ens betalings*evne*. Metoden er ikke opptatt av å redde flest mulig til gitt budsjett, men den er genuint opptatt av å følge det tredje allokeringskriteriet, nemlig at individers preferanser skal gjelde.

Den implisitte metoden tar overhodet ikke hensyn til inntekt og samsvarer følgelig med likhetskriteriet. Metoden beregner verdien på menneskeliv innenfor ulike sektorer ut fra tidligere politiske beslutninger og vil som sådan ikke si noe om framtidig allokering. Derimot vil metoden kunne

brukes til å beregne kostnadene pr. reddet menneskeliv innenfor ulike prosjekt og dermed vise hvor det er billigst å redde menneskeliv. Implikasjonen av metoden er følgelig i samsvar med kriteriet om å redde flest mulig innenfor gitt budsjett. Metoden tar ikke hensyn til individers preferanser.

3. Noen årsaker til forskjeller i verdien på menneskeliv

Den velferdsøkonomiske fortolkningen av de observerte forskjeller i verdsetting av menneskeliv vil være at betalingsviljen avhenger av såvel risiko*nivå* som størrelsen på risiko*endringen*, dessuten at konsumentene har preferanser for *måter* de ønsker å unngå å dø på, samt at de har ulik betalings*evne* for risikoreduserende produkter. La oss først se på *inntektsfordelingen*. Usher (1985) tar utgangspunkt i sikkerhet som et «personlig gode» med en stigende grensekostnadskurve for den enkelte; dess større grad av risikoreduksjon, dess høyere grensekostnader. Gitt samme risikoaversjon vil rike følgelig etterspørre mer risikoreduksjon enn fattige — sikre biler er ofte dyre biler. En høy implisitt pris på menneskeliv kan derfor være et uttrykk for at tiltaket reduserer dødsrisikoen for rike.

Mooney (1977) refererer til en studie som fant at mellom 80% og 95% ville dø plutselig og syntes å ønske dette for enhver pris. Respondentene til Jones-Lee et al (1985) ble spurt om hvilken dødsårsak de ville prioritere å redusere av trafikkulykker, hjertefeil og kreft. Hele 76 % ville redusere kreft. Altså, vi foretrekker å dø fort og greit framfor å pines ihjel. I henhold til konsumentsuverenitet-kriteriet vil det følgelig være rasjonelt å bruke mer ressurser pr. reddet menneskeliv på de dødsmåter vi aller helst vil unngå.

Det synes også å være et *angstelement* i å overlate kontrollen med ens liv til andre, selv om risikoen for å dø er mindre enn når en selv (tror en) har kontrollen; flyskrekk er mer utbredt enn bilskrekk! Angstelementet er også knyttet til ulykkers omfang — folk synes å ha sterk grad av «katastrofeaversjon». Betalingsviljen for en marginal reduksjon av risikoen for en katastrofe gir høyere verdi på menneskeliv enn hva betalingsviljen i sektorer med større sannsynlighet for småulykker gir.

Mens dødsmåten og angstelementet er bevisst og reflekterer *rasjonelle preferanser*, er persepsjonsproblemet et uttrykk for at en enten ikke vet at en utsetter seg for risiko eller at en ikke klarer å forholde seg til innholdet

i små risikoendringer. I den grad den subjektive risiko avviker fra den objektive av slike grunner, vil ens atferd reflektere *«irrasjonelle preferanser»*.

Så langt har vi diskutert psykologiske og økonomiske variable – rasjonelle preferanser, persepsjonsproblemet og inntektsfordeling – som forklarer de ulikheter som framkommer gjennom betalingsviljemetoden. Imidlertid vil det kunne være fruktbart å søke forklaringer på skjeivallokeringer i offentlig ressursallokering på dette felt i mer *statsvitenskapelig* teori. For politikere gir det større avkastning å symbolisere gjennom flaggsaker at de redder noen lidende enn å si at «en alternativ allokering vil kunne redde flere». En annen forklaring er media. Oppslag som «Han døde i hjertekø» har utvilsomt medvirket til de økte ressurser til hjerteoperasjoner.

Et tredje sett av forklaringer dreier seg om *etikk*. Dette vil kunne bidra til å forklare for det første en høyere verdsetting av yngre framfor eldre og for det andre høyere verdsetting av kjente liv framfor statistiske. I en nylig publisert undersøkelse (Lewis og Charny 1989), basert på et tilfeldig utvalg av befolkningen i Cardiff, spurte man «hvilken av to pasienter vil du behandle når bare alderen er forskjellig og du ikke kan behandle begge?» I valget mellom en femåring og en syttiåring prioriterte 94 % femåringen, 5 % ville ikke svare og kun 1 % prioriterte den gamle. Mer oppsiktsvekkende var at hele 70 % syntes det var lett å prioritere den yngste. Vi skal legge merke til at denne prioritering av barn framfor gamle *ikke* er teoretisk begrunnet men reflekterer en utbredt sosial etikk om at barns liv verdsettes høyere enn eldres liv.

En gjennomgående forutsetning i litteraturen om verdsetting av menneskeliv er som nevnt at temaet dreier seg om statistiske liv – ikke navngitte personer. Navngitte personer er begunstiget med vår «sosiale samvittighet» for å prioritere folk som går mot en sikker død uten behandling framfor å redusere risikoen for at alle andre skal dø. Dette berører vårt syn på den relative allokering mellom kurative og preventive tiltak i retning av høyere betalingsvilje for kurative tiltak.

4. Hvilke allokeringskriterier bør følges?

Innledningsvis etablerte vi tre allokeringskriterier som metodene for verdsetting av menneskeliv er diskutert i relasjon til. I dette avsnittet skal vi se

hvorvidt det er konflikt mellom disse, og i tilfelle hvilke(t) som bør overordnes. Mellom likhetskriteriet og maksimal nytte kriteriet ligger det ingen motsetning, fordi ens inntekt ikke er bestemmende for hvorvidt en bør reddes. Konsumentsuverenitet-kriteriet er derimot mer problematisk, da det står i konflikt med både likhetskriteriet og maksimal nytte kriteriet.

4.1 Likhet og/eller konsumentsuverenitet?

Som nevnt tidligere uttrykker betalingsviljen ens betalingsevne og ens preferanser, som varierer avhengig av ens inntekt og risikoaversjon. Høy betalingsvilje for risikoreduksjon skyldes høy inntekt og/eller høy risiko- aversjon. Lav betalingsvilje skyldes tilsvarende lav inntekt og/eller lav risikoaversjon. Dersom allokering av offentlige ressurser til livredding kun skal følge betalingsviljemetoden, som Mishan argumenterer for, vil det i realiteten si at mer penger bør brukes på å redde en rik enn en fattig. Som påpekt av Usher (1985) vil en slik allokeringsregel *øke* forskjellene i levestandard. Studien av Marin og Psacharopoulos (1982) illustrerer poenget, nemlig at verdien på en «nonmanual worker» er 3,5 ganger større enn en «manual worker». Nytte-kostnads implikasjonen er at til identisk kostnad pr. reddet menneskeliv vil det være samfunnsøkonomisk lønnsomt å prioritere «nonmanual workers».

Med den allment aksepterte holdning om at ens betalingsevne ikke skal påvirke ens tilgjengelighet i helsesektoren, er det vanskelig å finne aksept for at betalingsevnen skal gjelde i *andre* sektorer som dreier seg om liv og død. Det som da gjør det problematisk å bruke betalingsviljetilnærminga er at det er metodisk vanskelig å *trekke ut* den del av betalingsviljesigna- lene som skyldes ulik betalingsevne fra den som skyldes ulike preferanser.

At konsumenter har *ulike risikopreferanser* er åpenbart. I litteraturen skiller man mellom «risk lover» og «risk averter». Risiko «pirrer» risiko- elskerne, som er villige til å betale for å påføre seg en viss grad av risiko. Så lenge denne type atferd ikke har eksterne effekter, er det ingen økonomiske argumenter mot å tillate «risikopirrende konsum». Det som derimot kan virke urimelig er at folk med risikoaversjon blir avkrevd skatter for å redusere andres dødsrisiko ved risikopirrende konsum.

Offentlige prosjekter som er ment å øke innbyggernes sannsynlighet for å overleve har som påpekt av Dehez og Dreze (1982) samme egenskaper som et *kollektivt gode*. Et kollektivt gode kan imidlertid være begrenset,

eller «lokalt», i den forstand at det kun er gjennom et bestemt konsum eller bosted at man drar nytte av en offentlig finansiert risikoreduksjon. Ens sannsynlighet for å overleve er også i høy grad bestemt av konsumet av *private goder.*

Mishan (1971) deler risiko i en *frivillig* og en *ufrivillig* del. Frivillig risiko er knyttet til konsumet av et gode, som konsumenten er kjent med og dermed aksepterer i kraft av sitt konsum. Risikoen vil være tatt hensyn til gjennom ens etterspørsel, f.eks. røyking og fjellklatring. Etterspørselen etter godet vil følgelig øke dersom risikoen ved det kan reduseres. Ufrivillig risiko derimot er påtvunget.

Vi skal her tenke oss tre typer risiki, der de to første er ufrivillige, mens den tredje er frivillig. For det første har vi en *kollektiv risiko*, r_k, som er mindre dess mer av det kollektive godet risikoreduksjon som samfunnet allokerer ressurser på. Slike goder kan være vaksiner og sikkerhetsstandarder i sin alminnelighet. For det andre har vi en *lokal kollektiv risiko*, r_l, som varierer regionalt. Trafikksikkerhet og tilgang på helsetjenester er slike goder med utpregede regionale variasjoner. For det tredje har vi en frivillig risiko som er knyttet til variasjoner i *individers atferd*, r_p. Summen av disse tre risiki vil være ens totale dødsrisiko, r. Det er rimelig å anta at de tre sannsynlighetene er *uavhengige* av hverandre samt at summen normalt vil være betydelig mindre enn 1.

$$r_k + r_l + r_p = r$$

Denne oppsplittingen synliggjør to dilemma. Det første gjelder ens holdning til frivillig risiko. Det andre dreier seg om regionale variasjoner i risiko og drøftes i neste avsnitt. Til gitt budsjett vil man redde flere mennesker dess lavere r_p er. Det kan derfor argumenteres for at fellesressurser brukes til å redusere r_k og r_l, mens ressurser til å redusere r_p finansieres gjennom avgifter på risikabelt konsum og tvungen forsikring. F. eks. betaler røykerne avgift for å finansiere et helsevesen som kurerer dem.

Oppsummert er det to hovedproblem ved konsumentsuverenitet kriteriet i forhold til likhetskriteriet. For det første at betalingsevnen er ulikt fordelt og for det andre at graden av risikoaversjon varierer.

4.2 Maksimal nytte og/eller konsumentsuverenitet

Det tredje hovedproblem ved konsumentsuverenitet-kriteriet er allerede berørt flere steder, nemlig *persepsjonsproblemet*. En paternalistisk løsning er nærliggende, nemlig å overføre beslutningsmyndighet til en instans som vil forholde seg rasjonelt til de objektive risiki. Om dette skriver Mooney: «Provided — and it is a major proviso — that consumers are prepared to accept that there are certain areas where 'acting in their own interests' it is better to hand over certain decision-making, or in this particular case knowledge on which to base decision-making to the state (since by so doing they are likely to increase their **ex post** utility), then there is nothing to prevent the state in stepping in to correct the misestimates of probabilities by individuals.» (Mooney 1977)

På et område der likhetskriteriet er så allment akseptert kan det være fruktbart å gjøre som bl.a. Usher (1985) og tenke innenfor John Rawls' rettferdighetsteori. Utgangspunktet er hvordan fordelingsreglene i et samfunn ville utformes av dets framtidige medlemmer «behind a veil of ignorance»; dvs. dersom disse ikke vet hvilken plass i samfunnet de selv vil få og de har lik sannsynlighet for i hvem sine sko de vil havne. Spørsmålet inviterer til en «maksimin rettferdighet», dvs. en som maksimerer nytten til den minst begunstigede.

Linnerooth (1982) tolker Rawls' maksimin-rettferdighet slik at det impliserer en favorisering av de som til enhver tid har størst dødsrisiko, mao. kjente liv framfor statistiske. Implikasjonen er at mer ressurser vil bli brukt på å redde et kjent liv enn et statistisk. En allokering i henhold til Rawls' rettferdighet vil følgelig ikke samsvare med en maksimal nytte allokering.

Denne fortolkningen er ikke overbevisende. Det synes nemlig merkelig å anta at de framtidige samfunnsmedlemmer som i prinsippet vil inneha begge rollene — både som statistiske og kjente liv — vil etablere fordelingsregler som favoriserer kjente liv, og det på et tidspunkt da disse kjente liv pr. definisjon er statistiske. Nettopp derfor er det vanskelig å tenke seg at fordelingsreglene for allokering av ressurser til livredding og dødsrisikoreduksjon vil kunne ligge særlig langt fra et som innebærer like kostnader pr. reddet liv, fordi det er dette som gir det framtidige samfunnsmedlem et lengst mulig forventet liv.

La oss tenke oss at befolkningen erkjenner problemene med å signalisere sine preferanser gjennom en serie partielle betalingsviljemålinger og

i stedet stilles overfor ett spørsmål med følgende innhold, presentert av f.eks. sosial-, samferdsel- og kommunalministrene i fellesskap: Ønsker du at vi skal allokere våre totale livreddingsressurser slik at flest mulig menneskeliv kan reddes, eller har du aversjon mot spesielle måter å dø på som gjør at mer penger skal allokeres der, med den konsekvens at færre menneskeliv kan reddes og din samlede dødsrisiko dermed øker?

Igjen synes det vanskelig å tenke seg at allokeringen vil kunne ligge særlig langt fra et som innebærer like kostnader pr. reddet liv, mao. maksimal nytte kriteriet. Legg merke til at det ikke er noe i veien for at man samtidig uttrykker en preferanse for å bruke mer penger på å redde «known lives in present danger». Innenfor et gitt budsjett vil naturligvis en slik prioritering av kjente liv framfor statistiske innebære at færre liv kan reddes. Weale (1979) argumenterer mot denne skjeivallokeringen i favør av kjente liv:

«There is nothing right in itself or good on balance about discriminating in favour of known lives in present danger, if greater efficiency can be achieved otherwise» (Weale 1979).

Dersom efficiency defineres som en allokering der det ikke er mulig å redde flere menneskeliv uten tilførsel av mer ressurser, har Weale naturligvis rett. Imidlertid synes han å trekke poenget for langt. Dersom vår «sosiale samvittighet», eller det Culyer (1976) kaller «caring externality», foretrekker å redde «known lives in present danger» til en høyere grensekostnad enn statistiske liv, er dette *både* rett i seg selv og «good on balance».

Til slutt en utfordring og ett dilemma: *Utfordringen* ligger i «timing» og beslutningsnivå. Den faktiske allokeringen på spesifikke program må gjøres før ofrene blir ofre, dvs. *ex ante* før statistiske liv blir kjente navngitte. Det er nemlig ikke bare samfunnet som uttrykker lavere betalingsvilje for å redde statistiske liv, men de som blir *ofre* for en sykdom uttrykker også lavere betalingsvilje for å kurere den *før* de vet at de er blant ofrene. Da vil deres private betalingsvilje for en kort livsforlengelse reflektere at de selv ikke har særlig annet å bruke penger på. Samfunnet derimot kunne brukt de samme ressursene til å redde flere. Linnerooth (1982) har et godt poeng:

«the important point is that from the utilitarian's perspective, in contrast to the welfare economist's, the differential treatment of risk populations cannot bu justified by reference to an individual's *own* preferences for

reducing high risks to himself, but can only be justified by reference to a desire on the part of others to respond disproportionately toward the more threatening situation.» (Linneroth 1982)

«Timing» berører beslutningsnivå i den forstand at «tragiske valg» bør foretas på høyt plan *før* helsepersonell står ansikt til ansikt med pasienten. Tragiske valg er i henhold til Calabresi og Bobbitt (1978) beslutninger om hvem som skal få tilgang på begrensede livreddingsressurser.

Dilemmaet dreier seg om hvilken *type* likhet som skal gjelde. Så langt har vi behandlet likhet synonymt med lik kostnad pr. reddet menneskeliv uavhengig av inntekt. Et alternativ er *likhet i ufrivillig risiko*. I forrige avsnitt så vi at r_i vil variere. I den grad det eksisterer regionale variasjoner i å redusere kostnader ved f. eks. trafikkulykker − topografiske forskjeller som gjør det dyrt å rette ut farlige svinger i enkelte regioner − vil det være dyrere å redde et statistisk liv i slike regioner enn i andre regioner. Dersom allokeringskriteriet er samme kostnad pr. reddet liv, vil folk i slike regioner utsettes for større dødsrisiko. Maksimal nytte kriteriet tilsier at r_i skal variere, mens en «risk equity» norm tilsier samme eksponering for ufrivillig risiko.

Keeney (1982) er opptatt av at det er en sosial preferanse for risikospredning, at det intuitivt virker mer rettferdig at risiko spres enn at den er konsentrert. Denne rettferdighetsnormen eller preferansen for risikospredning reflekterer individenes oppfatning av grad av risiko en selv eksponeres for, nemlig avtakende grensenytte av etterfølgende risikoreduksjoner.

Oppsummert; vi har argumentert for at det kan være rasjonelt for individer å foretrekke en paternalistisk løsning som allokerer i henhold til maksimal nytte kriteriet. Imidlertid er det ikke noe i veien for at «pateren modereres» til å ta hensyn til befolkningens preferanser for dødsmåter og deres sosiale samvittighet for å prioritere kjente liv i fare. Slike preferanser kan imidlertid komme til syne på andre måter enn å måle betalingsvilje.

5. Oppsummering

Det som er så ufordragelig med økonomer er at vi på død og liv skal gjøre «tragiske valg» eksplisitte. Leger kan skjerme seg fra tragiske valg ved kun å forholde seg til pasienten foran seg i stedet for til hele køa.

Politikere lar helst være å ta stilling til tragiske valg av redsel for at neste valg da kan bli tragisk for dem selv!

Rhoads (1980) mener det er demoraliserende når samfunnet offentlig setter en verdi på menneskeliv og argumenterer for at temaet og vurderingene bør være mindre åpen. Er det mer moralsk å sette en verdi i det skjulte? At et tema er følsomt er i seg selv ikke et argument mot å beskjeftige seg med det:

«The claim that something is inherently complicated does not excuse us from explicitly stating the principles which should be brought to bear upon the decision.» (Marin 1983)

Formålet med artikkelen var å diskutere de ulike metoder for verdsetting av menneskeliv med eksplisitt referanse til tre kriterier for allokering av «livreddingsbudsjettet». Den teoretiske diskusjonen synes nemlig ofte å overse det faktum at dette budsjettet er underlagt eksplisitte fordelingsmål.

Så lenge vi ikke snakker om pris på *konkrete liv*, burde ikke selve problemstillingen være uetisk. Dog synes *human-capital metoden* å ha tvilsomme etiske implikasjoner. Metoden samsvarer ikke med noen av de tre allokeringskriterier vi introduserte innledningsvis og burde derfor vært avlivet som metode for å beregne verdien av menneskeliv. Likevel sniker den seg inn og brukes opportunistisk av helsepersonell som er opptatt av å bevise hvor samfunnsnyttig nettopp deres geskjeft er.

Betalingsviljemetoden beregner verdien av menneskeliv på grunnlag av hvordan folk faktisk forholder seg til risikoen for å miste det. Metoden er følgelig i samme gate som konsumentsuverenitetskriteriet. Men, betalingsviljen reflekterer ens betalings*evne* og er således i motstrid til likhetskriteriet. Metoden samsvarer heller ikke med maksimal-nytte-kriteriet, fordi den ikke pretenderer å beregne noen endelig eller gjennomsnittlig verdi på menneskeliv som budsjettet kan allokeres i henhold til.

«This is simply because the crucial dv/dr ratio varies, *inter alia*, with the level of risk r, with the magnitude of dr, with the interpretation placed on dr and also, of course, with the specific sort of death envisaged.» (Mishan 1985)

Siden «verdien på menneskeliv», V, som utledes fra dv/dr vil variere fra null til uendelig, følger det at *V må beregnes for hvert enkelt prosjekt*. Mishan (1985) sin implikasjon er at dette krever partielle nytte-kostnads analyser av de prosjekt som påvirker folks dødsrisiko, for å finne fram til

hvorvidt betalingsviljen for å redde menneskeliv er større enn kostnadene.

Gitt de alvorlige metodiske problemer forbundet med å måle konsumentenes sanne preferanser vil slike målinger mildt sagt måtte bli omfattende. For egen del tviler jeg sterkt på at den samfunnsøkonomiske nytten av slike studier oppveier de totale kostnader! Bl.a. derfor har jeg et mer pragmatisk syn på spørsmålet om å bruke en gjennomsnittlig verdi på menneskeliv. Det vil dessuten samsvare med både likhets- og maksimalnytte-kriteriet.

Den *implisitte metoden* er egentlig ikke en metode for verdsetting av menneskeliv men en metode som beregner *kostnadene pr. reddet menneskeliv* i ulike sektorer. Metoden viser hvor det er billig og hvor det er dyrt å redde et ekstra liv. Den kan dermed anvise mulige forbedringer dersom målet er å redde flest mulig innenfor et gitt budsjett. Dens åpenbare fordel ligger på policy-planet, i det den «will allow decision-makers to be more consistent in their decision-making and at the same time more efficient» (Mooney 1986). Kan vi forlange stort mer?

Vi kan forlange at metoden i tillegg skal si noe om budsjettets størrelse, og det gjør den ikke. Men dersom samfunnet av etiske hensyn *ikke* ønsker å utlede noen gjennomsnittlig verdi på innbyggerne (en ikke urimelig forutsetning?), vil anvisninger fra den implisitte metoden være et meget godt alternativ.

REFERANSER

Bergstrom, T.C. (1982): «When is a man's life worth more than his human capital?» i Jones-Lee, M.W. (ed): *The Value of Life and Safety*, North-Holland, Amsterdam

Blomquist, G. (1979): «Value of Life Saving: Implications of Consumption Activity», *Journal of Political Economy* 87 540–557

Blomquist, G. (1982): Estimating the value of life and safety: Recent development, i Jones-Lee, M.W. (ed) op cit

Brody, W. (1975): «Economic *Value of a Housewife*». Research and Statistics Note no. 9. Washington: U.S. Department of Health, Education and Welfare

Broome, J. (1978): «Trying to Value a Human Life», *Journal of Public Economics* 9 91–100

Calabresi, G. og Bobbitt, P. (1978): *Tragic Choices*, New York, Norton

Cooper, B. og Rice, D.(1976): «The Economic Costs of Illness Revisited», *Social Security Bulletin* 39 21–36

Culyer, A.J. (1976): *Need and the National Health Service*, Martin Robertson, Oxford

Dardis, R. (1980): «The value of Life: New Evidence from the Marketplace», *American Economic Review 70* 1077–82

Dawson, R.F.F. (1967): «Cost of Road Accidents in Great Britain», i Cooper, M.H. og Culyer, A.J. (eds.) (1973): *Health Economics*

Dehez, P. og Dreze, J.H. (1982): «State-dependent utility, the demand for insurance and the value of safety», i Jones-Lee, M.W. (ed) op cit

Dublin, L.I. og Lotka, A.J. (1946): *The Money Value of Man*, New York, Ronald Press

Graham og Vaupel (1981): «Value of a Life: What Difference Does It Make?» *Risk Analysis 1* 89–95

Hammerton, M. (1982): «Comments on Estimating the value of life and safety: Recent developments», i Jones-Lee, M.W. (ed) op cit

Hanlon, J.J. (1969): *Principles of Public Health Administration*, St Louis, C.V. Mosby

Hodgson, T. and Meiners, M. (1979) «Guidelines for Cost of Illness studies in the Public Health Service», Washington: Public Health Service Task Force on Cost of Illness Studies

Hodgson, T. and Meiners, M. (1982): «Cost of Illness Methodology: A Guide to Current Practices and Procedures», *Milbank Memorial Fund Quarterly/Health and Society 60* 429–62

Jones-Lee, M.W. (1974): «The value of changes in the probabilty of death or injury», *Journal of Political Economy 82* 835–849

Jones-Lee, M.W. (1976): *The Value of Life: An Economic Analysis*, Martin Robertson, London

Jones-Lee, M.W. (1979): «Trying to Value a Life: Why Broome does not sweep clean», *Journal of Public Economics 12* 249–56

Jones-Lee, M.W. (1982): «Preface» i Jones-Lee, M.W. (ed) op cit

Jones-Lee, M.W. (Ed) (1982): *The Value of Life and Safety*, North-Holland, Amsterdam

Jones-Lee, M.H., Hammerton, M., Philips, P.R. (1985): «Tha Value of Safety: Results from a National Sample Survey», *The Economic Journal 95* 49–72

Keeney, R.L. (1982): «Evaluating mortality risks from an organisational perspective», i Jones-Lee, M.W. (ed) op cit

Klarman, H. (1982): «The Road to Cost-benefit Analysis», *Milbank Memorial Fund Quarterly/Health and Society 60* 585–603

Lewis, P.A. og Charny, M. (1989): «Which of two individuals do you treat when only their ages are different and you can't treat both?» *Journal of Medical Ethics 15* 28–32

Linnerooth, J. (1982): «Murdering statistical lives...?» i Jones-Lee, M.W. (ed) op cit

Loomes, G. (1988): «Disparities between health state measures: an explanation and some implications», Paper presented to the UK HESG meeting, Brunel

Marin, A. og Psacharopoulos, G. (1982): «The Reward for Risk in the Labor Market:» Evidence from the United Kingdom and a Reconciliation with Other Studies, *Journal of Political Economy 90* 827–853

Marin, A. (1983): «Your money or your life» *The three Banks review 138* 20–37

McGuire, A. og Mooney, G. (1985): «Back to life: resurrecting the value of life debate», Discussion Paper No. 5, Health Economics Research Unit, University of Aberdeen

Melinek, S.J. (1974): «A method of evaluating human life for economic purposes», *Accident Analysis & Prevention 6* 103–114

Mishan, E.J. (1971): «Evaluation of Life and Limb: A Theoretical Approach», *Journal of Political Economy 79* 687–705

Mishan, E.J. (1981): «The value of trying to value a life», *Journal of Public Economics 15* 133–7

Mishan, E.J. (1985): «Consistency in the Valuation of Life: A Wild Goose Chase», In Paul, E. et al: *Ethics and Economics*, Blackwell, Oxford

Mooney, G. (1977): *The Valuation of Human Life*, Macmillian Press, London

Mooney, G. (1978): «Human Life and Suffering», i Pearce, D.W. (ed) *The Valuation of Social Costs*, London: George Allen and Unwin Ltd

Mooney, G. (1986): *Economics, Medicine and Health Care*, Wheatsheaf, Brighton

Reynolds, D.J. (1956): «The Cost of Road Accidents», *Journal of Royal Statistical Society 119* 395–408

Rhoads, S.E. (1980): «How Much Should We Spend to Save a Life», i Rhoads, S.E. (ed.) (1980): *Valuing Life: Public Policy Dilemmas.* Westview Press, Boulder, Colorado

Rice, D. and Cooper, B. (1967): «The Economic Value of Human Life», *American Journal of Public Health 57* 1954–66

Robinson, J.C. (1986) «Philosophical Origins of the Economic Valuation of Life», *Milbank Memorial Fund Quarterly/Health and Society 64* 133–155

Robinson, J.C. (1986b) «Hazard Pay in Unsafe Jobs: Theory, Evidence, and Policy Implications», *Milbank Memorial Fund Quarterly/Health and Society 64* 650–677

Scitovsky, A. (1982): «Estimating the Direct Cost of Illness», *Milbank Memorial Fund Quarterly/Health and Society 60* 463–91

Smith, V.K. og Devousges, W.H. (1987) «An Empirical Analysis of the Economic Value of Risk Changes», *Journal of Political Economy* Vol 98, 89–114

Thaler, R. (1982): «Precommitment and the value of a life», i Jones-Lee, M.W. (ed) op cit

Thaler, R.H. og Rosen, S. (1976): «The Value of Saving a Life: Evidence from the Labor Market», i Terleckyj, N.E.: *Household Production and Consumption*, Columbia University Press (N.B.E.R.): New York

Tversky, A. og Kahneman, D. (1981): «The framing of decisions and the psychology of choice», *Science 211*, 453–8

Usher, D. (1985): «The Value of Life for Decision Making in the Public Sector», In Paul, E.F. et al op cit

Viscusi, W.K. (1978): «Labor Market Valuations of Life and Limb: Empirical Evidence and Policy», *Public Policy 26*, 359–386

Weale, A. (1979): «Statistical life and the principle of maximum benefit», *Journal of Medical Ethics 5*, 185–95

Weisbrod, B. A. (1961): «Economics of Public Health, Philadelphia:» *University of Pensylvania Press*

Williams, A. (1985): «Economics of coronary artery bypass grafting», *British Medical Journal 291*, 326–9

Wiseman, J. (1963): «Cost-Benefit Analysis and Health Service Policy», *Scottish Journal of Political Economy 10* 128–145
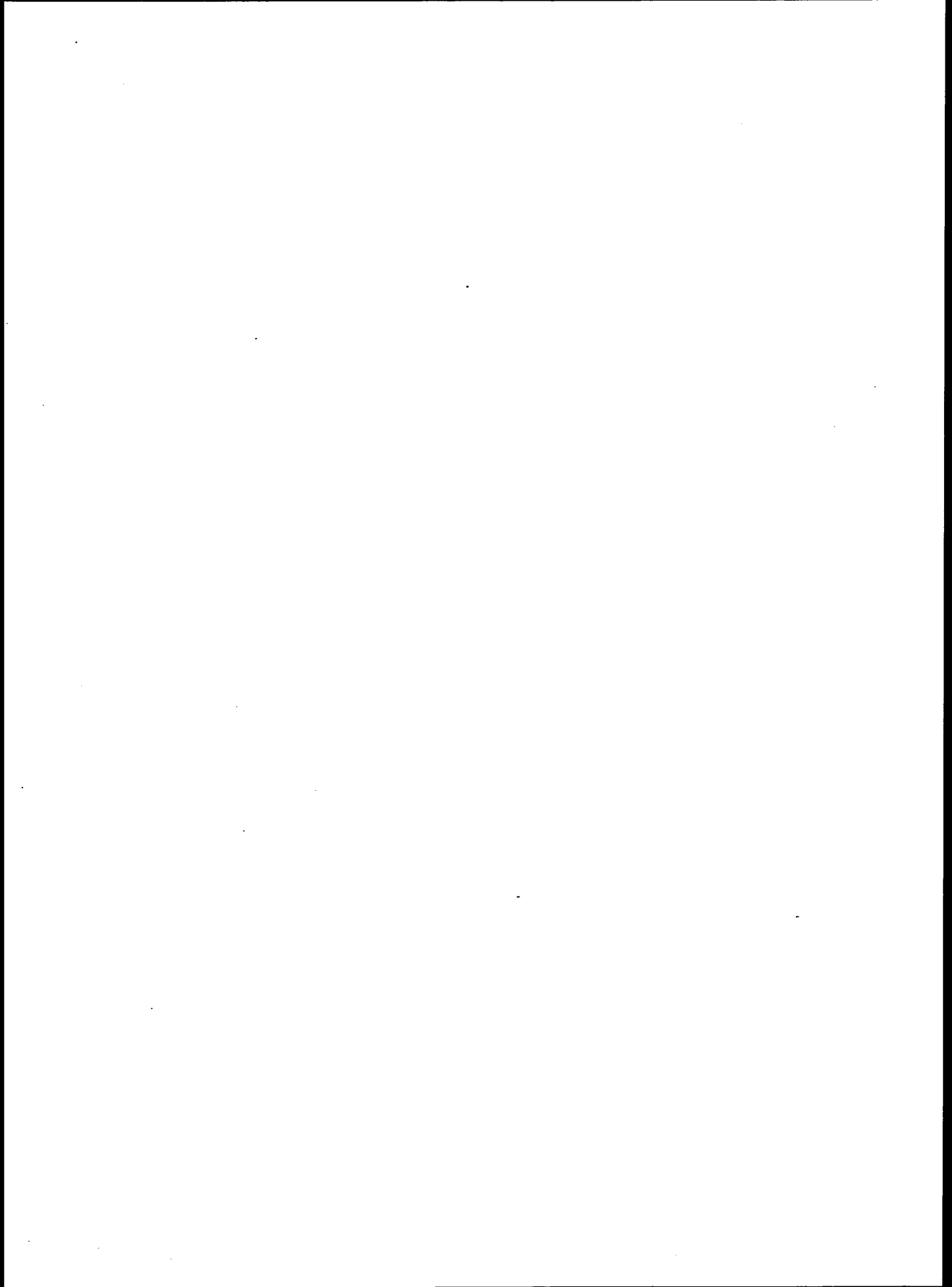
Paper 2
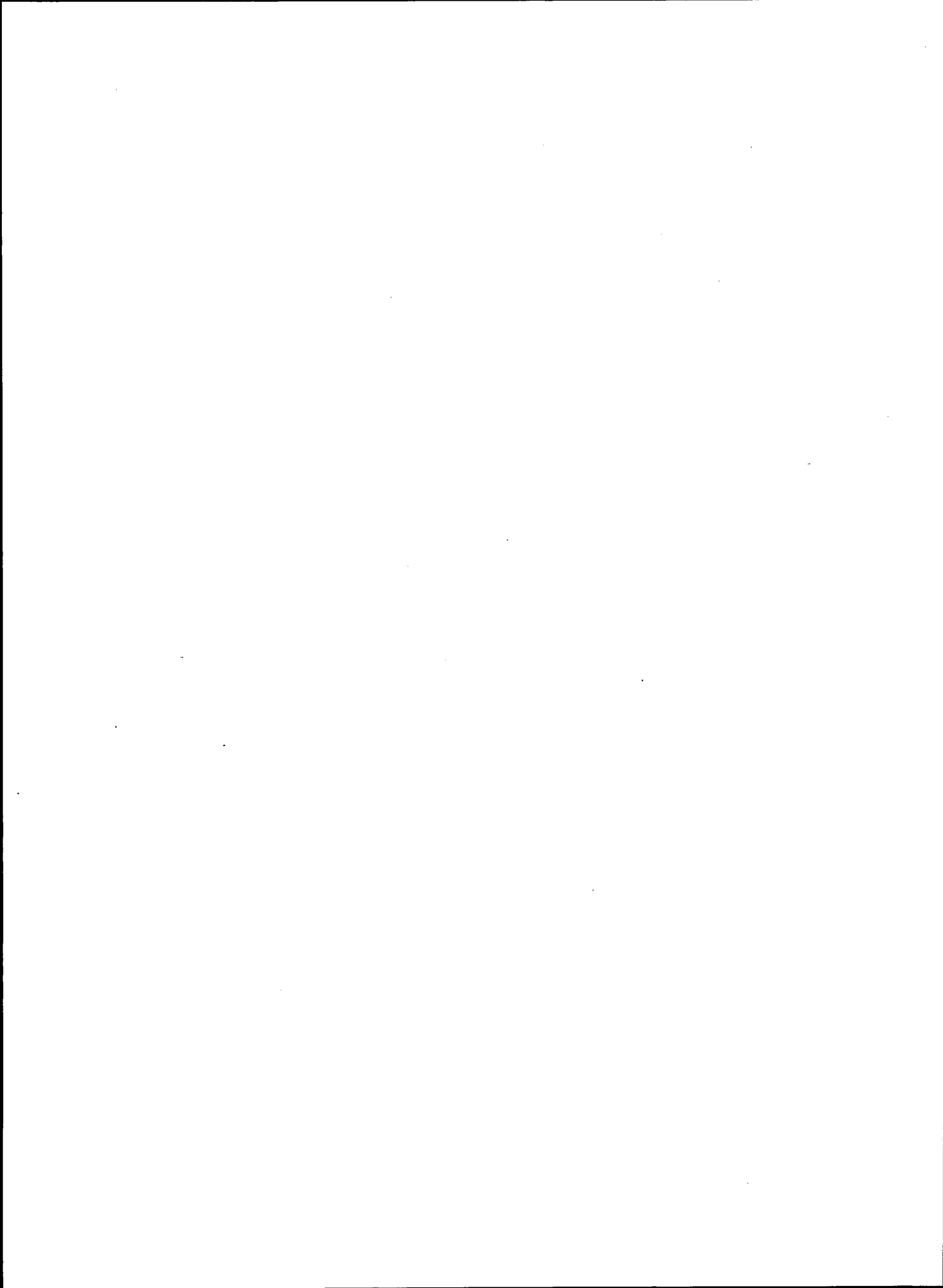
QALYs: Where Next?

GAVIN MOONEY* and JAN ABEL OLSEN**

1. Introduction

Why QALYs (quality-adjusted life-years)? The answer is that QALYs represent one way of measuring multi-dimensional outputs and that this in turn allows cost utility analysis (CUA) to proceed as an economic evaluative tool in health programmes. So why CUA? Primarily because CUA overcomes a major deficiency in cost-effectiveness analysis (CEA), namely that CEA can only deal with output which is uni-dimensional, such as lives saved. However, health programmes are frequently multi-dimensional and often reflect not just quantity *or* quality of life but both. CUA allows economic evaluation to reflect this multi-dimensional nature of the outputs—provided these different dimensions of quantity of life and quality of life can be measured and brought together in a single index. This is what QALYs aim to do.

If QALYs can be measured they can be used at different levels of decision making in health care, for example by clinicians in comparing one form of treatment against another for a particular disease or condition; or by health service planners and policy-makers to help to decide on priorities for where to spend the next increment in resources. It is in this latter context that marginal cost-per-QALY 'league tables' have been developed (Williams, 1985).

As economists we are interested in the measurement of health in the context of decision making with respect to resource allocation. This has two facets to it. The first is efficiency and the second equity, social justice, or, to use a more neutral term, distribution.

This chapter concentrates primarily on two questions. To what extent in the context of efficiency do QALYs adequately reflect what health services are trying to produce? And to what extent in the context of health output do QALYs adequately measure changes in health?

In addressing these issues it will be immediately clear that this is not a comprehensive review of QALYs, nor does it attempt to examine specific alternatives to QALYs such as healthy year equivalents (HYEs) as recently proposed by Mehrez and Gafni (1989). It also restricts attention to 'formal' health services while it is clearly the case that QALYs have potential relevance to other health producing (or diminishing) activities such as road safety, environmental pollution, and diet.

One of the central difficulties we see in discussions surrounding QALYs in both principle and practice is how efficiency is to be defined and measured and thereafter pursued or implemented. This may seem an odd comment for economists to make but it lies at the heart of the whole QALY debate. What do we want from our doctors? What do we want from our health services?

Additionally, QALYs tend to assume that the only concern of health services is health maximization, consequently that other benefits are not relevant and that distributional questions are not of concern. It is also normally assumed that how health enters the *social* welfare function is as a simple aggregation of how health enters *individuals'* utility functions. Thus, implicitly and sometimes explicitly, the proposal to use marginal cost per QALY as a basis for a league table in priority setting involves the assumption that *all* that is relevant in answering the above two questions is encapsulated in QALYs and that QALYs adequately reflect somebody's valuation of the benefits of health status gains.

What we are attempting in this chapter is to further discussion about ways of improving and developing QALYs and at the same time to point to their limitations. We suggest adding to them other relevant outputs or characteristics of health care—such as information—and giving due consideration to equity concerns. We then indicate some avenues for dealing with some of the existing measurement problems.

2. What Are Health Services Trying To Do?

(a) Health Care and Utility

For economists, it is reasonable to assume that the objective of the clinical doctor, as the patients' agent, is to maximize the utility of her patients subject to some resource constraint; similarly for a health service, that it is trying to maximize the welfare of the citizenry in some aggregate form (usually referred to as 'the social welfare function'), again subject to some resource constraint. One interesting question which can be posed at each of these levels is then: how

are the utility function and the social welfare function to be defined? What is it for the individual patient and for society that is to be maximized?

To try to address these questions, it is necessary to examine the nature of the demand for the commodity that we are dealing with, i.e. (primarily) health care. This has been summed up as follows:

The uncertainty generated by ignorance about health status, availability and effectiveness of treatment, etc. makes decision-making about the consumption of treatment difficult, especially as there may be substantial anxiety about making a wrong decision which could have serious adverse (ill-health) outcomes. Consequently, the consumption of health care—especially for life-threatening conditions—may also include the characteristic of being able to pass the burden of decision-making to the clinician. In other words, the demand for health care may include a demand to avoid having to make difficult decisions and bear the responsibility of such decision-making. These informational and decision-making features are central to the issue of the 'agency relationship' whereby the doctor acts as an agent on behalf of and in the interests of the patient. (McGuire *et al.*, 1988)

Given the characteristics of this commodity, the economists' standard approach to studying the market for goods and services (neoclassical economics) is not particularly helpful. Essentially this is because neoclassical economics assumes *inter alia* a degree of knowledge on the part of consumers which patients in health care simply do not have.

Thus the extent to which individuals can respond in an informed and rational way to health problems and thereafter demand health care in such a way as to maximize their expected utility (i.e. their future satisfaction) is severely limited. (It should be noted at the same time that it can be problematical to apply this 'expected utility theory' in other markets as well.) For a particularly good discussion of the problems in this context in health care, see Evans and Wolfson, 1980.

QALYs need to be viewed in the context of the nature of health care as a commodity. Rationality for the health care consumer is seriously 'bounded' as Simon (1961) described it, i.e. patients are 'intendedly rational, but only limitedly so'. There exists 'information impactedness' (Williamson, 1973) which results in patients simply being unable to cope with the complexities with which they are faced and thereby relying on the doctor for information.

One of the important emphases in expected utility theory is on consequentialism. The *only* considerations that are utility bearing in expected utility theory are the consequences (or outcomes) resulting from various choices. In utilitarian philosophy as Mill (1979) states 'actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness' or more crudely as it is often interpreted the goal is 'the greatest good of the greatest number'. It is the consequences or outcomes that count in judging the extent of happiness or of goodness. Further, the Mill view tends to assume that the social welfare function is simply an aggregation of individuals' utilities.

These factors rule out any good or utility associated with the 'process' or 'action', e.g. utility in choosing or indeed in not choosing, the latter perhaps being better expressed as not having to choose. It also says nothing about the desirability of different *distributions* of happiness. Further, it may, if interpreted too narrowly, rule out the inclusion of the utility of information. There is a need for care here.

Using the formulation of Hey (1979):

$$U(C_j) = \sum_{i=1}^{I} P_i U(A_{ij})$$

where $U(C_j)$ is the expected utility associated with the choice, C, for individual j; A_{ij} is the consequence if the individual chooses C_j and the state of the world, i, occurs; $U(A_{ij})$ is the utility associated with A_{ij}; P_i is the probability that the state of the world, i, occurs; I is the number of possible states of the world such that $i = 1, 2, \ldots I$. Here, then, the *only* source of utility is from the A_{ij}s, i.e. the consequences.

This assumption on consequences underlies much of the work on QALYs. It leads quickly to the view that

(1)   the only relevant arguments in the health care consumer's utility function are the consequences of that consumption, often reduced to health status changes *only* and thereby QALYs;

(2)   priorities in health care can be determined on the basis of marginal costs per QALY.

Additionally, when it comes to the question of considerations of the social welfare function, much of the QALY literature assumes that the only relevant arguments in that function relate to individuals' utility and only then from health and with no consideration of equity entering in. Consequently the social welfare function for health care for a society of n individuals is assumed to take the form

$$\sum_{j=1}^{n} U(C_j)$$

and where utility is derived solely from health status gains. In this context we want to consider three issues: information; decision making/autonomy; and equity/distribution.

Before turning to these considerations, it is relevant to comment briefly on the question of interpersonal comparisons of utility—for example, *can* we compare the satisfaction that different individuals get from health—and the related question of cardinal measurement of utility. We would make three basic points

here. First, QALYs in no sense resolve the problems in principle which surround interpersonal comparisons of utility. Second, no claim is made that QALYs do 'solve' this problem. Third, it can be argued that, in practice, the issue of interpersonal comparisons of utility is addressed daily by both clinical decision-makers and health-service planners and in cardinal terms. Decisions are made in practice which involve comparisons of one person's cardinal utility from treatment (or quicker treatment) with another's and which involve allocating extra resources to benefit one group of individuals, say, pregnant women rather than another, say, diabetics. (Of course it can be argued that doctors do not attempt such comparisons but simply impose their value judgements without even attempting to judge individuals' relative utilities. We believe doctors make some effort to understand, judge, and compare different individuals' utilities.)

Certainly, the whole idea of interpersonal utility comparisons has been met with much resistance in economic theory since the seminal paper by Robbins (1938). However, although we cannot, on scientific grounds, compare the utility of one person with the utility of another, it is possible to frame comparisons in the form of imaginary choices, such as: 'Would you rather be person A or person B given the choice?' (see e.g. Harsanyi, 1955).

When measuring health state utilities (of which more in section 3, and see also Drummond in this volume), for example, using the time trade-off (TTO) method, utility comparisons are made by such imaginary choices. 'Would you rather have X years in perfect health than Y years in a described inferior state?' However, the problem of interpersonal comparison is not being resolved. Although two individuals might share the same X/Y and thus be assigned the same cardinal utility, the *strength* of the preference for *both* X and Y might well be higher for one of them. The reason is that while everyone might agree that being dead has a utility of 0, the utility from a year in a specified (less than perfect) health state might well differ between individuals.

By assigning the same finite end points, 0 − 1, on the cardinal scale (as is normally done in QALYs—but see the discussion in 3(c) below), individual differences in their relative strength of preferences are being ignored. Hence, individual differences in the utility increment by moving from one inferior health state to an improved state are ignored in a QALY framework. These would *not* be ignored within a strict utilitarian framework where, *ceteris paribus*, the individual with the highest potential utility increment would be preferred. Thus, QALYs are normally at best and in practice 'quasi-utilitarian'.

It is thus in our view important to assert three things with respect to interpersonal comparisons of utility. First, the problems associated with such comparisons are not solved in QALYs and, in our view, cannot be solved. Second, this fact is not to be seen as an argument against QALYs but rather as a recognition that in the real world of health care interpersonal comparisons *are* made. Third,

that in this real world, such comparisons are made only or, at least, very largely implicitly—and not explicitly as QALYs require.

(b) *Information and Decision-Making*

It is quite apparent from what has been said about the nature of the commodity health care, that the consumer is not informed to the extent that would commonly be assumed by expected utility (EU) theory. The normal formulation of EU theory requires knowledge regarding: (1) the choices available; (2) the states of the world that are the consequences of these choices; and (3) the probabilities that the states of the world will occur.

The consumer of health care is lacking information on all three fronts. She does not know what set of treatments/procedures is available, how effective these are, nor what the probabilities are that the various possible outcomes will occur. To assist in the process of choice, the consumer may want to reduce the extent of her imperfection of knowledge on any or all of these issues—or may seek to default on decision-making, passing that responsibility to the doctor.

If it is assumed, however, that it is only consequences that are relevant in terms of being utility bearing and, beyond that, that health status changes are the only consequences that are relevant, then information from the doctor/agent is not utility bearing in itself.

This seems problematical. Culyer, in this volume, suggests that neither health care nor information 'are valued for themselves'. (While it might be simply semantics, it would seem better to include information as a characteristic of health care and therefore reword this to say instead 'neither treatment nor information'.) While we would agree that (normally) treatment will not be valued *positively* for itself (but we can envisage it quite frequently as a 'bad'—many treatments involve pain, for example), there are situations where information will be of value for itself. For example, the majority of individuals screened do not have any health improvement; however, those screened as true negatives get information which is valued for itself. It is of course possible to include in QALYs such factors as anxiety or reductions therein as Feeny and Torrance (1989) propose in the context of pre-natal screening. Information can be incorporated at least partly here. But there are situations where QALYs have not embraced this element and situations where information can itself be an argument in the utility function.

There may even be situations where information which confirms the presence of disease carries positive utility. This would appear to be the case where women agree to a prenatal diagnosis, for example, for polycystic kidney disease, when they have already decided that they will *not* abort if the test is positive. The benefits associated with information were also revealed in the study by Berwick and Weinstein (1985) which indicated women's willingness to pay for informa-

tion on various aspects of their pregnancy at least some of which had no bearing on changes or feasible changes in health status.

Certainly there will be other ways in which information will affect expected utility than simply as a consequence in the utility function. We do not want to dispute that. Rather, we simply want for our purposes to propose that health status improvements are *not* the only relevant states of the world.

A more detailed example may be helpful here. In a study in Denmark currently being undertaken by one of the authors, doctors were interviewed about various aspects regarding the decision to operate. One doctor stated clearly that in the end it was the patient's decision. However, when pressed on what he did when the patient persisted in wanting the operation when he thought this was not the best option, he stated: 'I then tell her about the possible adverse outcomes.' We believe this is an example of a situation where the disutility (as perceived by the doctor/agent) of information was such that full information was initially not disclosed and only when the 'wrong' decision was about to be made was the information forthcoming.

We do not recount this as an example of something necessarily being wrong in a moral sense with what this doctor said he would do. Rather, we believe it reflects a much wider consideration where at least some doctors do sometimes attempt to take account of the utility and disutility of informing patients. (They may not get it right in terms of utility maximization but that is something else!)

We would also want to distinguish between information and decision making. It is pertinent to suggest that perhaps because of the emphasis in EU theory on the presence of an informed, rational consumer, and the problems for many economists in rejecting this, the emphasis in health economics tends to remain too often on the same informed and rational idea, albeit via an agent. In other words, it is assumed that the perfect agent's task is to allow the patient/consumer to *end up* where she would have been if informed and rational.

In this context we want to quote from Williams (1988b) at some length.

The basic weakness of the idealized view of both [basically 'public' and 'private'] systems is the peculiar 'agency' role which doctors play in all health care systems. The essence of this problem is that the 'consumers' rely on doctors to act as their agents, in a system which ostensibly works on the principle that the doctor's role is to give the patient all the information the patient needs in order to enable the patient to make a decision, and the doctor should then implement that decision once the patient has made it. I am sure that the reader would find the above statement closer to his or her own experience if the postulated roles of patient and doctor were interchanged, so that the sentence would then read 'the *patient's* role is to give the *doctor* all the information the *doctor* needs in order to enable the *doctor* to make a decision, and the *patient* should then implement that decision once the *doctor* has made it.

Given the nature of health care, then the case has yet to be made that the role of the doctor/agent *is* necessarily idealized through giving the patient '*all* the in-

formation the patient needs' *or* through having the patient make the decision. This holds only if perfect information and consumer sovereignty are necessary ingredients of any theory with which we are operating—and, strictly, if the provision of information is costless. (Otherwise optimal information will be less than perfect information.)

But are they? It is certainly not certain that perfect or total information (even if costless to provide) will always be optimal to the patient's utility function. Being the decision-maker may not result in maximum utility to the patient—even where health status improvements (QALYs gained, if you like) are the same.

An alternative to consumer sovereignty (other than paternalism), and one to which many doctors subscribe, is autonomy. Defining autonomy is not un-problematical. We have adopted the (rather broad) definition of the principle of autonomy as stated by Beauchamp and Childress (1983): 'Autonomous actions and choices should not be constrained by others. ' They argue that this principle 'asserts a right of non-interference and correlatively an obligation not to constrain autonomous actions—nothing more but also nothing less.'

Our view of such autonomy in the current context is that, depending on the preferences of the patient and assuming that the patient is competent to form self-interested preferences, for some patients (1) autonomy will mean the same as consumer sovereignty with the informed patient choosing; for others (2) it will mean choosing to have someone else make the decisions regarding treatment; and for yet others (3) autonomy will lie somewhere in between. Note that in the case of (2) the position is not constrained by the level of information—the patient can choose to receive any level of information from zero to fully informed and then choose to have someone else make the decision.

We thus argue that the question of patient autonomy enters the patient's utility function in the form of a variable reflecting the patient's preferences for decision making. The job of the perfect agent is then not necessarily to inform *fully* to allow the patient to decide but rather to inform *optimally* and to allow *optimal* involvement in decision making—where optimal in both instances is defined according to the *patient's* preferences. (We would also endorse Beauchamp and Childress' view that the principle of autonomy 'does not apply to persons who are not in a position to act in a sufficiently autonomous manner'—but we will not pursue here how that is to be defined in practice.)

There are thus in our view three arguments in the patient's utility function: health, information, and decision making. The perfect agency relationship would allow the patient to maximize the utility associated with all of these arguments—and not just health. We are clear that for the patient there is more to health care than improved health status or QALYs gained. For those who doubt the possibility of the non-health variables being present in the patient's utility function, we would pose the following questions: Are there no circum-stances in health care where your utility as a patient would be unaffected by who

made the decision? Are there no circumstances where too little or too much information might not lead to a fall in your utility?

We find it difficult to see how the decision-making component (and perhaps even information *per se*) can be included in the *consequentialist* utility of expected utility theory (but see Dowie, 1989). We would suggest that, however one handles these other aspects, the crucial point is that they *are* additional to considerations of only health.

How information is incorporated is, however, a relevant issue. The answer is at least partly dependent on the decision environment. For example, if annual check-ups picked up, let us say, no health defects but increased individuals' utility by reducing their anxiety, is such utility to be counted in? More generally in the context of information, what is the role of information with regard to individuals' perceptions or misperceptions of probabilities?

In the proposal on prenatal screening by Feeny and Torrance (1989) mentioned earlier, they imply either that respondents' perceptions of the probabilities involved (e.g. of some problem in the foetus) are 'accurate' or that they are irrelevant. They state that the expected utility of diagnostic tests can be estimated by combining the utilities of the relevant health states 'with estimates drawn from the trial results on the frequency and duration of each state'.

QALYs need to be set more clearly in the decision-making frame of the clinical setting and health service planning, particularly with respect to information and patient autonomy. Just how these factors affect the patient's utility function is not clear. We have indicated that information and decision making seem to be important arguments in at least some instances. More research is needed to determine how to incorporate these—and any other factors that the patient wants included—into the patient's utility function.

(c) *Equity*

Most health services include, at least implicitly, but often explicitly, equity as a goal. Many of these express this goal in terms of equality of access (often with the qualification, 'for equal need'). One of us has argued elsewhere (Mooney, 1986) on the merits of the Margolis (1982) 'fair shares' approach in addressing the issue of equality in health care—at least in part because it endorses the most frequently found health service equity goal, i.e. it is access-related.

Margolis draws a distinction between 'selfish' and 'group' utility. The former is the 'normal' outcome or consequentialist utility. The latter is based on the idea that individuals derive utility from *doing* their 'fair share' (hence the name of Margolis' model) for some group, e.g. some community or society generally, of which they are themselves members.

According to this theory, if we examine it in the context of health care, the concept of 'fair sharing' would lead individuals to want to participate in the provision of health services which were available to all. Since in the Margolis

model participation utility is derived not from the outcome for the group but from the participation *per se*, it follows that it is the provision of *access* to health care which is the basis of the utility here. The approach is discussed by Culyer at some length in this volume.

However, the essential point on equity in the context of QALYs is that it is very clear that for most (all?) health services, equity in the form of access is a consideration in the social welfare function for health care. This is not currently reflected in QALYS, which are primarily about individuals' utility from health rather than about the social welfare function *per se*.

What is particularly relevant here in the specific context of QALYs is that if the Margolis model were accepted and with it the notion of participation altruism then caring about access would enter the individual's utility function and through that could then be incorporated into the social welfare function.

The question of what is meant by caring in this context is important. Earlier Culyer (1980) wrote of the 'caring externality' to explain the way in which others' health entered an individual's utility function. Culyer's caring does not itself bear utility; nor in fact does being cared for bear utility.

Indeed there would seem at times to be some confusion about the notion of caring. Our view is summarized by Wright (1987):

> When we speak of caring, often a subtle ambiguity creeps into our thinking, for caring can be quite good, in the technical sense of meeting physiological needs and correctly utilizing science and technology, and yet not be caring at all ... If I am seeing to your every physiological need, I am certainly giving you care or *taking care* of you, but I may be doing this without *caring for* you at all. I may be seeing what I do simply as my job, something that must be done to guarantee my paycheck. Although such a view is appropriate for fixing cars or manufacturing spoons, it is not appropriate in dealing with other persons, because then the person is treated as an object.

We would want to put forward the view that caring (and being cared for) and equity, specifically with respect to access, are also possible arguments in the social welfare function regarding health care. We would also suggest that these are perhaps better accommodated in some utility function that more readily accepts participation or process utility in addition to consequentialist utility than does expected utility theory.

What this means, we believe, is not that QALYs should be abandoned. Rather, there is a need to examine first what, in addition to health, is in the patient's utility function, so that these can be added to the QALYs when using health care outputs, and second, what the nature of the relevant social welfare function is. Certainly if it is proposed, as Culyer (1988), does, that health services exist to promote health only or as he states 'given the resources available to the health services, the health of the community should be maximized' then any measure of health status—QALYs or whatever—potentially covers everything relevant to the individual's consumption of health care. But Culyer's subsequent

comment that 'egalitarianism equals health maximisation' can only hold if the assumption is made (and others are also necessary) that health care is *only* about health and it can hold if we retain the assumption that the relevant social welfare function is simply a summation of individuals' utility functions (in which health is the only argument).

What are the implications of accepting a broader based social welfare function based not purely on the summation of individuals' utility functions in which only health enters? At a clinical level, this might mean accepting that doctors spend time listening to patients and trying to assess how much information they wish to have and how sovereign as consumers they wish to be. At a planning level it might mean that geographically remote areas might get higher priority in terms of resource allocation if equity considerations were included in health care priority setting than if they were not. And depending on such considerations as economies of scale this equity consideration could alter rankings of marginal cost per QALY, i.e. if the geographical location of the relevant technologies were determined not just on efficiency grounds but on equity grounds. (We would hazard the guess that the additional costs to meet some sort of geographical equity criterion would be relatively greater in the case of heart transplants than pacemakers.)

In terms of social class, too, the inclusion of some equity criterion is likely to affect the rankings of marginal costs per QALY. Thus if the costs considered include those falling on potential patients and not just health service costs, then in instances where the *only* difference in terms of marginal cost per QALY between choosing to treat a social class I patient and a social class V patient was the access cost to the patient, the patient with the higher access cost would have a higher marginal cost per QALY.

3. Measurement Issues

Even given the limitations of QALYs as indicated above, it is none the less important to consider, within their own lights, how adequate QALYs are as measurements of health. In this section of the chapter, beyond a brief description of the main methods for measuring QALYs, we consider five measurement issues that arise with QALYs that we would claim are both important and to date not satisfactorily resolved.

The most frequently used methods in the QALY literature are the standard gamble (SG), the time trade-off (TTO) and category scaling (CS). Only a brief explanation of the methods is given here. For a more complete exposition see Torrance (1986, 1987), Drummond *et al.*(1987), and Drummond in this volume.

SG is based on von Neumann and Morgenstern (1944). Respondents are faced with the choice between a certain prospect and a lottery with two possible

outcomes, one of which is better than the certain prospect and the other is worse. The question is then to find the probability level, p, in the lottery, at which the respondent is indifferent between entering the lottery and accepting the certain prospect. In health status terms, the preference value of the certain health state, i, is then $h_i = p$.

A different version of the classic SG is the use of 'certainty equivalents' (McNeil *et al.*, 1978). Acknowledging the difficulties people have in understanding gambles other than the flip of a coin, respondents are asked to specify the period of certain survival that is considered to be equal to entering a gamble with a 50 per cent probability of dying and a 50 per cent probability of survival in a given period. If the certainty equivalent is lower than the expected utility of the gamble, this indicates risk aversion.

The practical problems with SG are, first, that it is difficult for untrained respondents to understand it and, second, that it is highly sensitive to the actual framing of the questions (Kahneman and Tversky, 1979). Further, it can be costly to apply and the revealed values may contain at least an element of people's attitudes to risks.

The TTO was developed by Torrance *et al.* (1972). The method aims at excluding the influence of respondents' attitudes to risk by presenting two alternative health outcomes which are to occur with certainty: either Y years in a described (inferior) state or X ($<Y$) years in perfect health. The respondent is asked to trade off some years in the inferior state for a shorter life in perfect health, so that Y years in the inferior state is equated to X years in perfect health.

Category scales are sometimes termed category rating or rating scales. A related method is magnitude estimation. In this approach, respondents are asked to assign relative values or utilities to describe health states in relation to fixed end-points where normally death is worst and assigned a value of 0, whereas the best imaginable state is assigned a value of 1. As Torrance (1987) describes it: 'The remaining health states are placed on the line between these two, in order of preference, such that the intervals between the placements correspond to the differences in preference as perceived by the subject.'

Using CS, respondents' utilities are derived *explicitly*. Using SG and TTO, the utilities—or health state values—are elicited *implicitly*, based on raters' responses to decision situations. In all instances the utilities are measured cardinally.

(a) Utility Evaluation Bias

There appear to be systematic disparities between the values obtained by the different approaches, what Loomes and McKenzie (1989) call the 'utility evaluation effect'. Torrance (1976) compared the values obtained from the three methods and found that CS values were lower that SG and TTO. A study by

Llewellyn-Thomas *et al.* (1984) comparing SG and CS, supported Torrance's findings that the CS values were lower than SG values.

Read *et al.* (1984) arrived at a different conclusion from Torrance. They found that the SG method gave significantly higher values than TTO and concluded that $SG \geq TTO \geq CS$.

A possible explanation for the relatively low CS values may be the choiceless context of the CS method as opposed to TTO and SG. In CS, subjects are asked merely to assign relative values to the quality of inferior health states. A respondent might well suggest that the quality of a given health state is 1/3 of perfect health, but not imply from this that she is willing to trade 3 such years with 1 perfectly healthy year. By its very nature, CS does not press for a 'tragic choice' between a long life in an inferior state and a short life in a perfect state. In CS the respondents do not have to 'pay', either by giving up life-years or by undertaking gambles which might involve losing.

In other words, we suggest that there is a likelihood of some strong 'preferences for living'—that life is considered very or even too precious in any trade-off with improved quality. Such an attitude was revealed in a study by Pliskin *et al.* (1980). Although acknowledging that angina was not a perfect health state, some of their respondents were not very ready to trade off life years to obtain perfect health. This observation might explain why TTO values are higher than CS values.

In TTO the two alternatives are presented with certainty, whereas SG involves gambling. It seems to be generally accepted that individuals are risk-averse with respect to choices involving health outcomes (see e.g. Hellinger, 1989), which is likely to serve as an explanation for the higher SG scores.

What, then, are the relative strengths and weaknesses of the different methods? The SG approach appears to have its strongest advocates among medical decision theorists. Llewellyn-Thomas *et al.* (1984) call it 'widely accepted and theoretically sound'. Mehrez and Gafni (1989) call it the 'gold standard' measure. At the same time, Loomes (1988) contends that measures of choiceless utility are not appropriate when decisions have to be made but adds 'neither are the values generated by the standard gambles, if they fail to take account of any regret and rejoicing that may be involved.'

The CS appears not to be posing the right questions at a planning level, where the health aim is to maximize health improvements measured by quality and length of lives within a given budget. The reason is that when revealing their preferences through CS, the respondents are not faced with the same trade-off considerations that decision-makers face when allocating scarce resources.

This choice issue is important in choosing between measurement techniques and on that basis alone it would seem that the TTO is the 'true' QALY method. Whichever method is used, some interesting implications for resource allocation arise. For example, the low CS values will favour improving lives compared

to extending lives, whereas the high SG values will favour extending *vis-à-vis* improving lives. What we would conclude is that, in many situations, the TTO poses the right QALY-question. Whether it gets reliable answers is another question!

(b) *Should Attitudes to Risk Be Taken Into Account?*

One of the potential problems about the measures is that the CS does not offer a choice, the TTO offers choice with certainty and the SG, while introducing risk, is based on von Neumann–Morgenstern which explicitly assumes risk neutrality. Yet individual decision-makers, whether they be patients, clinicians, or planners may well have non-neutral attitudes to risk.

However, if SG is applied in practice to derive the index values, what emerges will inevitably measure not only the utilities of living in the inferior conditions, but also any aversion to undertaking risky treatments that are likely to improve the states. Whether it is relevant, within a QALY health state utility framework at the planning level, to 'boost' the utility values to reflect individuals' risk aversion if they are offered improvement, is an issue of debate. We see no reason to exclude this factor.

However, the concept of risk aversion is more ambiguous than that which follows from conventional wisdom. An individual is normally classified as risk-averse if she prefers the certain outcome to a gamble with the same expected nominal value. By introducing the notion of strength of preferences, Dyer and Sarin (1982) offer a competing explanation to such behaviour, namely the diminishing marginal value placed on extra units of the goods being offered. Their illustrative example is where there is indifference between three oranges for certain and a gamble involving a 50 per cent chance of eight oranges and 50 per cent of none. An analogous example in health would be being indifferent between the certainty of three years in a given inferior state or a gamble involving a 50 per cent chance of eight years in the same state and a 50 per cent chance of dying. The conventional explanation of such behaviour would be risk aversion. However, within the Dyer and Sarin (1982) framework, this could be explained by the likelihood that the person assigns decreasing marginal values to each succeeding year in the particular state—an explanation that violates the assumption of constant health state utilities which is assumed in most of the QALY literature (and which we consider in the next section).

How to include risk attitudes in practice does not yet appear to have been satisfactorily resolved. What may be particularly problematical here is the question of *whose* attitudes to risk should be incorporated. It may seem obvious that, at least in a clinical setting, these should be those of the patient. However, given the earlier discussion (in section 2) about the utility to be derived by some patients in passing the burden of decision making and thereby at least some of the elements of risk taking to the clinician, it is not so clear that it *is* the patient's risk

attitudes that are relevant. In principle, the answer might be for patients to choose clinicians who share their attitudes to risk; in practice a rather difficult solution to obtain!

On the question of the relevance at a planning level of individuals' attitudes to risk, as stated above we see no reason to omit these. It would only be if the goal of health service planning were health maximization with no concern about patients' attitudes to risk or if patients were risk-neutral that it would seem justified at a planning level to omit considerations of patients' attitudes to risk.

(c) *Constant Health State Utilities?*

In TTO, the question posed seeks an indifference point between Y years in a described inferior health state and X years in perfect health. When a respondent reveals her indifference point, the analyst assigns a cardinal utility equivalent to X/Y throughout the Y years, i.e. constant health state utility is assumed. The respondents' underlying considerations are, however, likely to be influenced by a whole range of 'time-related variables' that will give a more complicated utility scenario than the X/Y assumption does (see, for example, Nord, 1988).

The aim with the various health state measurement techniques is to assign a health state value to a described health state. Thus, the health state is intended to be an objective description of the state the individual is thought to remain in throughout the whole period. The health state value is an average (= X/Y when TTO is used) of the described health state.

This is open to question. First, there is the issue of different utilities depending on life stages. Although it seems generally accepted in the literature that we consider some life stages to be more important than others, there is scant empirical evidence. An exception is a preliminary attempt by Williams to examine the issue. Using data from a questionnaire survey, he elicited some 'entirely hypothetical' weights against an average index of 1.0 on the relative importance of different life stages. 'As infants' and 'when bringing up children' were life stages that were considered much more important than the average, respectively 2.75 and 3.31 (Williams, 1988a). While Williams stresses that the numbers are purely indicative, it does seem likely that if someone has a life stage of bringing up children in front of her, when responding to a TTO or SG question, that individual would be less willing, *ceteris paribus*, to trade off these life-years.

However, the question of the relative importance of different life stages is not directly relevant to the weighting of QALYs *per se* in cost utility analyses. Here the relevant consideration is the value attached to gains in quantity and/or quality of life. We are interested in marginal gains (and losses) and the fact that a life stage is judged to be more important than the average does not necessarily tell us anything about the relative valuation of some reduction or increment in health status at that life stage as compared with that at the 'average' life stage.

Secondly, there is the issue raised by many authors (for example, Nord, 1988) that the health state utility of inferior health states is likely to decrease because individuals will get weary and satiated after some time in a dysfunctional state. The explanation put forward for the diminished utility is that, after some time, people will have problems in coping mentally with their dysfunctional state. However, for some dysfunctional states, it is possible that the actual health state description will change. The physical disability may get worse as well as the distress.

Sutherland *et al.* (1982) introduced the concept of 'maximal endurable time' (MET), beyond which 'additional time spent in that state was regarded as a penalty and assigned a negative value relative to death'. The more dysfunctional the health state, the shorter was the MET.

When questioning the assumption of constant health state utilities, the literature seems to be concerned only with the possibility that they might decrease. However, it might be that when individuals learn to cope with their new situation, they adjust to and tolerate it, thereby increasing their health state utility as compared with the initial set-back.

While acknowledging the relevance of these non-constancy possibilities at the clinical level, a pragmatic counterargument to the validity of assuming constancy can be presented for the planning level. Given that the objective of QALYs is to serve as an aid for resource-allocation decisions at the planning level, and that this is a continual process, the average health state utilities revealed by individuals in any period would be a representative reflection of the utility at the group level.

In an interesting consideration of the treatment of utility over time, Richardson *et al.* (1989) presented respondents with three health states spread over a period of 16 years, the last year being particularly unpleasant. While their results can be interpreted in various ways, their preferred explanation is 'that future prognosis affects—contaminates—the assessment of previous health states in such a way that the holistic and composite approaches to measurement are incommensurable... the knowledge of future suffering and death casts a shadow over—it devalues—the enjoyment of earlier life years.'

The measurement problems here are not conceptual but purely empirical. We simply need more empirical evidence to be able to judge the validity of the constancy position.

(d)   *Is There a Time Preference for Health?*

The observation that the utility in chronic dysfunctional states may decrease has led some authors to argue for positive time preferences for health. We would suggest that this has nothing to do with time preference *per se*. Sutherland *et al.* (1982) claimed that the existence of a 'maximal endurable time' (MET) means that the time preference rate for health in these dysfunctional states increases as

one is approaching the MET, i.e. the time preference rate (TPR) is non-monotonic. The observation that an individual 'values less each succeeding year in a chronic dysfunctional state simply because he is getting weary or satiated with the situation' made Gafni and Torrance (1984) introduce the notion of a 'quantity effect' as an argument for increasing the discount rate for health.

However, accepting that the health state utilities are decreasing with succeeding years in a dysfunctional state implies that we are no longer comparing two identical utility levels taking place in two different time periods, which is the relevant consideration when assessing individuals' time preferences. Rather, we are comparing different utility levels in different time periods, or to put it in Böhm-Bawerk's (1888) original terminology: 'the true intensity of their future marginal utility' is *changing*. An important distinction should be made: the decreasing future utility in a chronic dysfunctional state may be due to a real decrease in utility as time goes by. It may not be due to the remoteness in time of the future utility, which essentially is the impatience argument for assigning lower utility levels to future consumption.

Although Gafni and Torrance (1984) explicitly separate the 'quantity effect' from pure time preference, the introduction of this effect contains two implicit assumptions. First, it is assumed that the utility-decrease from getting wearied and satiated follows a smooth monotonic pattern in the same way as the time preference rate. A study by Sutherland *et al.* (1982) does not support this view. Secondly, it is implied that the health state utility will always decrease over time. As discussed in 3(c) above, the utility might well *increase* as compared with the initial setback as the individual learns to cope. These sorts of problems would seem to support our suggestion for sticking to a narrow understanding of the time preference concept, and in addition for being explicit with respect to time-related utility changes which have no time preference element.

When a positive time preference rate for health is observed using gambling approaches (McNeil *et al.*,1978, 1981; Pauker and McNeil, 1981; Sutherland *et al.*, 1982), a competing explanation could be found in people's attitudes to risk. McNeil *et al.* (1978) found that 'life during the next few years is much more important than life many years later'. This might be due to a positive TPR. However, it might well be due to changing attitudes to risk because, according to Mehrez and Gafni (1987), risk aversion is more common for shorter periods while risk seeking increases with the length of the time period considered.

In a recent study by Hellinger (1989), fifty individuals were asked whether they valued each of the next thirty years, assuming good health, equally. Of the nineteen who did so, and thus revealed no time preference, eighteen were risk-averse.

While it might sound peculiar to use the concept of pure time preference for 'life *years* in a given health state, perfect or inferior, it is more meaningful to use the concept for temporary health gains or losses. Lipscomb (1989) observed

positive time preference by varying the delay of onset of a temporary inferior health state and also the duration of the inferior state.

Llewellyn-Thomas *et al.* (1984) observed a non-monotonical time preference rate—increasing as the patient approached the time when pain set in. Fuchs (1982) refers to a study which indicated that the time preference rate was lower when assessing the relative preferences between two points both in the future than between the present versus the future. In other words, short-sightedness is a short-sighted phenomenon!

To date there are poor empirical estimates of any TPR for health. The claimed positive TPR may often be due to risk aversion or decreasing utility in an inferior health state due to satiation. Measures of TPR for health seem to be highly conditional on the specific situation. The revealed TPR in one situation might be irrelevant in another. An individual's short-term preference for immediate pain relief might indicate a very high TPR, while her healthy life-style might indicate a low TPR for health in the long run.

Not only is the revealed TPR situation specific, it might also be sensitive to the actual framing of the question. Say that an individual values her life-years equally in the next thirty years, the implication being that there is no time preference. The message to the QALY tool-kit is to leave QALYs undiscounted when ranking alternative programmes. However, say that two programmes, A and B, have identical net present costs and identical undiscounted numbers of QALYs, i.e. the cost per undiscounted QALY is the same. The only difference between the two programmes is that A saves QALYs in the present period, whereas B saves QALYs in ten years. Which of the two programmes would the individual vote for? Intuitively, we would expect that A is preferred.

Thus, although one might reveal an individual TPR for health equal to zero, the social TPR to use in QALY decisions at the planning level might still be positive. If society is not indifferent between the timing of its QALY gains, the social TPR for health would serve as the appropriate measure for the relative strength of its preferences for present versus future health gains. Note that it is not *a priori* given that this social TPR for health would be identical to that applied as the social TPR elsewhere in the economy—but clearly this creates some problems in allocating resources between health and other sectors.

Weinstein and Stason (1977) suggested that since health benefits are being valued relative to monetary units, then, for the sake of consistency, they have to be discounted at the same rate as costs. Keeler and Cretin (1983) elaborated on this point and showed that if health benefits were discounted at a lower rate, it would have a 'paralysing effect' on decision-makers, as it becomes better always to postpone any project. The point is clearly relevant for intertemporal budget allocations. However, this conventional wisdom does not, in our view, undermine the idea of ranking health programmes according to net present costs per

undiscounted QALY (or per discounted QALY using a lower rate than on the cost side) within the *current* budget.

(c) *Whose Values?*

A key measurement issue is quite simply whose values to elicit. There is little agreement in the literature on this point—but this may be all right as it is probably not for the analyst to decide. Analysts need, however, to accept that the answer to the question 'Whose values?' will tend to vary with the decision environment. Mulley (1989) in this context makes the following pleas for future research:

Utility theorists need to pay more attention to the departures from the normative model and the cognitive heuristics and biases that patients use in its stead if they are to help clinicians help patients make their best decision. The clinician needs to better understand the shared nature of both decision-making and decision-making responsibility so that patient autonomy can be realized while still offering protection from unnecessary anxiety and regret.

And he then poses two key questions: 'Whose utilities should be used to develop consensus values for appropriateness? To what extent can variability in utilities for health outcomes be used to distinguish between health care "needs" that reflect societal consensus and health care "wants" about which consensus may not be possible?'

As one possible way forward on this front we would suggest that a random sample of the population should be asked not only for their values but also what weight should be attached to their values (and perhaps also what weight should be attached to others' values). Thus in the context of screening women for breast cancer, we two males might be part of the random sample but might then choose to have zero or low weights attached to our responses and we might also want to have some say in proposing that women should have a higher weight than men, and women with breast cancer the highest weight of all.

4. *Que Faire?*

On measurement issues, we have tried to indicate some of the current difficulties in QALYs. None of these is in our view overwhelmingly problematical. But there is a need for greater empirical work in the area, especially, we believe, in respect of the choice of the basic measuring technique and of the question of time preference.

These problems of measurement are not surprising. Given the work by Kahneman and Tversky (1979) on the variations in responses as a result of different ways of framing questions, it is perhaps inevitably difficult not to get different answers emerging from different measurement methods.

Beyond that, perhaps the underlying theme for the future research agenda on QALYs should be the decision environment in health care at both the clinical and the planning/health policy level. There is too little understanding of the behaviour of the key actors in the system, what is in their utility functions, how they make their decisions and how different stimuli affect their decisions. Simply as examples of this lack of understanding, we still know little today about the nature of GPs' decision-making and the impact on that of different forms of remuneration (Donaldson and Gerard, 1989). In the hospital sector, most of the economic research adopts models that seem to show little appreciation of what goes on in hospitals (McGuire, 1987).

Much that is relevant to the peculiarities of health care relates to information, consequently information processing and decision making. More than a quarter of a century ago, Arrow (1963) argued that the medical profession specialized in information. What is clearer today is that the information processing of medical doctors and of patients and the decision-making procedures based on this are particularly complex phenomena that economists might usefully appraise more than they have to date. There is something missing at the very centre of health economics when we present models of supplier-induced demand and of the agency relationship and yet still argue that the only relevant output from health care is health and that the measuring processes for health should be based on expected utility theory. More research is needed by economists on decision making in health care, perhaps looking to the economics of institutions (Simon, 1961; Williamson, 1973; Langlois, 1986) where the extent of information and of rationality is recognized as being limited. What QALYs currently appear to lack—despite their strengths and despite the improvements they offer on what exists in health care today—is a better understanding of the decision-making milieu of health care and a wider view of the nature of the health care social welfare function.

Loomes and McKenzie (1989) have previously drawn attention to the potential merits of alternatives to expected utility (EU) theory in the form of either prospect theory or regret theory. Their proposals need to be seen against a background of considerable evidence that EU maximization appears not to be a general description of how individuals behave and that the axioms of EU theory as expressed earlier are frequently violated (Schoemaker, 1982). There do seem to be several attractive features of these alternative theories that merit examination. Prospect theory has in our view the decided advantages of giving much more consideration to the specific context of decisions; the framing of questions; and the certainty effect (which results in states of the world which occur with certainty looking disproportionately large as compared with uncertain states of the world). Perhaps most important, prospect theory echoes the view of Simon (1955) that gains and losses are better or more easily viewed in terms of some

reference point, specifically emphasizing that the utility function is likely to be convex for losses and concave for gains.

In regret theory, allowance is made for the context of choice. Loomes and Sugden (1982) do not include the process utility of choosing per se but do attempt to incorporate the idea of regret or rejoicing associated with making a bad or good, respectively, decision. These approaches are worthy of closer examination especially as the context of health service decision making, particularly at the clinical level, is so far removed from the neoclassical paradigm.

There is more than health and therefore more than QALYs in the patient's utility function, whether this is seen through the patient's eyes directly or through the doctor's as the patient's agent. And there is more to the relevant social welfare function than the aggregation of individual patients' utilities from health improvements. Health economists currently may be falling into the medical trap of thinking that health services are about health only rather than some wider based notion of utility.

Emphasizing the decision-making frame, and thereby considering other avenues than expected utility theory and other arguments in the utility function than simply health, would seem to be the way to go with QALYs. Given the nature of health care for the consumer and the role of the doctor-agent, issues of information and decision making need to be considered more when defining the patient's utility function. And given the nature of health care for the citizen, issues of equity need to be reflected more in any health care social welfare function.
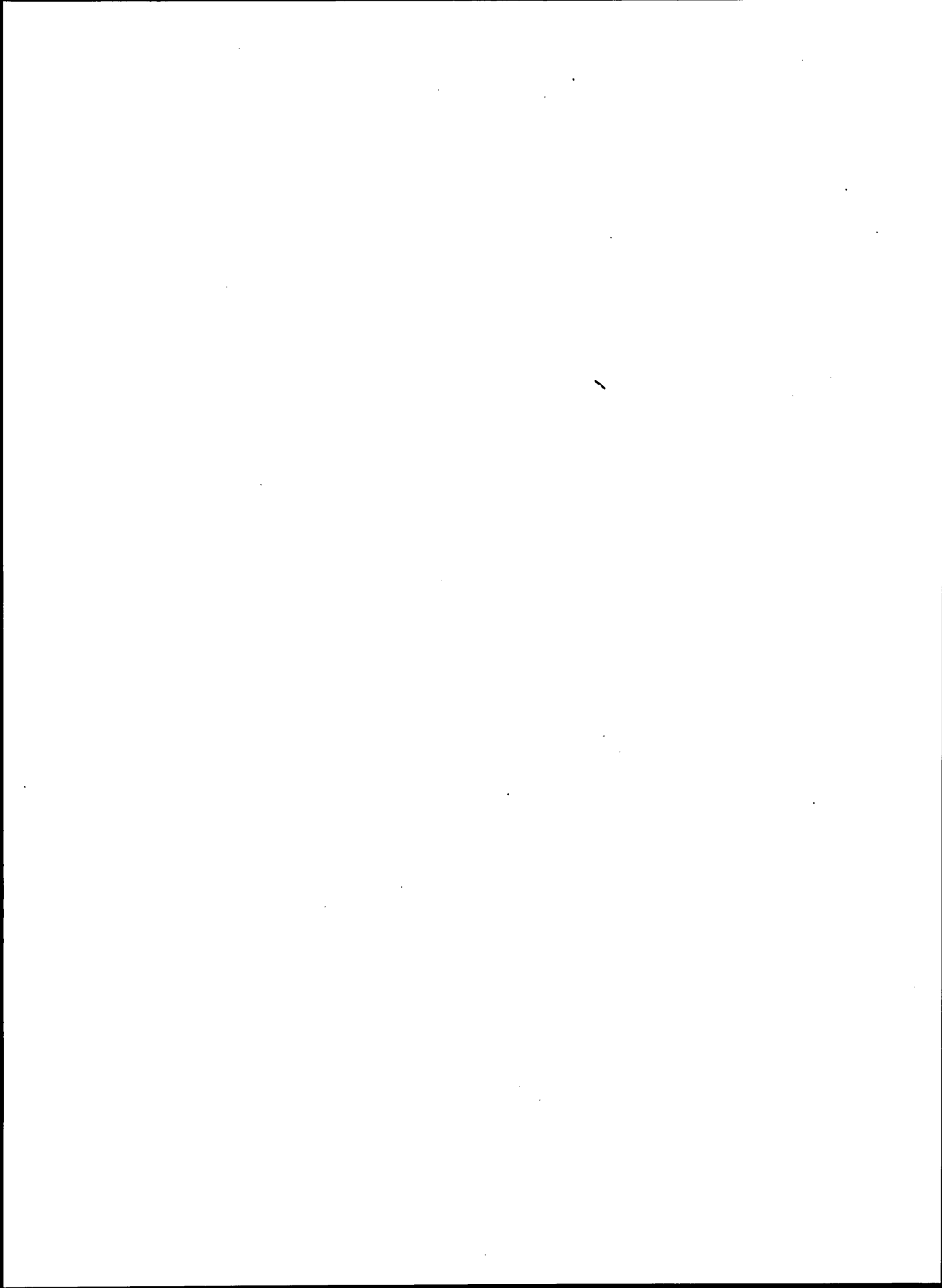
References:

Arrow, K.J. (1963), Uncertainty and the welfare economics of medical care, American Economic Review, 53, 941-73

Beauchamp, T.L. and Childress, J.F. (1983), Principles of Biomedical Ethics, Oxford: Oxford University Press.

Berwick, D.M. and Weinstein, M.C. (1985), What do patients value? Willingness to pay for ultrasound in normal pregnancy, Medical care, 23, 881-93.

Bøhm-Bawerk, E. von (1888), The Positive Theory of Capital, trans. W. Smart, G.E. Stechert and Co, 1891.

Culyer, A.J. (1980), The Political Economy of Social Policy, Oxford: Martin Robertson.

Culyer, A.J. (1988a), Health Care Expenditures in Canada: Myth and Reality; Past and Future, Toronto: Canadian Tax Foundation.

Donaldson, C. and Gerard, K. (1989), Paying general practitioners: shedding light on the review of health services, Journal of the Royal College of General Practitioners, 39, 114-7.

Dowie, J.(1989), Professional Judgment, Text 5, Milton Keynes: Open University.

Drummond, M.F., Stoddard, G.L., and Torrance, G.W. (1987), Methods for the Economic Evaluation of Health Care Programmes, Oxford: Oxford University Press.

Dyer, J.S. and Sarin, R.K. (1982), Relative risk aversion, Management Science, 28, 8, 875-86.

Evans, R.G. and Wolfson, A.D. (1980), Faith, hope and charity: health care in the utility function, mimeo, Vancouver: Department of Economics, UBC.

Feeny, D.H. and Torrance, G.W. (1989), Incorporating utility-based quality-of-life assessment measures in clinical trials: two examples, Medical Care, Supplement, 27, 3, 190-204.

Fuchs, V. (1982), Time preference and health: an exploratory study, in V.R. Fuchs (ed.) Economic Aspects of Health, Chicago: University of Chicago Press.

Gafni, A. and Torrance, G.W. (1984), Risk attitude and time preference in health, Management Science, 30, 4, 440-451.

Harsanyi, J.C. (1955), Cardinal welfare, individualistic ethics, and interpersonal comparisons of Utility, Journal of Political Economy, 63, 309-21.

Hellinger, F.J. (1989), Expected utility theory and risky choices with health outcomes, Medical Care, 27, 3, 273-9.

Hey, J.D. (1979), Uncertainly in Microeconomics, Oxford: Martin Robertson.

Kahneman, D. and Tversky, A. (1979), Prospect theory: an analysis of decision under risk, Econometrica, 47, 2, 263-91.

Keeler, E.B. and Cretin, S. (1983), Discounting of life-saving and other non-monetary effects, Management Science, 29, 3, 300-306.

Langlois, R.N. (1986), Economics as a Process, Essays in the New Institution Economics, Cambridge: Cambridge University Press.

Lipscomb, J. (1989), Time preference for health in cost-effectiveness analysis, Medical Care, Supplement, 27, 3, 233-53.

Llewellyn-Thomas, H., Dutherland, H.J., Tibshirani, R., Ciampi, A., Till, J.E., and Boyd, N.F. (1984), Desccribing health states: methodological issues in obtaining values for health states, Medical Care, 22, 543-52.

Loomes, G., McKenzie, L. (1989), The use of QALYs in health care decision making, Social Science and Medicine, 28, 299-308.

Loomes, G. and Sugden, R. (1982), Regret theory: an alternative theory of rational choice under uncertainty, Economic Journal, 92, 805-24.

McGuire, A.J. (1987), The measurement of hospital efficiency, Social Science and Medicine, 24, 719-24.

McGuire, A.J., Henderson, J.B. and Mooney, G.H. (1988), The Economics of Health Care, London: Routledge and Kegan Paul.

McNeil, B.J., Weichselbaum, R. and Pauker, S.G. (1978), Fallacy of the five year survival in lung cancer, New England Journal of Medicine, 299, 1, 397-401.

McNeil, B.J., Weichselbaum, R. and Pauker, S.G. (1981), Tradeoffs between quality and quantity of life in laryngeal cancer, New England Journal of Medicine, 305, 982-7.

Margolis, H. (1982), Selfishness, Altruism and Rationality, Cambridge: Cambridge University Press.

Mehrez, A. and Gafni, A. (1987), An empirial evalution of two assessment methods for utility measurements for life years, Socio Economic Planning Science, 21, 6, 371-5.

Mehrez, A. and Gafni, A. (1989), Quality-adjusted life-years, utility theory and healthy years equivalents, Medical Decision Making, 9, 142-9.

Mill, J.S. (1979), Utilitarism, On Liberty, and Essay on Bentham, New York: New American Library.

Mooney, G.H. (1986), Economics, Medicine and Health Care, Brighton: Wheatsheaf

Mulley, A.G. (1989), Assessing patients' utilities: can the ends justify the means?, Medical Care, 27, 3, S 269-81.

Nord, E. (1988), Prioriteringer i helsevesenet ut fra nyttevurderinger, Avd. for helsetjenesteforskning, Oslo: SIFF.

Pauker, S.G. and McNeil, B.J. (1981), Impact of patients' preferences on the selection of therapy, Journal of Chronic Diseases, 34, 77-86.

Pliskin, J.S., Shepard, D.S. and Weinstein, M.C. (1980), Utility functions for life years and health status, Operations Research, 28, 1, 206-24.

Read, J.L., Quinn, R.J., Berwick, D.M., Fineberg, H.V. and Weinstein, M.C. (1984), Preferences for health outcomes: comparison of assessment methods, Medical Decision Making, 4, 315-9.

Richardson, J., Hall, J. and Salkeld, G. (1989), Cost utility analysis: the compatibility of measurement techniques and the measurement of utility through time, paper presented to the 11th Annual Conference of the Australian Health Economists' Group, ANU, 7-8 September.

Robbins, L. (1938), Interpersonal comparisons of utility, Economic Journal, 48.

Schoemaker, P.J.H. (1982), The expected utility model: its variants, purposes, evidence and limitations, Journal of Economic Literature, xx, June, 529-63.

Simon, H.A. (1955), A behavioral model of rational choice, Quarterly Journal of Economics, 69, 174-83.

Simon, H.A. (1961), Administrative Behavior, New York: Macmillan.

Sutherland, H.J., Llewellyn-Thomas, H., Boyd, N.F. and Till, J.E. (1982), Attitudes towards quality of survival - the concept of "maximal endurable time", Medical Decision Making, 2, 3, 299-309.

Torrance, G.W. (1976), Social preferences for health states, an empirical evalution of three measurement techniques, Socio-Economic Planning Science, 10, 129-36.

Torrance, G.W. (1986), Measurement of health state utilities for economic appraisal, Journal of Health Economics, 5, 1-30.

Torrance, G.W. (1987), Utility approach to measuring health-related quality of life, Jornal of Chronic Diseases, 40, 593-600.

Torrance, G.W., Thomas, W.H. and Sackett, D.L. (1972), A utility maximisation model for evalution of health care programs, Health Services Research, 7, 2, 118-33.

Von Neumann, J. and Morgenstern, O. (1944), Theory of Games and Economic Behavior, New York: Wiley

Weinstein, M.C. and Stason, W.B. (1977), Foundation of cost-effectiveness analysis for health and medical practices, New England Journal of Medicine, 296, 31, 716-21.

Williams, A. (1985), Economics of coronary artery bypass grafting, British Medical Journal, 291, 326-9.

Williams, A. (1988a), Ethics and efficiency in the provision of health care, in J.M. Bell and S. Mendus (eds.), Philosophy and Medical Welfare, Cambridge: Cambridge University Press.

Williams, A. (1988b), Priority setting in public and private health care. A guide through the ideological jungle, Journal of Health Economics, 7, 173-83.

Williamson, O.E. (1973), Markets and hierarchies: some elementary considerations, American Economic Review, 63, 316-25.

Wright, R.A. (1987), Human Values in Health Care, New York: McGraw Hill.

Paper 3

PRODUCTION GAINS: SHOULD THEY COUNT IN HEALTH CARE EVALUATIONS?

*Jan Abel Olsen**

I INTRODUCTION

In various economic evaluations of health care, it is frequently suggested that curing productive people is of higher value to society than curing those who will not produce anything after treatment. Hence, it is often contended that this economic gain from cure should be taken into account in economic evaluations. This is an old argument in the literature on health investment (see e.g., Mushkin, 1962), the human capital approach to valuing human lives (see e.g., Rice and Cooper, 1967) and more recently in 'cost of illness' studies (see e.g., Hodgson and Meiners, 1982; Henke and Behrens, 1986). Although these branches of the literature have been the subject of much criticism (see e.g., Wiseman, 1963; Mooney, 1977; Shiell *et al.*, 1987) it appears that the distinction between 'direct costs' (i.e. health care costs) and 'indirect costs' (i.e. production losses) has survived in the current health economics literature. Similarly, a distinction is often made between direct and indirect benefits in the widely used cost-effectiveness and cost-utility analysis (CEA/CUA). Health benefits, which are normally measured in quality-adjusted-life-years (QALYs), are often referred to as the 'direct benefits', while benefits beyond health per se are referred to as 'indirect benefits' and can be defined as 'production gains to society because more people are well, or alive, and able to return to work' (Torrance, 1986). Production gains in CEA/CUA are identical with avoided production losses.

While there is considerable intuitive appeal to both economists and lay-persons in the idea that production gains should be taken into account, their inclusion has certain controversial implications for priority setting in health care in that, *cet par*, patient groups with high present values of future earnings will be given priority over other groups. This contradicts a widely held objective of most public health services, that of 'equal access for equal need independent of the patient's economic status'.

If the aim of the health service were to maximize GDP, the decision rule with respect to cure/not cure would be whether the present value of expected future production were higher than the cost of curing the patient. While this is trivial, its policy implications are rather extreme. In a thought-provoking article in this journal, Jack Wiseman (1963) gave two illustrative examples. First, it implies that one would provide euthanasia for those who can no longer work; and second, in poor countries in which malnutrition is a serious health problem,

*Department of Economics, University of Tromsø, Norway

JAN ABEL OLSEN

cannibalism could be incorporated by letting the productive 'eat those whose contribution to growth was negative' (Wiseman, 1963). His examples were deliberately chosen to be extreme to illustrate the dangers of 'single objectived' analysis.

However, the issue is not whether the analysis is 'single objectived', but rather what the objective(s) is (are), and whether the analysis correspond with the objective(s). In most public health services the objective is not to be a means to the end of maximizing GDP, but to maximize health, or perhaps to contribute to increasing welfare in general. It is widely agreed that the economic analysis which best correspond with such objectives are based on the cost-benefit approach.

Textbooks on economic evaluation of health care programmes tend either to recommend that production gains should be taken into account (e.g. Luce and Elixhauser, 1990), or they are more ambiguous (e.g. Drummond *et al.*, 1987). When it comes to empirical cost-utility studies, relatively few analysts seem to include such items. In her review of 51 cost-utility studies, Gerard (1991) found that only 6 had included indirect costs. However, the issue has recently received much attention among health economists (see e.g. Gerard and Mooney, 1993; Ratcliffe, 1993). Some experts in the field (Drummond *et al.*, 1993) hold that there is a low level of agreement in the literature concerning the inclusion of production gains and losses. According to Alan Williams (1992), British health economists have got themselves out of this dilemma through the use of an economic argument: due to the high level of unemployment in the British economy, employees who are sick can be replaced by fit unemployed people. Since GDP 'will be virtually unchanged', production gains are ignored. Williams however, leaves unanswered the question of what to do in countries which enjoy near-full employment. To my knowledge no attempt has so far been made in the literature to provide more theoretically sound recommendations to health programme analysts.

The aim of this paper is to unravel the conditions under which production gains should be included in economic evaluations of health care programmes. In doing so, a model is introduced which distinguishes between the pure production possibilities and the normative aspects. Emphasis is placed on the extent to which production gains are welfare improving, or, how the definition of the welfare function restricts the range of relevant production gains. It is shown that, when health is to be maximized (which is assumed in CEA/CUA), the relevant production gains to be subtracted from the direct health care costs is the fraction of the increased output which increases the health care budget. When the Pareto-criterion is used and only health matters, it is demonstrated that production gains are relevant if the increased health care output arising from having productive people return to work exceeds the health care needed to cure them. Further it is shown that increased production of consumption goods due to a redistribution of health care from the non-productive to the productive, represents potential Pareto-improvements as long as the output exceeds the minimal compensation claimed by the non-productive for foregoing health care.

II   Production Gains and Welfare Gains

In the context of this paper a production gain is defined as the value of the increased output—of whatever good—which is attributable to the treatment of a member of society. Beyond that, the question of whether production gains represent welfare gains will depend on the definition of the welfare function.

The structure of the general model[1]

There are basically three sets of relationships in this model: definitions and production functions, utility functions and objective functions. The focus of the model is health care, i.e. that part of the economy formally devoted to the production of health. While it is accepted that there are other goods and activities that are health promoting, for ease of exposition all of these are subsumed within the term 'health care'.

The model assumes that all individuals in society belong to one of two separate groups: 'the productive' group which produces the two marketable goods, health care and consumption goods; and 'the non-productive' group which does not produce these goods. Treatment of members of the non-productive group will therefore not lead to increased production. Further, the model assumes full employment, i.e. there are no fit unemployed to replace the sick productive. The productive group, N_P^*, represents the constraint in the model. It consists of two subgroups, those who do work and those who could work if they were fit. The first are the $\underline{F}$it and $\underline{P}$roductive members, N_P^F, who do not need health care. The second are the $\underline{S}$ick and potentially $\underline{P}$roductive members, N_P^S, who would return to work if health care were provided for them (equation (1)). When they are cured, they thereby become members of the group of fit productive.

Production, X, consists of health care, H, and consumption goods, C (equation (2)). Total health care, H, is to be allocated between the productive, H_P, and the non-productive, H_N (3); consumption goods are to be similarly allocated (4). Health care production is some fraction, β, of total production (5). Total production is a function of the number of fit and productive members (6), which in turn depends on the number of potentially productive members who are cured. So the more health care that is provided to the potentially productive, the higher becomes the number of fit productive (7), because of what follows from (1): $N_P^F = N_P^* - N_P^S$:

$$N_P^* = N_P^F + N_P^S \tag{1}$$

$$X = H + C \tag{2}$$

$$H = H_P + H_N \tag{3}$$

$$C = C_P + C_N \tag{4}$$

$$H = \beta X, \qquad 0 \leqslant \beta \leqslant 1 \tag{5}$$

[1] This section is based on a model presented in Nygaard and Olsen (1992).

$$X = f(N_P^E), \qquad f' > 0, \; f'' < 0 \tag{6}$$

$$N_P^E = g(H_P), \qquad g' > 0, g'' < 0. \tag{7}$$

Some simplifying assumptions are made about health care in the utility functions. Both the effectiveness of health care on the individual's health status (HS), $\partial HS / \partial H$, and the marginal utility from a given change in health status, $\partial U / \partial HS$, are assumed to be the same for all individuals. Thus, the utility of health care, $U_i = U_i(H_i)$, is assumed to be the same for members of both the productive and the non-productive groups. Further, it is assumed that there are no externalities in health care use.

There are two equally large patient groups; the productive, P, and the non-productive, N. Instead of viewing H_P and H_N as different amounts of health care, one can alternatively view them as expected amounts of care. Given an illness j, the $p_P H_j$ is expected care for P and $p_N H_j$ is expected care for N. If $p_P H_j > p_N H_j$, this has to be because $p_P > p_N$. Thus, we assume the same treatment for the same medical need, but the access may differ.

Equations (8) and (9) express the group aggregate of the individuals' utility functions. Decreasing marginal utility of health care is often assumed in the health economics literature. It is again here:

$$U_P = u(H_P, C_P), \qquad u_H' > 0, u_H'' < 0, u_C' > 0 \tag{8}$$

$$U_N = u(H_N, C_N), \qquad u_H' > 0, u_H'' < 0, u_C' > 0. \tag{9}$$

The analysis is restricted to just two collective objectives: the utilitarian and the Pareto-criterion. The social welfare function which is normally used in health economics is an additive function of individuals' utilities of health only, i.e. it is in the Benthamite utilitarian tradition. The objective is to maximize welfare in terms of total health utility:

$$\text{Max } W = u(H_P) + u(H_N). \tag{10}$$

While this utilitarian objective function in (10) determines one single welfare optimum, there may be a number of distributions which are Pareto-optimal. Assuming that health care is initially distributed equally between the two patient groups, the question is whether an allocation which gives more health care to the productive will be compatible with the Pareto-criterion for welfare improvement:

$$\text{Max } u(H_P) \text{ given } u(H_N) \geqslant u(H_N^I), \tag{11}$$

where H_N^I is N's initial endowment of health care. The possible range, and types, of production gains and the optimal solutions can be more easily illustrated by separating the gains in terms of first, health care and second, consumption goods. With (10) and (11) as objective functions, it follows that the production and distribution of consumption goods lie outside of our concern.

When focusing on consumption goods, equation (12) assumes that the total health care budget is given as H^0. The production gains from allowing the potentially productive to return to work are then realized solely as consumption

goods (then $\beta = 0$ in equation (5)). Due to the problems of measurement and interpersonal comparisons of utility, the idea of adding individuals' utility has less support in the welfare economics literature than has adding together health gains in the health economics literature. Hence, optimal solutions are discussed in terms of only the Pareto-criterion:

$$\text{Max } U_P \text{ given } U_N \geqslant U_N^0 \text{ and } H = H^0. \tag{12}$$

Before we turn to the analysis, some brief remarks regarding two assumptions are needed. First it is assumed that curing non-productive members who are outside the workforce has no impact on the level of production, only on health per se. However, there are situations when the ability of a productive person to work is dependent at least in part on fit non-productive people; e.g., if a housewife is sick, the husband may have to stay at home to take care of the children. Curing the wife could thus have the same impact on the production level as curing the man. Second, there is a supply side assumption in the model, as the production is constrained by the availability of fit and productive people. However, in economies where fit unemployed can replace the sick productive members, the relationship between health care to the productive and the level of production will be accordingly weaker.

Increased production of health care only

Implicit in the cost-utility methodology for evaluating health care programmes is the objective function which consists of health only, equation (10). Given the assumption that health care is the only good that affects health, then only increased production of health care can be deemed relevant.

The production possibilities

The resource constraint and the production functions give the maximum attainable health care that can be distributed between P and N—independent of their preferences and how the welfare function is defined. Inserting equations (5), (6) and (7) in (3) and using the total derivative gives an expression of how changes in health care to P affect health care to N:

$$\frac{dH_N}{dH_P} = \beta \frac{dX}{dN_P^F} \frac{dN_P^F}{dH_P} - 1 = \beta f' g' - 1 \tag{13}$$

where

$$f' = \frac{df}{dN_P^F} \quad \text{and} \quad g' = \frac{dg}{dH_P}.$$

Allocating a marginal unit of health care to the productive is to the benefit of the non-productive as long as equation (13) is positive, i.e. $\beta f' g' > 1$. In that case the increased health care output following treatment of the productive, more than outweighs the health care required for their cure. The maximum health care to the non-productive is attained when $\beta f' g' = 1$. When $\beta f' g' < 1$,

the productive will no longer 'pay their way', i.e. P's marginal contribution to health care production will be less than what is sacrificed to make them fit and productive.

The distribution possibilities of health care between P and N can be illustrated by a 'consumption possibility curve for health care', Figure 1. H_P is measured along the horizontal axis, H_N along the vertical axis. The 45° 'equity line' illustrates points of equal distribution between the productive and the non-productive, $H_P = H_N$. The curve starts at point A where all health care, now produced by the group of fit and productive, is distributed to the non-productive. When some health care is given to the potentially productive, the total amount of health care increases, since $\beta f' g' > 0$. Maximum health care is attained in point B when all is given to the productive and nothing to the non-productive.[2] In this analysis the initial distribution is at point I where the available health care is distributed equally between P and N, $H_P^I = H_N^I$.
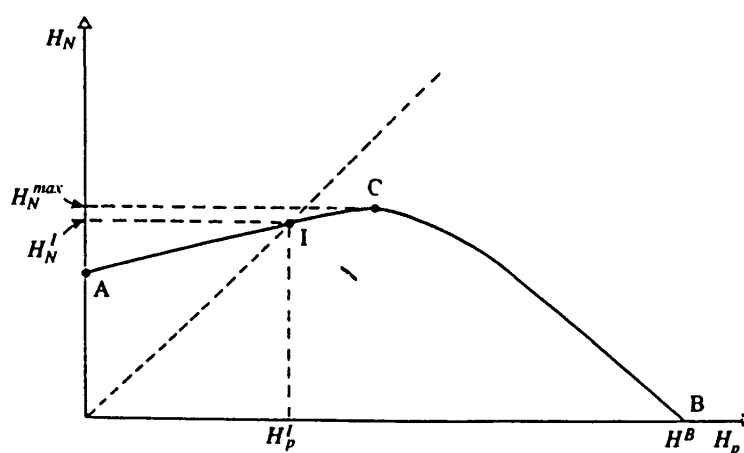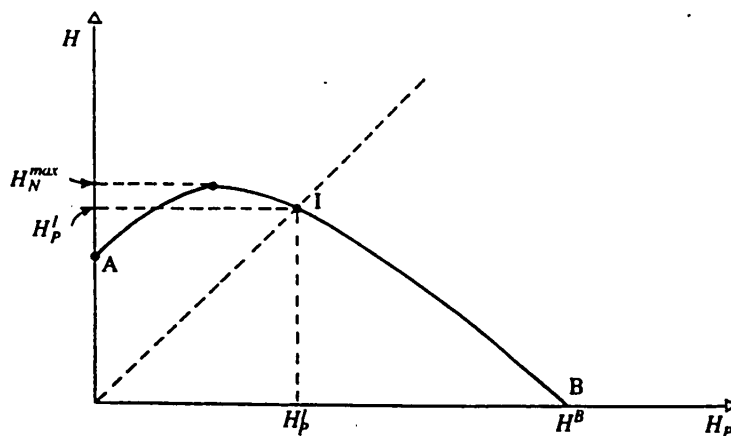
Figure 1a.

Figure 1b.

[2] If there were no increase in health care output from treating the productive, this consumption possibility curve would be a straight line from point A with a slope of -1.

The second derivative of (13) shows that H_N is a concave function of H_P. If (13) has an internal maximum, the crucial question is where this point lies compared with $H_P = H_N$, i.e. if the point of maximum attainable health care for the non-productive, H_N^{max}, lies to the right of the equity line or not. In Figure 1a it does so, while Figure 1b illustrates a situation where H_N^{max} is to the left of the point of equity. (However, H_N^{max} could also be located at the point of equity or, if no internal solution, at the corner point A.)

From Figure 1 it has been indicated that the maximum attainable health care production gains are the difference between H^B and $(H_P^I + H_N^I)$, i.e. all health care goes to the productive. But how does the welfare function restrict the range of production gains, i.e. which gains are welfare improving?

The optimal solutions

Within the context of possible Pareto-improvements, the only way that the non-productive will *not* be made worse off compared with the initial point I is if there are points on the consumption possibility curve to the North-East of I, i.e. if $\beta f' g' > 1$ at I, and, thus, H_N^{max} lies to the right of the equity line. In Figure 1a there is an area where the non-productive will not be made worse off from seeing relatively more health care allocated to the productive.

The form of Figure 1a is replicated in Figure 2. Moving East on the dotted line from I is Pareto-improving until we reach point D. If we maximize P's utility (equation (11)), point D is the Pareto-optimum. And if we maximize N's utility, point $C = H_N^{max}$ is the Pareto-optimum.[3] (All points on the consumption possibility curve between C and D are Pareto-optimal.) Thus, given that we start at I and use the Pareto-criterion, the inclusion of production gains is Pareto-improving until point D, i.e. in the interval I-D the productive group will 'pay their way' in terms of health care. Beyond point D, the increased health care production is less than what is required for their cure. Hence, production gains resulting from making the productive return to work are irrelevant and should not be taken into account. When starting at point I in Figure 1b, we note that there is no Pareto-improving production gains.

However, in the context of the utilitarian welfare function, it is not required that H_N^{max} lies to the right of the equity line for production gains to yield welfare improvements. From the welfare function in equation (10), the slope of the isowelfare curve is:

$$\frac{dH_N}{dH_P} = -\frac{U_P'}{U_N'} = -1, \text{ when } H_P = H_N. \tag{14}$$

By maximizing equation (10) with respect to H_P, we get:

$$-\frac{U_P'}{U_N'} = \beta f' g' - 1 \tag{15}$$

which of course is where an isowelfare curve is tangential to the consumption

[3] As it happens, C also represents the Rawls-solution of 'minimax'.

possibility curve (see equation (13)). For this optimum we can see that the slope must be less than unity in absolute value. Hence, point I is always inferior to the utilitarian optimum, in which more health care is allocated to the productive than to the non-productive. The optimal solution will be between C and B in Figure 2 and between I and B in Figure 1b.

Rearranging (15) to: $U_P' = U_N'(1 - \beta f'g')$, we can see that at the welfare optimum the marginal utility of health care to the productive is lower than the non-productives' marginal utility, implying that the productive receive more health care. The intuitive explanation is that the welfare gain of the productives' health care consumption is derived not only from their own health utility per se but also from the fact that their health care consumption enables them to increase the total available output of health care to society. So the greater the productives' contribution to increased total health care, the greater is the difference between the marginal utilities, which implies that the greater is the priority given to the productive ahead of the non-productive. And if the productive did not contribute to increased health care (i.e. $\beta f'g' = 0$), then the utilitarian optimum would be characterized by a point where the marginal utilities were equal, that is, on the equity line.

In Figure 2, we have introduced two isowelfare contours: that which maximizes welfare, W^1, is tangential to the consumption possibility curve at point E, which then represents the optimal allocation. At this point, the gain in terms of increased health care production equals the lost health utility because P's marginal utility of health care is lower than N's marginal utility. W^0 crosses the initial allocation at I. The difference between these two isowelfare contours is, according to the utilitarian welfare function, the welfare gain arising from the unequal distribution of health care between the productive and the non-productive.
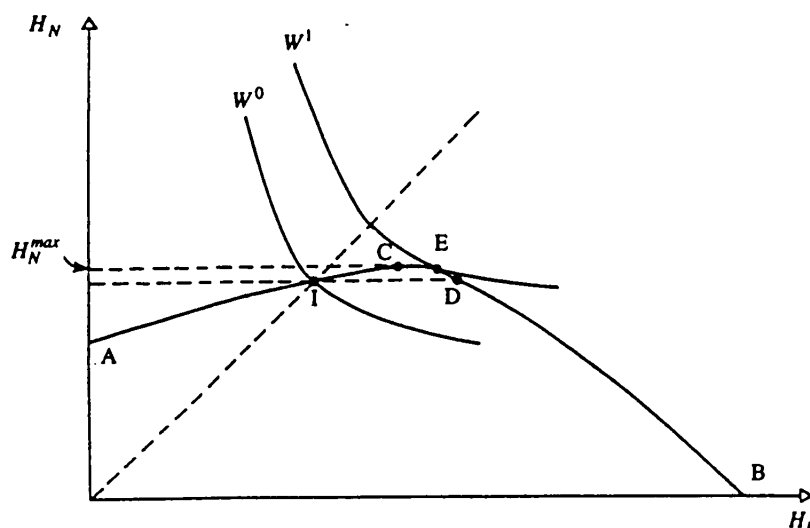

Figure 2.

Compared with the initial allocation of equal amounts of health care, the productive will always be better off with the allocation which maximizes total health utility. With a consumption possibility curve like in Figure 1b, the non-productive will be worse off, i.e. there is no Pareto-improvement. In the case of Figure 2, we cannot tell whether the non-productive will be better or worse off, because we do not know whether point E lies above or below a horizontal line through point I. Thus, we cannot tell if the 'Benthamite- optimum' in point E represents a Pareto-improvement compared with the initial allocation in point I. What we can tell, though, is that counting production gains beyond point D is not Pareto-improving, and beyond E is not welfare improving. This demonstrates how the definition of the welfare function restricts the range of relevant production gains.

It was assumed that both the effectiveness of health care on health status, $\partial HS/\partial H$, and the utility of health, $\partial U/\partial HS$, be the same for the productive and the non-productive. Relaxing these assumptions could move the optimal solutions in either direction. Since the non-productive are liable to be sicker, the effectiveness of health care on their health status may be higher. In that case the marginal utility of a given treatment is higher for the non-productive than for the productive; $\partial U_N/\partial H_j > \partial U_P/\partial H_j$. That would alter the isowelfare contours in Figure 2 so that the ulitarian optimum would shift to the left of point E on the consumption possibility curve.

But this point will take us in the opposite direction if the productive gain utility from working. Then their capacity to benefit from treatments which enable people to return to work, is higher than that of the non-productive. When a $\partial U_P/\partial H_j > \partial U_N/\partial H_j$, the optimal solution will be to the right of point E. Note that this is a utilitarian argument for prioritising the productive, which is a different argument from that of increased output.[4]

In this section, we have followed the health care evaluation tradition, in which the objective function includes health only. Increased output of consumption goods has not been considered as production gains, simply because such goods are assumed not to affect health. However, although the non-productive may be made worse off in terms of health care, they may be willing to accept compensation in terms of increased consumption.

Increased production of consumption goods only

To simplify the following analysis, it is now assumed that the total production of health care is fixed, which could be explained by public budgetary inertia or shortage of skilled health care personell in the short run. Then, how should the health care budget be allocated between the two groups if the productive

---

[4] Such considerations influenced the priority setting made by the Chief Consultant in the orthopaedic department in one of the biggest hospitals in Norway. When waiting time was nearing the critical time for those on sick leave to have a reasonable chance of getting their job back, they were prioritised ahead of the non-productive. He firmly rejected the suggestion that this was due to considerations of GPD-gains. Rather, it followed from 'taking account of people's satisfaction from working'.

produce only consumption goods? Because most people make trade-offs between health and consumption goods, it is relevant to analyse the extent to which the increased production of consumption goods, made available by real-locating health care from the non-productive to the productive, is large enough to compensate the non-productive. The question is formulated as in equation (12). From the assumption $H = H^0$, it follows that at the margin $\beta = 0$. By maximizing equation (12) with respect to H_P, C_P, H_N and C_N we get:

$$\frac{\dfrac{\partial U_P}{\partial H_P}}{\dfrac{\partial U_P}{\partial C_P}} = \frac{\dfrac{\partial U_N}{\partial H_N}}{\dfrac{\partial U_N}{\partial C_N}} - f'g'. \tag{16}$$

The Pareto-optimal allocation is characterized by different marginal rates of substitution. The higher the product of the two marginal productivities, the bigger will this difference be. Assuming equal initial distributions of H and C and identical preferences, P will end up consuming more health care and fewer consumption goods as compared with N.

An interesting situation emerges when considering an Edgeworth Box, Figure 3. Given that the initial allocation, I, is where $H_P = H_N$, then by re-allocating health care from the non-productive, N, to the productive, P, the horizontal side increases by ΔC so that more consumption goods are available in total. Hence, the distribution of the given H^0 between the two consumers affects the total endowment of the other good, C.

Figure 3 illustrates the situation in which C_p is held constant at C_P^I. Then, N is willing to exchange H_N for C_N as long as the increased consumption made

Figure 3.

available by giving health care to P exceeds the minimal compensation. The optimal solution is illustrated at point B^1 which N considers equally as good as the initial combination at I. This can be seen from N's two indifferences curves: the curve U_N^0 based on the initial origin O_N represents the same combinations for N as does the curve U_N^1 based on the new origin O_N^1. [The horizontal distance between the two curves equals ΔC, i.e. the production gain which has gone to N. It is their compensation for giving up $(H_N^I - H_N^B)$. Since N was indifferent between B and I and B^1 represents the same commodity bundle for N as B did, then N must also be indifferent between B^1 and I. The welfare gain in terms of increased utility has gone solely to P, who has moved from the initial utility level U_P^0 to U_P^1, consuming the same C_P^I but more of H: H_P^B rather than H_P^I.

Note that the optimal allocation at point B^1 does not represent any technical boundary, beyond which production gains cannot be pushed further. Rather, it represents a point where the increased production of consumption goods will not be large enough to compensate the non-productive for foregoing more health care. Thus, further reallocations of the given health care resources to the productive would mean that the production gains were not within the range prescribed by the Pareto criterion for welfare gains. However, the range of welfare improving production gains is likely to be restricted even further than this, because the externalities in health care consumption suggest that society values an individual's health care relative to her/his consumption of other goods higher than what the individual consumer does.

In most European countries the redistribution of income from the productive to the non-productive, as well as the distribution of health care, are governed by public institutions. Transfer payments represent the mechanism for compensating the non-productive. In Norway and Sweden (and I believe other countries as well) there has recently been much political focus on the idea of having 'the sickness fund' buy treatment for those reported sick who will return to work after treatment. The next section attempts to adapt the above model to such a context in which a public institution regulates the distribution of health care and sickness benefits. This is not to confuse transfer payments and real resource costs.

III   PRODUCTION GAINS TO THE REST OF SOCIETY[5]

From the perspective of a health sector budget holder, production gains would be seen as indirect benefits to society—an externality which lies outside one's concern. But if a health authority were to pay sickness benefits, some of these intersectoral externalities could be internalized. To what extent would productive patients then be prioritized ahead of the non-productive? If the budget is to be spent on health care and social security, then social security savings arising from allowing the productive to return to work can be spent on health care.

---

[5] The model is outlined in the Appendix.

Consider the situation in which both groups of non-productive and potentially productive people get the same per capita support before cure. After cure the former group represents the same 'budget burden' as before, whereas the latter live from their own income. Assume that the authority's expenditures are financed through an earmarked income tax. If the extra health care resources arising from the social security savings and the increased tax-base, exceed the costs of curing a productive person, then the net gain can be spent on increasing health care for the non-productive. If this is the case, the appropriate shape of 'the health care budget possibility' curve would be as in Figure 1a.

However, in addition to the gain from increased health care there are indirect benefits to society arising from the productives' increased consumption. If the institution chooses to take account of the productives' utility from increased consumption (the second term on the RHS of equation (A12) in the Appendix), that involves an even more unequal distribution of health care in favour of the productive. If there is no positive effect on the rest of society of productive people's own consumption, then increased C_P would not make non-productive people voluntarily give way in terms of health care use. What will affect their utility in the sense that they have an incentive to give way to productive people's access to health care, is that this latter group finance the health care through taxes and thereby can increase the budget, and that the social security savings can be spent on increased health care production. Note that the social security savings will be the same no matter whether a productive person dies or returns to work. Thus, for life saving treatments, the only variable that matters to the rest of society with respect to production gains is their net contribution in terms of taxes.

In the case of unemployment, sick productive people could be replaced by fit unemployed. If this institution also pays unemployment benefits, and these were the same per capita as sickness benefits, then there will be no savings on the social security part of the public budget which can be transferred to health care. Furthermore, as soon as the sick productive person is replaced, the temporarily lost tax revenue will be made up. Thus, the relevant production gain to the rest of society is the avoidance of extra social security payments and lost taxes during the 'friction period' until the sick productive person is replaced. This conclusion accords with the approach adopted by Richardson (1991); that the source of the potential compensation to the rest of society lies in the person's net-contribution beyond own consumption.

IV DISCUSSION

The allocation of health care that gives the maximum attainable production gains occurs when all health care is distributed to the productive, and nothing to the non-productive. This is the implication of those analyses which Wiseman (1963) called 'single objectived'. However, this extreme distribution does not accord with any commonly applied or adopted welfare function.

This paper has demonstrated how the definition of a welfare function

restricts the range of welfare improving production gains. First, it was assumed that only health matters. As the total production of health care depends on the number of fit and productive people, there is a production gain from curing the potentially productive. Applying the utilitarian objective function of maximizing the sum of utility from health, there are welfare gains up to the point where the extra health gain arising from the increased availability of health care equals the loss due to unequal marginal utilities of health care between the productive and the non-productive. Hence, when the CEA/CUA approach is used, the increased health care output which might arise from treating potentially productive people will then be the opportunity cost which are to be subtracted from the direct health care costs of curing the productive. The implication is that, *cet par*, productive people are prioritised ahead of non-productive, simply because the net health care costs of treating productive people are lower.

In the context of Pareto-improvements, and given the assumption of equal amounts of health care initially, the condition for welfare gains is that the increased health care output arising from allowing potentially productive people to return to work exceeds the health care needed to cure them. The productive should only be given priority ahead of the non-productive for so long as they can 'pay their way' in terms of financing extra health care, i.e. when the net health care costs are non-positive.

It is hard to suggest a general figure for that part of the total production gain that will go to health care. However, one might get a rough idea by considering the proportion of GDP spent on health care and the elasticity by which health care spending increases compared with a 1% increase in GDP. The most recent OECD-figures indicate that the average health care spending in OECD countries is 7·3% of GDP and the elasticity is 1·3 (Schieber and Poullier, 1990). Thus, out of any marginal production gain, roughly 10% ($\approx 7\cdot3\% * 1\cdot3$) will be spent on increased health care production. If this represents the marginal fraction spent on health care from a given increased production, then the appropriate magnitude of the 'indirect benefits' which are to be subtracted in a CEA/CUA would be 10% of the value of the possibly increased production.

In the second version of the model, in which increased production was in terms of consumption goods only, it was shown that although the non-productive may be worse off in terms of health care, they can be compensated by more consumption goods. If health care is redistributed from the non-productive to the productive, then as long as the increased production of consumption goods exceeds the minimal compensation claimed by the non-productive for foregoing health care, there are potential Pareto-improvements, and thus relevant production gains. However, it was suggested that the relevant production gains would be smaller than that indicated by this free exchange solution, if there are positive externalities in individuals' health improvements.

Although the production gains are big enough potentially to compensate the non-productive, this does not mean that compensation will actually take place. In countries where the distribution of health care and the redistribution of

income are publicly regulated, the non-productive's conceivable source of compensation would be from public budgets. If there is no compensation in terms of direct transfers of consumption goods from the productive to the non-productive, then consumption gains to the productive are irrelevant to the non-productive. The relevant source of compensation is the net contribution beyond the productive's own consumption.

This paper has used a static model. In reality, the production gains will of course arise after the sick productive people have been cured. Hence, there is at least a two-period setting; sick and unproductive before treatment, and fit and productive after treatment. While this issue could also be discussed using a two-period model, in this paper we have chosen a static cost-benefit framework, simply because this approach has the closest links to the framework adopted in economic evaluations of health care programmes (see Drummond *et al.*, 1993).

It is tempting to agree with Williams (1992) that in the case of unemployment, production 'will be virtually unchanged', and production gains can therefore be neglected. However, there are many firms operating in economies with unemployment that still choose to offer free occupational health services for their employees. An obvious reason is that of high replacement costs. If the value of the lost output during absenteeism exceeds the costs of preventing it happening, or the costs of cure which reduces absenteeism, then a profit-maximizing firm would provide its own internal health services, simply because such services contribute to increased profit. If the sphere of the health service devoted to occupational therapies comes as a supplement to a public health service, then the productive will be made better off without the non-productive being made worse off. However, interestingly there is no reason for undertaking CEA/CUAs in this sphere, because the aim of this type of health service is not concerned with health per se.

APPENDIX

Equations (A1) and (A2) indicate three groups of people; the non-productive, N_N^*, the sick potentially productive, N_P^S and the fit productive, N_P^F. The public budget, G, is to be spent on health care and social security, S, (A3). Social security savings arising from making the productive return to work are to be spent on health care. (A4) suggests that the social security expenses are a product of a given per capita social support, $\bar{s}$, and number of recipients. G is financed by a proportional income tax, (A5). The total consumption for the group of productive people is the sum of social security received by the sick productive and the net income for the fit productive, (A6). The production and utility functions (A7)–(A10) are similar to those in Section II, except for the given consumption level, $\bar{s}$, for the non-productive. (A11) differs from the utilitarian function in Section II in that it includes utility from consumption:

$$N = N_N^* + N_P^* \tag{A1}$$

$$N_P^* = N_P^F + N_P^S \tag{A2}$$

$$G = H_N + H_P + S \tag{A3}$$

$$S = \bar{s}(N_P^S + N_N^*) \tag{A4}$$

$$G = \bar{i}X, \qquad 0 < \bar{i} < 1 \tag{A5}$$

$$C_P = N_P^S \bar{s} + N_P^F[(1 - \bar{i})X] \tag{A6}$$

$$X = f(N_P^F), \qquad f' > 0, f'' < 0 \tag{A7}$$

$$N_P^F = g(H_P), \qquad g' > 0, g'' < 0 \tag{A8}$$

$$U_P = u(H_P, C_P), \qquad u_H' > 0, u_H'' < 0, u_C' > 0 \tag{A9}$$

$$U_N = u(H_N, \bar{s}), \qquad u_H' > 0, u_H'' < 0 \tag{A10}$$

$$\text{Max } W = U_P + U_N. \tag{A11}$$

Maximizing (A11) with respect to H_P, H_N and C_P gives the condition for the optimal allocation of the public health care budget:

$$\frac{\partial U_P}{\partial H_P} = \frac{\partial U_N}{\partial H_N} (1 - \bar{i}f'g' - \bar{s}g') + \frac{\partial U_P}{\partial C_P} [(1 - \bar{i})f'(g')^2]. \tag{A12}$$

ACKNOWLEDGEMENTS

REFERENCES

DRUMMOND, M. F., STODDART, G. L. and TORRANCE, G. W. (1987). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford.

DRUMMOND, M. F., BRANDT, A., LUCE, B. and ROVIRA, J. (1993). Standardising methodologies for economic evaluation in health care: Practice, problems, and potential. *International Journal of Technology Assessment in Health Care*, 9, 1, pp. 26–36.

GERARD, K. (1991). A review of cost utility studies: assessing their policy making relevance. HERU Discussion Paper 11–91. University of Aberdeen.

GERARD, K. and MOONEY, G. (1993). QALY league tables: handle with care. *Health Economics*, 2, pp. 59–64.

HENKE, K. D. and BEHRENS, C. S. (1986). The economic cost of illness in the Federal Republic of Germany in the year 1980. *Health Policy*, 6, pp. 119–143.

HODGSON, T. and MEINERS, M. (1982). Cost of illness methodology: a guide to current practices and procedures. *Milbank Memorial Fund Quarterly*, 60, pp. 429–462.

LUCE, B. R. and ELIXHAUSER, A. (1990). *Standards for Socioeconomic Evaluation of Health Care Products and Services*. Springer Verlag, Berlin.

MOONEY, G. (1977). *The Valuation of Human Life*. Macmillan, London.

MUSHKIN, S. J. (1962). Health as an investment. *Journal of Political Economy*, LXX 5, 2.

NYGAARD, E. and OLSEN, J. A. (1992). A model for discussing the consequences of
    including production gains in economic evaluation of health care programmes.
    Working paper, Folkehelsa, Oslo.
RATCLIFFE, J. (1993). The measurement of indirect costs in economic evaluation: a
    critical review. Paper presented to the UK Health Economists' Study Group
    meeting, University of Strathclyde, Glasgow, 30th June–2nd July.
RICE, D. and COOPER, B. (1967). The economic value of human life. *American Journal
    of Public Health*, 57, pp. 1954–66.
RICHARDSON, J. (1991). *Economic Assessment of Health Care: Theory and Practice.*
    The Australian Economic Review, pp. 4–21.
SCHIEBER, G. J. and POULLIER, J. P. (1990). Overview of international comparisons of
    health care expenditures, *OECD Social Policy Studies*, 7.
SHIELL, A., GERARD, K. and DONALDSON, C. (1987). Cost of illness studies: an aid to
    decision-making? *Health Policy*, 8, pp. 317–323.
TORRANCE, G. W. (1986). Measurement of health state utilities for economic appraisal.
    Journal of Health Economics, 5, pp. 1–30.
WILLIAMS, A. (1992). Cost-effectiveness analysis: is it ethical?, *Journal of Medical
    Ethics*, 18, pp. 37–42.
WISEMAN, J. (1963). Cost-benefit analysis and health service policy. *Scottish Journal of
    Political Economy*, 10, pp. 128–45.

SMALL RESEARCH GRANT SCHEME

Applications are invited from members of the Scottish Economic Society
for grants of up to £1,000 to assist with research expenses. Requests for
grants for other scholarly purposes, e.g. to cover publication costs, will
be entertained, but in general preference will be given to research appli-
cations. There will be no limitations on the areas of research involved,
having regard to the aim of the Society to advance the study of economic
problems "on the widest basis, in accordance with the Scottish tradition
of political economy inspired by Adam Smith"; but should selection be
necessary, preference may be given to research on Scottish topics, other
things being equal.

Application for grants should be made by 30th April and 30th
November in each year. Application forms and further information may
be obtained from the Secretary of the Society, Mr F. G. Hay, Depart-
ment of Political Economy, University of Glasgow, G12 8QQ.

Paper 4
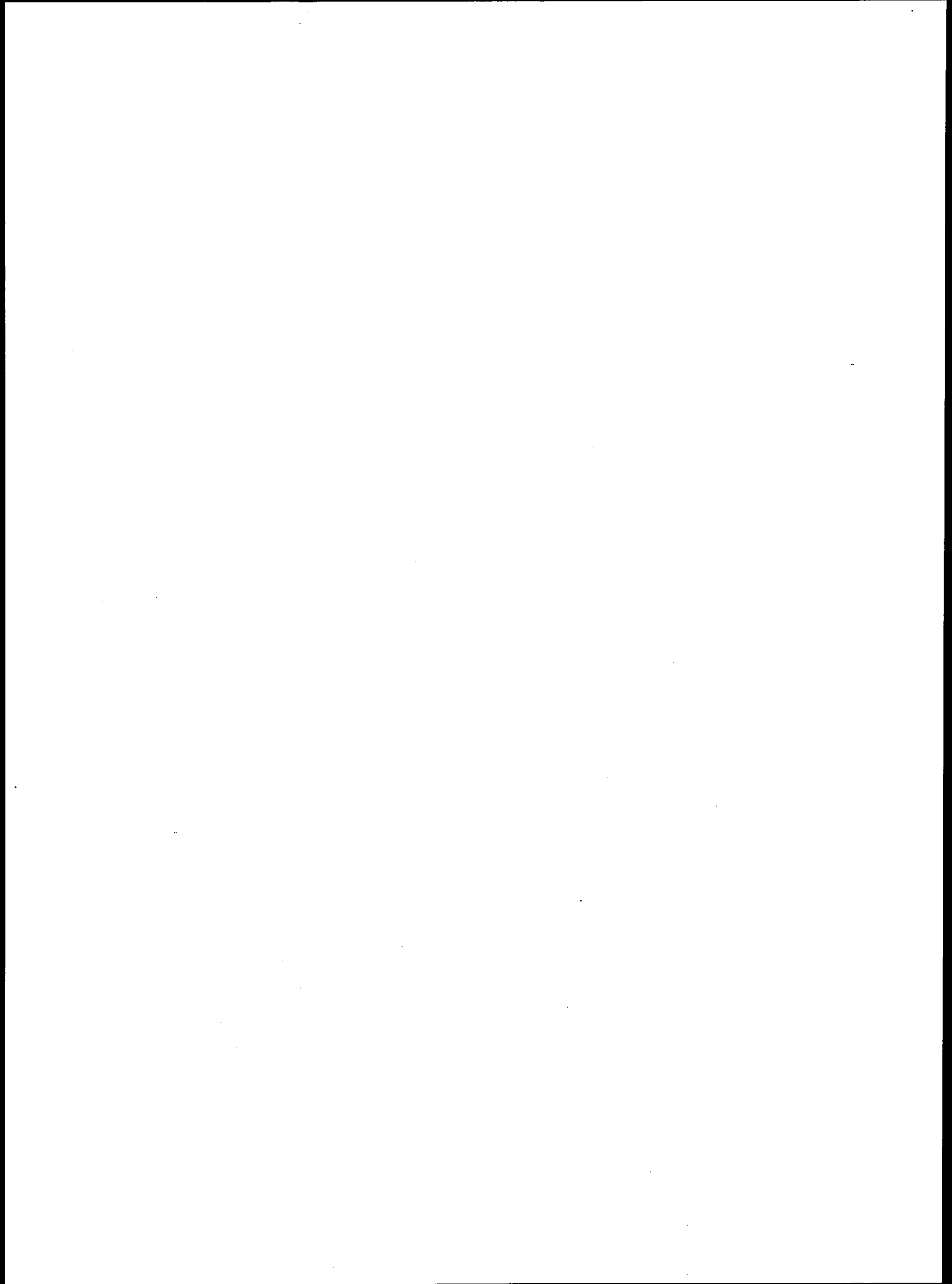
On what basis should health be discounted?

Jan Abel Olsen*

Department of Economics, NFH, University of Tromso, Tromso, Norway

The discount rate to be applied in health care programmes should be based on the time preference rate for health, and this same rate should be applied to costs as well. Due to the limited tradeability of health, when eliciting its time preference rate, the intertemporal choices must be framed in such a way as to resemble as closely as possible those facing health planners and decision makers.

1. Introduction

Discounting future costs and benefits to their present value is an essential feature in project evaluation. In economic evaluations of health care programmes there is, however, no consensus on the assigning of lower values to lives, life-years or QALYs gained in the future compared to those gained in the present (see Robinson, 1990). It is often contended that health benefits should be left undiscounted or be discounted at a lower rate than monetary items (see e.g. Parsonage and Neuburger, 1991).

The more the time profiles of the health benefits differ, the more important is the discount rate for the recommended ranking between competing programmes, and thus for priority setting in health care. Ceteris paribus, the higher the discount rate, the lower the priority given to the health of future generations and to preventive care. A study of the prevention of heart disease (Cretin, 1977) illustrates the point. Screening children for excess cholesterol concentration involved a programme in which costs were immediate, but health benefits delayed for 40 years or more. For 10-year-old boys, the costs per life-year saved were $ 67,000 when discounting both costs and benefits at

10%, $ 9,400 when discounting at 5%, and only $ 1000 if only costs were discounted at 5% and benefits were left undiscounted.

Although Fuchs and Zeckhauser (1987) express the rather strong view that: 'Self-respecting economists should not ... use different rates because it is health that is being valued', the health economics literature has as yet no unique solution or recommendation for handling this discounting issue. Drummond et al. (1987) state that the selected discount rate in the methodology of cost effectiveness- and cost utility analysis (CEA/CUA) should either be consistent with 'current practice' or be the government recommended rate. They suggest neither a lower discount rate in the health sector, nor that health benefits should be discounted at a lower rate than the costs. In a recent paper on standardising the CEA/CUA-methodology, Drummond et al. (1993) contended that there is low agreement on the choice of the discount rate.

In an empirical study by Moore and Viscusi (1990), the key questions raised were whether 'rates of time preference for health differ from the interest rates for trading financial resources and, if so, what is the nature of the discrepancy?'. Their study revealed an implicit time preference rate for health of 2%, which was consistent with financial market rates for the period. Concluding that they found 'no clear evidence of systematic differences between discount rates for health and financial rates of time preference', the second issue was then not discussed. In his comment on their paper, Portney (1990) did not 'place much stock in this finding' but believed it was more an artifact of the data and the estimation techniques they had used. A different model applied on the same data set gave a time preference rate of 11% (Viscusi and Moore, 1989). Portney argued that the rate at which non-monetary benefits and costs should be discounted is a question which is 'not at all moot'.

The underlying view in this paper is that it is a priori at least possible that there are differences between our time preferences for health compared to those for ordinary consumption goods. Further time preferences for *own* health may well differ from those, which individuals – in the role of citizens – have for *social* intertemporal choices of health.

The paper has two aims. First it attempts to answer the interrelated theoretical questions that have bedeviled health programme analysts for some time: Should health be discounted at a different rate from costs? And should health care programmes be discounted at a different rate from that used in other sectors of the economy? The standard argument in the CEA/CUA literature is to say *no* to the first question for reasons of consistency. However, drawing on the general economics literature on discounting, Section 2 shows that while the conclusion remains the same, the argument is different. Further, this literature suggests that the only theoretically legitimate argument for proposing a different discount rate in health care programmes

is that the time preference rate for health (TPR_H) is different from the time preference rate for consumption goods (TPR_C). Before embarking on empirical research to elicit the time preference rate for health, economists would require sound theoretical reasons for assuming that intertemporal preferences for health are different. Section 3 discusses this question.

The second aim is to draw up a research agenda for eliciting the TPR_H. The focus is on some important lessons which should be taken into account before doing empirical research. The limited tradeability of health creates considerable problems in eliciting time preference rates from individuals' behaviour and makes it difficult to translate an implicit rate revealed in one context to a different one. Therefore, it is argued that what is required is to frame decision contexts which reflect accurately and closely the nature of the issue; that is, how health programme analysts should weight health benefits accruing at different points in time. Since individuals do not seem to discount the future at a constant rate independent of the length of the time period involved, it is also concluded that instead of aiming for one single TPR_H, a non-monotonic time preference function may be more appropriate.

2. Discounting health differently?

If health is to be discounted at a different rate, the question immediately arises: different from what? There are two separate issues here. The first is whether the denominator – in terms of non-monetary health – should be discounted at a different rate from the numerator – in terms of resources. The second issue is whether to apply a discount rate in health care programmes (but the same for the benefits and costs) which is different from that used in other public sectors.

2.1. Discounting health benefits differently from costs

The 'classic' argument in favour of applying the *same rate* to both health benefits and costs is as follows. Since health benefits are being valued relative to monetary units, then for the sake of consistency, they have to be discounted at the same rate (see e.g. Weinstein and Stason, 1977). Within the framework of intertemporal budget allocation, Keeler and Cretin (1983) show the paralyzing effect on decision makers if the benefits are left undiscounted: 'For any attractive program, there is always a superior delayed program which should be funded first. The result is that no program with a finite starting point can be selected.' Of course, the same *paralyzing paradox* applies when using a lower discount rate than that used for costs.

The paralyzing paradox is sometimes used in the health economics literature as 'the' grand argument for not applying a different discount rate for health. That is too hasty. First, Keeler and Cretin themselves stress that

the paralyzing paradox relates to budget allocation over time. It was not their contention that undiscounted benefits are meaningless as such. Second, their argument does not require that the discount rate applied in health care programmes be the same as the rate used elsewhere, but simply that the same rate has to be applied to health as to costs.

Parsonage and Neuburger (1992) hold that the Keeler & Cretin-argument becomes irrelevant in the context of setting priorities based on CEA/CUA. They argue that cost per QALY measures are not used for making timing decisions, but for selecting projects from a ranked list, when the budget is restricted by other criteria. Furthermore, when health planners have no scope for deferring current funds to future periods, it cannot be paralyzing for the decision makers to rank alternative projects according to costs per undiscounted QALY.

When the matter at hand is the ranking of health care programmes within a constrained current budget, it is easy to agree with the Parsonage & Neuburger-view that one *could* apply a different discount rate in the denominator from the one used in the numerator of CEA/CUA. However, the lessons from the general literature on discounting suggest that one *should* not.

2.2. Lessons from the literature on discounting

There are two sources for a positive discount rate. First, the social opportunity cost of capital (SOC) reflects the marginal rate of return in the private sector, adjusted by a risk premium (Arrow and Lind, 1970), while second time preference reflects the rate at which consumers claim compensation for delaying consumption or, more precisely, the marginal rate of substitution of consumption next year for consumption this year. The literature assumes that, due to tax distortion and risk, the TPR is lower than the SOC. On which of these two should the social discount rate be based? And to what extent should this *social* discount rate be lower than the *private* discount rate?

In his seminal article, Baumol (1968) provided no very definite conclusion in favour of either the TPR or SOC. But given the *tangible loss* in transfering resources from a high rate of return use to one in which their yield is low, Baumol saw an argument for choosing the SOC as the social discount rate. For health it is hard to use the same argument if the TPR_H is different from the SOC, simply because health benefits by definition are 'intangibles' in CEA/CUA, and thus not measured on the same scale.

In an elaboration of Baumol's ideas, Sandmo and Dreze (1971) concluded that the public sector's discount rate should be a *weighted average* of the rate facing consumers and the tax-distorted rate used by firms. The weight would

depend on the proportions of funding drawn from consumption and from investment. Such an approach would give *different discount rates* to the extent that a project displaced more or less consumption as compared with investment.

If health sector funding displaces consumption only, the appropriate discount rate would be the consumers' time preference rate. This rate, however, can be derived in different ways. The first is the rate consumers face in financial markets. But according to Lind (1990) and the behavioural school of thought (see Section 4 below), this rate may tell us nothing about people's rates of time preference. A second interpretation would be the time preference rate for the consumption goods that are being displaced, and a third would be the time preference rate for the good that is to be produced. If we choose the last interpretation, then the appropriate discount rate for health care programmes emerging from this 'weighted average' approach would be similar to that from the 'shadow price of capital' approach.

There is wide consensus now that theoretically the 'shadow price of capital' approach is most defensible (JEEM 18S 1990). Since it is the flow of consumption that is important to welfare, it is the time preference for these benefits that matters (see e.g. Feldstein, 1964; Lind 1982, 1990). All costs and benefits are converted to consumption equivalents and a single rate of discount is then applied, namely the consumer's rate of interest (the TPR). As the consumer's time preference rate varies, the appropriate discount rate will then be the situation-specific time preference rate. The core of the approach is that all relevant items be discounted at the *same rate* based on the intertemporal preferences for the particular good that is to be consumed. In the health care context, this would mean that *both* non-monetary health benefits and costs should be discounted at a rate based on the time preference for health.

Consequently, the key issue is whether to base the social discount rate on the SOC, the TPR, or some weighted average of the two. The issue is *not* whether to apply a different rate on the cost side (e.g. the SOC) from that used on the benefit side (e.g. the TPR). The unambiguous lesson for health programme analysts is that no matter which approach to discounting is adopted, each implies that health benefits should be discounted at the *same rate* as costs, which is the same conclusion as that from the consistency argument in the health economics literature.

While the conclusion is the same, the arguments differ. Let us assume that the recommended discount rate is based on the TPR_C, and that costs are discounted at this rate. According to the consistency-argument, health benefits must then be discounted at this same rate. In CEA/CUA, where benefits are in terms of health only, the 'shadow price of capital' approach suggests that the TPR_H is the appropriate rate on which to base discounting in health care programmes. Given that $TPR_H \neq TPR_C$, the discount rate

implied by the 'shadow price of capital' approach becomes different from the rate implied by the consistency argument.

Hence, the only theoretically legitimate reason for applying a different discount rate in health care programmes is that the time preference rate for health is different. While this is entirely an empirical question, economists would require sound theoretical reasons to believe in such a discrepancy before embarking on empirical research.

3. Time preference – 'An essential attribute of human nature'

This section suggests that some important intertemporal preferences for health are of a different nature from those for wealth. It also shows that the interpretation and use of the time preference concept in a health context often depart from its precise meaning in economic theory.

3.1. The time preference for wealth

A hundred years ago Böhm-Bawerk claimed: 'to goods that are destined to meet the wants of the future, we ascribe a value which is really less than the true intensity of their future marginal utility' (Böhm-Bawerk, 1888). von Mises contended that a positive time preference is 'an essential attribute of human nature' (The New Palgrave, 1987). Pigou (1932) held that: 'everybody prefers present pleasures or satisfactions of given magnitude to future pleasures or satisfactions of equal magnitude, even when the latter are perfectly certain to occur. But this preference for present pleasures does not – the idea is self-contradictory – imply that a present pleasure of given magnitude is any *greater* than a future pleasure of the same magnitude.'

When individuals reveal their time preference, they express the aggregate of three distinct behavioural variables; (1) the *pure* time preference rate, which refers solely to remoteness in time and, thus, reflects the degree of impatience, (2) the rate at which the marginal utility of the increased future consumption diminishes, and (3) uncertainty. The concept of *pure* time preference is taken from Olson and Bailey (1981) (O&B) who used the symbol η. They explained it by considering a household with *identical* consumption in two periods and *no risk* involved. The household would then be willing to pay some positive interest rate, η, to increase consumption in the initial period. Due to diminishing marginal utility of consumption, the interest rate will increase if future consumption exceeds present; $r > \eta$.

This distinction between the *pure* time preference and that which arises as a result of diminishing marginal utility of the expected increase in future consumption, becomes clearer within a framework applied by Leif Johansen (1967). He considered two streams of payment X^1 and X^2, which are

identical but for the fact that a certain amount, $Y > 0$, is deferred one period in X^2 as compared with X^1:

$$X^1 = (X_0, ..., X_{i-1}, X_i, X_{i+1}, ..., X_m)$$

$$X^2 = (X_0, ..., X_{i-1}, X_i - Y, X_{i+1} + Y, ..., X_m)$$

'The impatience axiom' states that $X^1 \, P \, X^2$, which implies that the discounted utility of X^1 is higher than that of X^2, $DU(X^1) > DU(X^2)$, i.e. we prefer money or consumption goods sooner rather than later. However, if the streams of payment involve increased consumption over time, then a preference for advancing some fraction of the consumption in a future period may also be due to decreasing marginal utility of consumption in each period. Hence, within the framework applied by Johansen (1967), the consumption rate of interest, σ, depends on the *pure* time preference rate, η (Johansen used the term 'subjective interest rate' and the symbol, r), and the rate at which marginal utility decreases, β. β is a product of the elasticity of the marginal utility, $\tilde{u}$, and the consumption growth rate, G. Thus:

$$\sigma = \eta + \beta + \eta\beta \approx \eta + \beta. \tag{1}$$

Although Pearce and Turner (1990) use different symbols, the σ in Equation (1) has a parallel in what they call the *social time preference rate*. The similarity with O&B can be seen by considering a two-period case with constant consumption; then $\sigma = \eta$. If consumption in the future period were to exceed present consumption, then the difference between r and η would be the second term of Equation (1), i.e. β.

The third variable which affects individuals' preferences for the timing of their consumption is uncertainty, u. While both σ and η are assumed to be positive, O&B suggest there is more ambiguity with respect to the sign for uncertainty. It creates an incentive for present consumption in the event of an early death, but it can also 'create an incentive to save more to provide undiminished consumption in the event of an unexpectedly long life'.

However, a more commonly held view is that the existence of uncertainty increases the desire for present consumption, mainly due to what Feldstein (1964) called 'a quite rational fear of death'. We assume that the time preference rate, as revealed by individuals, has the following functional form:

$$TPR = f(\eta + \beta, u). \tag{2}$$

3.2. The time preference for health

While time preference is the conventional way of describing individuals'

choices of consumption goods over time, Gafni and Torrance (1984) (G&T) apply the concept 'risk attitude' to describe intertemporal choices of health. This risk attitude embodies three effects: a time preference effect, a quantity effect and a gambling effect. The time preference effect in health 'represents a preference structure in which the individual prefers his health gains earlier as opposed to later', a concept which is synonymous with what O&B called *pure* time preference. The gambling effect depends on one's attitude to risk, but 'risk aversion with respect to additional nonhealthy time' is assumed. The gambling effect seems to have its parallel in what O&B called uncertainty, the quantity effect in diminishing marginal utility of consumption. Hence, there appears to be a striking parallel between the concepts used by G&T and those used by O&B.

The pure time preference rate for health, η_H

Temporary health changes can be framed in the form of an intertemporal *choice* between alternative health streams (see Lipscomb, 1989). In those contexts it is therefore meaningful to apply the concept of pure time preference. However, there are not good grounds for hypothesizing about the ordinal relationship between η and η_H.

In general, though, there appear to be some qualitative differences between choices for alternative streams of health as opposed to choices for wealth. First, while lower consumption now may enhance future consumption, it is rather the case that *greater* current health enhances future health. As Cairns (1992) put it: 'It seems more natural to view individuals as selecting the rate of depreciation of their stock of health and augmenting their stock but not as trading it over time'. Second, when medical interventions provide permanent improvements, this health stream (H^2) would dominate that with no intervention (H^1):

$$H^1 = (H_0, H_1, ..., H_i, ..., H_m)$$

$$H^2 = (H_0, H_1 + \Delta H_1, ..., H_i + \Delta H_i, ..., H_m + \Delta H_m)$$

Compared with the choice between X^1 and X^2 in Section 3.1 there is no intertemporal trade-off between H^1 and H^2. In this context, then, the idea of a pure time preference for own health is at best strange.

The rate at which the marginal utility of increased future health diminishes, β_H

G&T suggest that a quantity effect in health is revealed when an individual who is offered various durations of life extension in a chronic dysfunctional health state 'values less each succeeding year in the chronic dysfunctional state simply because he is growing weary or satiated with the situation'. There are empirical studies in support (see e.g McNeil et al., 1978,

1981; Sutherland et al., 1982). However, the quantity effect refers to a diminishing marginal utility of *additional life years* – not to a diminishing marginal utility of an increased health gain in each succeeding year. In Pigovian terms: 'the present pleasure is greater than the future pleasure' in health streams of this type. Hence, this 'rate of utility depreciation' due to satiation does not bear theoretical links to the β variable, simply because the underlying intertemporal preferences are of a different nature. Furthermore, in certain chronic states individuals learn to *cope* with their new life situation, which will reverse the sign for this satiation effect, i.e. a utility-*increase* over time.

The uncertainty factor for health, u_H

A distinction should be made between uncertainty per se and the gambling effect. The former relates to the possibility that one will not be able to benefit from the promised future pleasure. While accepting O&B's contention that the influence of uncertainty may create an incentive to defer consumption in the event of a long life, individuals would not, for reasons of uncertainty, have incentive to defer a temporary health improvement. First, the marginal utility of a deferred health gain would not increase in a later period compared with the same gain now, provided that one's health state is not changing. Second, for consumption there is some utility from bequest in case of an early death, which of course has no meaning in the health context. A third point is that one cannot insure against the loss of one's own life in the same way as one can insure against the loss of one's own future consumption. Thus, the impact on the revealed time preference rate of the same 'fear of death' is different between health and wealth.

The gambling effect is revealed from individuals' responses to lottery questions concerning risky operations. By comparing gambling behaviour, Hellinger (1989) found that people were more risk averse with respect to choices involving years of life than they were to those involving monetary outcomes. However, given that an operation is equally as risky in the future as it is at present, then such risk-preferences are different from the 'future is uncertain'-concern which is included in the revealed time preference rate.

To sum up: the meaningfulness of the notion of a pure time preference rate, η, for one's own health has been questioned. Although there may well be a quantity effect due to satiation, the rate at which utility decreases over time has no theoretical links to the diminishing marginal utility variable, β. For reasons of uncertainty, u, individuals appear to have a stronger preference for expediting a given health gain than a consumption gain. Hence, it can be concluded that the underlying intertemporal preferences for health are of a different nature from those for wealth, and one would therefore expect the TPR_H to differ from the TPR_C.

4. A research agenda on discounting health

From the above discussion we can conclude, first, that the same discount rate should be applied to costs as to health, and second, that there are theoretical grounds for believing that this rate would be different in health care programmes than that used elsewhere. Beyond that, how do we proceed with empirical research? What messages are there to take into account in such research? In this final section a research agenda is set out for examining certain issues on eliciting the discount rate for health.

The key issues are that (1) it is hard to generalize from a TPR_H revealed, or expressed, in one specific context, (2) one should operate with social intertemporal choices, (3) the trade-offs should be between numbers of identical units of health gains at different points in time, and (4) the time interval should be chosen in such a way as to avoid concern for known lives.

4.1. Social intertemporal choices

In theory, a rational consumer would, because of the substitution possibilities within his/her available budget, reveal the same TPR_C for all goods in all situations. However, recent studies · have shown huge variations in the implied time preference rates (see e.g. Loewenstein and Thaler, 1989). A behavioural explanation is given for this seemingly irrational behaviour. Due to lack of self-control, consumers divide their assets into separate budgets. The danger of 'overspending' on one type of consumption is reduced by establishing self-imposed budget barriers. Thus, different TPR_Cs may be observed for each budget.

For health, the 'budget barriers' appear to be even higher. The actual utilization of health care is to a large extent decided by doctors, who then impose *external* 'budget barriers' between health and other goods. Secondly, consumers will generally have more problems in making smooth substitutions between health and other goods, than they have between two ordinary consumption goods in the utility function. Furthermore, different aspects of health behaviour may signal different TPR_Hs. An individual's short term preference for some element of immediate pain-relief might indicate a high TPR_H, while her healthy life-style might indicate a low TPR_H. To the consumer, these are such diverse contexts that the idea of equalizing her TPR_Hs is irrelevant. And the number of relevant substitution opportunities between various health goods/-bads are very small compared with the large number of substitution opportunities between consumption goods within a current budget period. Thus, the restricted opportunities to trade health suggest that the revealed TPR_H is even more situation specific, and thereby harder from which to generalize, than is the revealed TPR_C.

This conclusion strengthens the argument for operating with intertemporal

choice contexts which are close to those which health planners and decision makers face. Such choices deal with the allocation of health care resources to competing programmes, where benefits accrue at different points in time. These are social – not private choices. The individual – as citizen – may well have different time preferences for goods in a social context from that revealed in the context of private consumption (see e.g. Sen, 1982). Thus, a Social TPR would be best elicited by making individuals respond to explicitly social intertemporal choices.

Research Statement 1: Choices should be framed explicitly as social choices and should be between competing programmes which yield health benefits at different points in time. Each programme should cost the same, with all costs occurring in the current period.

4.2. Identical units at different points in time

When decision makers are faced with intertemporal trade-offs in practice, they are essentially comparing 'apples at time 1 with oranges at time 2', i.e. they compare one type of gain for one patient group in time 1 with another gain for another group in time 2. This involves a wide range of other concerns (e.g. different age groups and illnesses) and not just those related to the time aspect.

Problems arise in the context of prevention vs cure, if one describes a preventive programme and a curative programme and lets respondents express their trade-offs between number of future and present lives saved. First, preventive goods may yield utility in anticipation (Cohen and Mooney, 1984) beyond the utility bearing final outcome in terms of reduced mortality. Second, preventive programmes not only reduce mortality, but they often improve health as well (e.g. prevention of heart disease). Individuals will therefore tend to value a prevention programme for more than its final outcome. Simultaneously with their time preferences, the answers will be 'contaminated' by their subjective weighting of the differences in health outcomes. Thus, when framing intertemporal health choices, it is crucial not to mix up utility from other things with time preference.

If we aim at describing 'realistic' choices we are trapped by the problem of generalisation. Loewenstein (1987) concluded that 'discount rates estimated in specific contexts ... cannot be generalised beyond the domain of behaviour in which they were derived'. Hence, it is tempting to operate with gains that are in CEA/CUA units. Although respondents may perceive such choices as being very hypothetical, this problem has to be traded off with that of limited generalisability.

Research Statement 2: The benefits at different points in time should be

expressed in the same units, e.g. lives saved or number of individuals benefiting from a given health improvement.

By sticking to established CEA/CUA units of outcome, one avoids issues of life stage considerations (see Williams 1988, Charny et al., 1989) and preferences for particular health outcome streams. Such weighting of QALYs depending on which age group that benefits is not a discounting issue per se.

4.3. Limiting the time length

While environmental programmes yield benefits in the very long run, health care programmes would normally be termed long-run if benefits accrue after 20 years. In the short run, one must avoid the possible influence of the fact that individuals value known lives higher than statistical lives (see e.g. Mooney, 1977). Also medical ethics may favour the saving of individuals in immediate need. However, in our context of CEA/CUA, such analysis deals with assessing whether to implement new programmes or expand existing ones – impacts are rarely seen until after one year.

Research Statement 3: The interval within which individuals are asked to make their intertemporal trade-offs should be 1–20 (or 30) years.

Since it is consumers' preferences that are our concern, a random sample of the population must be selected. They should be asked to put themselves in the position of a health policy maker and express the number of units of future health gain which would make them indifferent to a given number of units in the coming year. With respect to framing, questions could be introduced as follows: 'Imagine that you are in a health planner's shoes ...'. The questions themselves could be, for example, 'how many lives saved after 20 years would be as good as saving 1000 lives next year?'

4.4. Hypothesis

From the discussion in Section 3, we have no hypothesis as to the ordinal relationship between individuals' time preferences for health as compared with those for wealth. However, there are at least two pointers which suggest that there is no such thing as one single Social TPR_H. First, the work of e.g. Strotz (1956), Ainslile (1975, 1991), Christensen-Szalanski (1984), Loewenstein (1987), and, most recently Cropper et al. (1991) suggest that constant rate discounting may not be the appropriate functional form for modeling individuals' time preferences for health. The TPR tends to decline with length of time, so that the correct discount rate for a given health care programme would depend on the duration of the programme.

In order to estimate such non-monotonic time preference functions,

respondents will, for each choice context, have to make at least two intertemporal trade-offs. The hypothesis that we do not discount health at a constant rate could be tested by operating with time periods of two different lengths, e.g. 1 vs 5 years and 1 vs 20 years.

Second, the problem-specific nature of our time preferences suggests that our Social TPR_H for life saving may differ from that for improvement of health states. So far empirical studies on time preference for health have focused either on life saving (Moore and Viscusi, 1990; Horowitz and Carson, 1990; Cropper et al., 1991) or on health changes (Lipscomb, 1989; Cairns, 1992). Because no study has applied the same approach to elicit individuals' TPR_H for life saving as for improving health, we do not know whether there are systematic discrepancies between the two. I would hypothesize that there are some discrepancies between the two, but not which is the highest.

4.5. Concluding remarks

The basis for this research agenda is the claim made by Lind (1990) that the only reasonable way to determine the social rate of time preferences is to elicit time preferences *at the individual level.* A crucial question arises: for which type of intertemporal health choices do we wish to elicit individuals' preferences? Our concern is with the 'weight(s)' that should be applied in CEA/CUA when comparisons are to be made between health benefits accruing at different points in time.

The empirical studies in this field have, except for Cropper et al. (1991), either used hypothetical choices about own health or described concrete policy scenarios. However, the lessons from behavioural economics are rather frustrating as to the generalisability of the results from these studies. It was the acknowledgement that individuals have diverse intertemporal preferences which led Lind (1990) to conclude that no single discount rate will be appropriate for all applications, and that the rate one should use therefore depends on the problem that is being analyzed.

The problem here concerns priority setting between competing health care programmes with different time profiles of the expected health gains. Hence, the heart of the issue which should be the baseline for empirical research in the field is: How would the present generation of citizens, with their current health care resources, prioritize between the saving of lives now vs in the future, and between improving the quality of lives now vs in the future? It is the Social TPR_Hs implied from such intertemporal choices which should form the basis for discounting in economic evaluations of health care programmes.

·Of course we are faced with the now classical problem of aggregating individual preferences. However, this is not a problem unique to that of

deciding a social discount rate. The simple solution in previous studies and in the CEA/CUA literature is to take the median or the mean of individuals' answers. Although this type of aggregation may be theoretically questionable, the procedure to aggregate individuals' preferences for the *timing* of outcomes should be similar to that of aggregating their *values* of these outcomes.

There is no a priori reason for expecting the implied Social TPR_H to be lower than the 'current practice' rate, which may be frustrating for those who favour a lower discount rate to be applied in health care programmes. A recent study in the US suggested that, on average, people discount future lives saved at 8.6% if the time horizon is 25 years, and at 3.4% if it is 100 years (Cropper et al., 1991). A preliminary study, along the lines suggested in this section, among 54 undergradute economics students gave an implied median Social TPR_H of 13.9% when the time horizon was 5 years and 8% when it was 20 years. More empirical research is needed to arrive at discount rates for health care programmes that are 'more moot' than what appears to be the current wisdom of implicitly assuming that the time preference for health is identical with that for wealth.

References

Ainslile, G., 1975, Specious reward: a behavioural theory of impulsiveness and impulse control, Psychological Bulletin 82, 463–509.

Ainslile, G., 1991, Derivation of 'rational' economic behaviour from hyperbolic discount curves, American Economic Review 81(2), 334–340.

Arrow, K.J. and R.C. Lind, 1970, Uncertainty and the evaluation of public investment decisions, American Economic Review 60 (3).

Baumol, W.J., 1968, On the social rate of discount, American Economic Review 58 (4), 364–378.

Böhm-Bawerk, E. von, 1888, The positive theory of capital, trans. W.Smart, G.E. Stechtert and Co., 1891..

Cairns, J., 1992, Health, wealth and time preference, Project Appraisal 7(1), 31–40.

Charny, M.C., Lewis, P.A. and S.C. Farrow, 1989, Choosing who shall not be treated in the NHS, Social Science and Medicine 28(12), 1331–1338.

Christensen-Szalanski, J.J.J., 1984, Discount functions and the measurement of patients' values, Medical Decision Making 4(1), 47–58.

Cohen, D.R. and G.H. Mooney, 1984, Prevention goods and hazard goods: A taxonomy, Scottish Journal of Political Economy 31(1), 92–99.

Cretin, S., 1977, Cost-benefit analysis of treatment and prevention of myocardial infarction, Health Services Research 12.

Cropper, M.L., Aydede, S.K. and Portney, P.R., 1991, Discounting human lives, Discussion Paper CRM 91-05 (Resources for the Future, Washington, DC).

Drummond, M.F., G.L. Stoddart and G.W. Torrance, 1987, Methods for the economic evaluation of health care programmes (Oxford University Press, Oxford).

Drummond, M.F., Brandt, A, Luce, B. and Rovira, J., 1993, Standardizing methodologies for economic valuation in health care: Practice, problems and potential, International Journal of Technology Assessment in Health Care 9(1), 26–36.

Feldstein, M.S., 1964, The social time preference discount rate in cost-benefit analysis, Economic Journal 74, 360–379.

Fuchs, V. and R. Zeckhauser, 1987, Valuing health – a 'priceless' commodity, American Economic Review 77(2), 263–268.

Gafni, A. and G.W. Torrance, 1984, Risk attitude and time preference in health, Management Science 30(4), 440–451.

Hellinger, F.J., 1989, Expected utility theory and risky choices with health outcomes, Medical Care 27, 273–279.

Horowitz, J.K. and Carson, R.T., 1990, Discounting statistical lives, Journal of Risk and Uncertainty 3, 403–413.

Johansen, L., 1967, Investeringskriterier fra samfunnsøkonomisk synspunkt (Finansdepartementet, Oslo).

Keeler, E.B. and S. Cretin, 1983, Discounting of life-saving and other nonmonetary effects, Management Science 29(3), 300–306.

Lind, R.C., 1982, A primer on the major issues relating to the discount rate for evaluating national energy options, in: Lind, R. ed., Discounting for time and risk in energy policy (Johns Hopkins University Press, Baltimore).

Lind, R.C., 1990, Reassessing the government's discount rate policy in light of new theory and data in a world economy with a high degree of capital mobility, Journal of Environmental Economics and Management 18, S8–28.

Lipscomb, J., 1989, Time preference for health in cost-effectiveness analysis, Medical Care 27, S233–S253.

Loewenstein, G., 1987, Anticipation and the valuation of delayed consumption, Economic Journal 97, 668–684.

Loewenstein, G. and R.H. Thaler, 1989, Intertemporal choice, Journal of Economic Perspectives 3(4), 181–193.

McNeil, B.J., R. Weichselbaum and S.G. Pauker, 1978, Fallacy of the five-year survival in lung cancer, New England Journal of Medicine 299, 1397–1401.

McNeil, B.J., R. Weichselbaum and S.G. Pauker, 1981, Tradeoffs between quality and quantity of life in laryngeal cancer, New England Journal of Medicine 305, 982–987.

Mooney, G.H., 1977, The valuation of human life (Macmillan, London).

Moore, M.J. and W.K. Viscusi, 1990, Discounting environmental health risks: New evidence and policy implications, Journal of Environmental Economics and Management 18, S51–62.

The New Palgrave, 1987, A dictionary of economics (Macmillan Press Ltd, London).

Olson, M. and M.J. Bailey, 1981, Positive time preference, Journal of Political Economy 89(1), 1–25.

Parsonage, M. and H. Neuburger, 1992, Discounting and health benefits, Health Economics 1, 71–76.

Pearce, D.W and R.K. Turner, 1990, Economics of natural resources and the environment (Harvester Wheatsheaf, London).

Pigou, A.C., 1932, The economics of welfare, 4th ed., (Macmillan, London).

Portney, P.R., 1990, Comments on 'discounting' session, Journal of Environmental Economics and Management 18, S63–64.

Robinson, J.C., 1990, Philosophical origins of the social rate of discount in cost-benefit analysis, The Milbank Quarterly, 68(2), 245–265.

Sandmo, A. and J.H. Dreze, 1971, Discount rates for public investment in closed and open economies, Economica 38(152), 395–412.

Sen, A., 1982, Approaches to the choice of discount rates for social benefit-cost analysis, in: Lind, R., ed., Discounting for time and risk in energy policy (Johns Hopkins University Press, Baltimore).

Strotz, R.H., 1956, Myopia and inconsistency in dynamic utility maximization, Review of Economic Studies 23, 165–180.

Sutherland, H.J., H. Llewellyn-Thomas, N.F. Boyd and J.E. Till, 1982, Attitudes toward quality of survival – the concept of 'maximal endurable time', Medical Decision Making 3, 299–399.

Viscusi, W.K. and M.J. Moore, 1989, Rates of time preference and valuations of the duration of life, Journal of Public Economics 38, 297–317.

Weinstein, M.C. and W.B. Stason, 1977, Foundations of cost-effectiveness analysis for health and medical practices, New England Journal of Medicine 296(31), 716–721.

Williams, A., 1988, Ethics and efficiency in the provision of health care, in: Bell, J.M. and Mendus, S., Philosophy and medical welfare (Cambridge University Press, Cambridge).

Paper 5

ECONOMIC EVALUATION

TIME PREFERENCES FOR HEALTH GAINS: AN EMPIRICAL INVESTIGATION

JAN ABEL OLSEN
Department of Economics, University of Tromsø, Norway

SUMMARY

The purpose of this study was to elicit the implied discount rates to be used in economic evaluations of health care programmes. The paper presents results from two Norwegian surveys in which a random sample of the population and a sample of health planners were asked to prioritise between alternative health care programmes, and make trade-offs between future health gains and more immediate gains. The questionnaire had four hypothetical choice situations; two for life saving and two for health improvement.

KEY WORDS—Discount rates, empirical evidence.

In the cost-effectiveness and cost-utility methodology for evaluating health care programmes, there is wide agreement that 'a life is a life is a life' and 'a QALY is a QALY is a QALY'. But this holds only as long as the benefits occur at the same point in time. When comparing future health gains with more immediate ones, a discount rate is used, which essentially operates as a 'time weight'. However, it is argued by some that health gains should be left undiscounted or be discounted at a lower rate than the one used elsewhere in the economy. There is the argument that all generations should be given equal weights.[1] Again there is the argument for discounting non-tradables (e.g. human lives) in their own terms,[2] which is assumed to be at a lower rate than tradables. Most recently, Parsonage and Neuburger[3] suggested in this journal that QALYs should be discounted at a rate close to zero.

Although, among health economists, there is a high level of agreement on discounting in principle,[4] there is, however, less agreement as to the size of the discount rate. In her review of CUA's, Gerard[5] found that the rate used was in the range 3%–10%. Then, with such different views and practices, economists with an affinity to the consumer sovereignty principle would ask which discount rate best corresponds to individuals' intertemporal preferences for health gains. Do people believe that health gains should count equally no matter when they arise? Do they have a relatively low time preference for health, or is it the case that they more strongly devalue future health gains relative to present gains?

The primary aim of the study reported here was to explore which implicit discount rates people have for social choices with respect to health care resource allocation. More specifically this paper seeks to answer four questions: 1) What are the implied time preference rates for health gains?, 2) Do the rates differ depending on the length of the time horizon over which they have been elicited? While these two questions add to previous empirical research, the following have not been investigated before: 3) Are time preferences different for life saving than for health improvements?, 4) Are time preferences among informed health planners different from those among 'less informed' opinion?

Address for correspondence: Jan Abel Olsen, Department of Economics, University of Tromsø, S. Tollbugt. 7, 9037 Tromsø, Norway.

PREVIOUS STUDIES AND RECEIVED WISDOM

Since the exploratory study by Fuchs,[6] the issue of the time preference for health has received considerable attention in the health economics literature. However, it is only fairly recently that empirical studies have been carried out. Of the five published studies in the area, three deal with people's implied discount rates for life saving, and two with that for health changes.

Moore and Viscusi[7,8] estimated individuals' time preference rates for future years of life implicit in observed trade-offs between wages and occupational risk. Two different models, using the same data set, gave quite different rates; 2% and 11%. Lipscomb[9] tried to identify a time preference for health on a group of students by varying the delay of onset of a temporary inferior health state and its duration. Although implicit rates were not estimated, the results indicated a strong time preference for health; e.g. a delay of onset of three years and a duration of three years in the inferior state gave a higher preference score than no delay but with a duration of one year.

Cairns[10] adopted a similar approach, also on a group of students, but using a simpler design than Lipscomb. The implicit time preference rates for health were zero or small but positive, in the range 0.4%–3%. Cairns notes that, since individuals value good health differently depending on age, what appears as time preference may be the result of different life stage considerations rather than futurity as such. According to Williams,[11] people tend to value good health considerably higher at the life stage 'when bringing up children' than the life stage 'when starting work'. It is likely therefore that the implicit rates expressed by Cairns' respondents have been reduced, reflecting the view that a temporary inferior health state in 20 years is worse than the same state in 2 years. Thus, when framing intertemporal choices related to the respondents' *own* health, life stage considerations may 'contaminate' the expressed time preferences.

While the above studies framed the questions in terms of *private* choices, in both Horowitz and Carson[12] and Cropper *et al.*[13] *social* choices between alternative life saving programmes were considered. Using again a group of students (evidently the cheapest and most obedient experimental animals there are) Horowitz and Carson uncovered median discount rates of 4.5% in a

flight safety experiment, 4.7% for occupational safety and 12.8% for traffic safety improvements.

The study by Cropper *et al.* differs from the previous ones in that it was the most extensive; three telephone surveys of a total of 3200 households. Their intertemporal choices concerned environmental programmes which would save lives. From earlier findings that people do not discount at a constant rate, different time horizons were used to test for this. Their findings suggested that the discount rate follows a negatively sloped, convex pattern. With a time horizon of 5 years, the estimated median discount rate was 16.8%. On a 25 years basis, it was 7.4%.

The empirical studies mentioned above and other studies of intertemporal choices suggest that the implied discount rate is not only time length dependent but situation specific as well (see e.g. Loewenstein[14]). In the context of priority setting in health care, it becomes crucial, then, to frame the choices as closely as possible to the issue at hand, i.e. how a citizen would wish to see society trade off—at the margin—future health gains relative to present gains. The rate to elicit is the marginal social time preference for health gains. (See Olsen[15] for a more detailed discussion of the research agenda chosen here.)

The situation specific nature of previously observed time preferences implies that differences might also exist between time preferences for health improvements as compared with life savings. In that case, rates elicited from intertemporal trade-offs between lives saved might not be the appropriate ones to use in programmes dealing with morbidity. Different rates might also exist for losses as opposed to gains. However, this study focuses on gains only.

The methodological objectives with regard to validity and generalizability require that the choices will have to be framed in a constructed—and hypothetical—format, which may be problematical for respondents to perceive. Assuming that health planners are more familiar with the issues, a parallel sample of health planners and decision makers was chosen to compare a better informed sample with a random sample of the Norwegian population.

METHOD

The questionnaire had two parts, one dealing with life saving, the other with health improvement.

The two parts use the same approach and have an identical format. In each part, there are two time horizons, five years and twenty years. The reference time is set at one year from now, in order to avoid the possible influence of individuals' higher valuations of known lives than statistical lives. Each of these four choice contexts opens with an ordinal choice between 1000 benefit units (lives saved or number of people's health improved) in one year or in the future. If the respondent prefers the closer of the two options, they are asked to indicate the number of benefit units that the more distant programme would need to produce to make it equally good as 1000 in one year. Those who found it difficult to give a finite number were allowed to indicate a range within which the number indicating indifference would lie. (See the appendix for the context for the first choice.) Furthermore, the questionnaire had a first full page of text explaining the purpose of the study and focusing on the policy choice of prioritising between competing health care programmes when their benefits occur at different points in time. A final page asked for the respondent's age, sex and education. The respondents' implied time preference rates were estimated as follows:

$$r = \frac{\ln(X/1000)}{t}$$

where X is the future number of benefit units, and t (4 or 19) is the *time interval* between the future and closer options. When a range was indicated instead of a finite indifference value, X, the midpoint of the range was taken as their indifference value.

This study is based on two surveys which took place in March 1992. In the first survey, the sampling and handing out of questionnaires were administered by a private opinion poll organization ('ScanFact'). An envelope containing the questionnaire, an accompanying letter and a stamped addressed envelope was handed out personally to half of their 'Mini-Norway' sample (i.e. their random sample of the Norwegian population aged 15 and over). It was believed that this personal delivery to 550 individuals would increase the response rate. As a further incentive to respond, a lottery ticket worth £2 was promised to everyone who returned the questionnaire. The replies were anonymous and no reminder was sent.

In the second survey the questionnaire was sent to every chief health manager of Norway's nineteen counties, and to all employees above a certain position level ('konsulent') in the Department of Health and the Directorate of Health. (Furthermore the questionnaire was sent to members of the health and social affairs committee in the Parliament. With only 3 out of 16 responding, this group was omitted.) We believed that an anonymous approach would increase the response rate. Again no reminder was sent.

RESULTS

In the opinion survey, 250 answered, which is a response rate of 45%. Comparing the three background variables, our sample appears to be fairly representative of the Norwegian population. When compressed into three age groups, the sample has the same proportion as nationally (when excluding people under 15) of the group aged less than 30 years (29%), a higher proportion of those between 30 and 60 (54% compared with 45%) and, hence, a lower proportion of those 60 and older. In the sample 51.6% were women. With respect to educational level, the sample is upward biased; 27% have a college or university degree, compared with 15% in the Norwegian population.

In the survey of the health planners, 77 out of 209 responded (i.e. 37%). Due to lack of information as to how this total population is distributed by age, sex and education, we cannot tell how representative the respondents are. Compared with the random sample, one third were women, 80% were between 30 and 60 years old and all had a college or university degree.

Positive time preferences

Each of the four choice contexts started with an ordinal question; which of the two programmes would you choose? The frequencies in the random sample are given in Table 1.

With a time horizon of five years, 84% prefer the closer, whereas 90% prefer the closer when the time horizon increases to twenty years. In other words, only 16% and 10% respectively have zero or negative time preference rates. An intuitive interpretation would be that when comparing benefits after one year vs after five years, both alternatives look like in the near future, whereas twenty years ahead represents 'a jump' to

Table 1. Distribution in the random sample preferring the different options

Time horizon	Life saving T = 5	Life saving T = 20	Health improvement T = 5	Health improvement T = 20
Prefer closer	207	228	212	223
Prefer distant	6	5	13	9
Indifferent	35	17	25	18
Missing	2	0	0	0
Total	250	250	250	250

the distant future. The differences between life saving and health improvement are small, except that more people prefer the distant options for health improvements. Among the few who prefer the distant options, there is a majority of young people. One explanation for their choice could be that, while their current health is good, their own need for health improving resources is greater later in life.

Among the health planners, more than 90% preferred the closer, while none preferred the distant.

Strong time preferences

Those who preferred the future programme were not asked how many fewer than 1000 benefit units they considered equally as good as 1000 in one year. The reason for omitting this, however logical, option was that it would have complicated the questionnaire.

Instead we used 999 as their indifference values. Of course the true figures may well have been lower, but we would not speculate on this. However, lower alternative numbers would not change the median rate which we are mainly concerned with.

In the random sample, there are 41 missing cases in the first choice context and 44 in the remaining three. Of those missing, there were 6 blank in the first and 9 in the other three. The majority of missing cases are a group of 35 respondents who were omitted due to a strong suspicion that they had misinterpreted the trade-offs. They had consistently answered 5000 to the two options with a T = 5 and 20,000 when T = 20, which suggests that they may have perceived the programmes as lasting over the whole period rather than as providing a 'one-shot' gain in one

future period. But still, the figures 5000 and 20,000 were relatively attractive choices. For the choice of life saving in 5 years (L5), 32 respondents chose the value 5,000, and for health improvement in 5 years (H5), 31 chose it. 14 chose 20,000 in L20 and 19 in H20.

In the sample of health planners, 4 respondents were omitted for the same reason as above. The proportion is 5% as compared with 14% in the random sample. Given the assumption that health planners are better informed to interpret the choice contexts correctly, this gives support to our classification of these answers as misinterpretations, and hence to exclude them from the analysis. The number of blank varied between 4 and 8.

In order to play down the influence of extreme values answers *over* 10,000 for T = 5 and 30,000 for T = 20 were replaced by 10,000 and 30,000 respectively. (In the random sample, there were 7 extremes in L5, 11 in L20, 5 in H5 and 9 in L20. In the planners' sample, there were no extremes in L5, but 4 in L20, 2 in H5 and 5 in H20.)

Testing the *mean* values of the implicit rates in the random sample showed that the short term rates are significantly greater than 15%, while the long term rates are significantly greater than 7% ($p < 0.0001$). Using a binomial test,[16] the observed significance level that the short term *median* rates are greater than 15% were $p = 0.0043$ for L5 and $p = 0.0009$ for H5. The long term median rates are greater than 7% ($p < 0.0001$). In the planners sample the long term median rate for H20 are greater than 5% ($p = 0.0128$).

Some indications of the sensitivity of the results following from adjusting the extreme values and the assumption of misinterpretations can be seen in Table A1 of the appendix.

Non-constant time preference rates

Comparing the implicit rates for the two life saving choices, 84% of our respondents revealed a decreasing rate; on the health improvement choices, 87% revealed a decreasing rate. The Wilcoxon signed rank test of the short term compared with the long term rates showed that the short term rates were significantly different ($p < 0.0001$) from their respective long term rates for all combinations (i.e. L5 with L20, and H5 with H20 within both samples). This Wilcoxon test is the nonparametric analogue of the t-test for

Table 2. Implied time preference rates (%) in the two samples

Choice	L5		L20		H5		H20	
Sample	rand.	plan.	rand.	plan.	rand.	plan.	rand.	plan.
N	209	69	206	66	206	68	206	65
Mean	22.1	13.6	8.9	6.6	23.3	17.2	9.2	7.6
Std dev	18.5	16.0	5.9	5.6	18.4	17.3	6.0	5.8
Std err	1.3	1.9	0.4	0.7	1.2	2.1	0.4	0.7
Median	**17.3**	**6.6**	**9.4**	**5.8**	**22.9**	**11.8**	**10.2**	**7.3**
25 pctile	2.4	0.1	3.6	1.1	4.6	1.5	3.6	2.5
75 pctile	40.2	17.3	14.2	10.6	40.2	27.5	14.3	12.1

matched samples, but it does not require the differences to be normally distributed.[17]

Highest rates for health improvements

In Table 2 we note that for all comparisons between life saving and health improvements (L5 with H5 and L20 with H20) in the two samples, both the mean and the median rates are lower for life saving. A non-parametric test indicated that most differences were significant, Table 3.

The observed significance levels of Table 3 support the conclusion that time preferences are stronger with regard to health improvements than for life savings, which is a novel finding in this literature.

Health planners have lower rates

Furthermore, Table 2 reveals that for all comparisons between the random sample and the planner sample, the latter gave the lower rates. Using a two-sample median test gave the following observed significance levels for the differences between the medians; $p < 0.0001$ for L5, $p = 0.0017$ for L20, $p = 0.0026$ for H5 and $p = 0.1650$ for H20.

Table 3. Two tailed significance probabilities that the implied median time preference rates are the same for life saving as for health improvements (Wilcoxon signed rank test)

Samples	rand.	plan.	Merged samples
L5 with H5	0.1351	0.0055	0.0095
L20 with H20	0.0462	0.0158	0.0041

Explanatory power of age, sex and education

In the random sample, a multiple regression analysis was used for each of the four dependent variables; L5, L20, H5 and H20, to test the explanatory power of the three independent background variables; sex, age and education. There were hardly any differences in time preferences between men and women (the two-tailed significance levels that they differed (sig T) were in the range 0.6 to 0.7!). The linear relationship were weakly positive for age, but not much significant (sig T < 0.10 for L5 and H20). The education variable had been recorded on five levels (1 = o-level, 2 = vocational training, 3 = A-level, 4 = 1–2 years professional education, and 5 = college/university degree). However, rather than entering it as a continuous variable in the regression analysis, it was introduced as a dummy comparing the highest educated with the rest. Those with college or university degrees had significantly lower rates than the rest (sig T < 0.01 for L5, <0.05 for L20 and <0.10 for H5 and H20).

A closer examination of the background variables was made using the two-sample *median test*. There was a striking similarity between the sexes, except for H5 for which women had higher median rates than men ($p = 0.0963$). Comparing the median rates in the age group under 30 ($n = 61$) with the rest gave lower values, significantly at the 10% level except for L20. Although the age group 60 and over ($n = 36$) had higher time preference rates than the rest of the sample, the differences were far from significant. Interestingly, the oldest age group 70 and over ($n = 14$) had the same median rates as the rest of the sample for L5 and H5, and; although not significant, a *lower* rate for H20.

The education group whose median rates came out most significantly different from the rest was the one with college/university degree. (L5 and L20 were lower at the 1% level, H5 at 5% and H20 at the 10% significant level.) Again using the two-sample median test on a comparison of the rates in this highest educated group with the planners sample indicated some striking similarities. Although the planners still had slightly lower median rates, they were far from significant.

DISCUSSION

The magnitude of the time preference rates elicited in this survey might appear higher than anticipated—they reflect a high degree of 'impatience' with respect to when individuals wish the benefits of health care programmes to occur. Individuals' impatience for own consumption has for long been something about which many economists have felt uneasy. Pigou[18] wrote that 'our telescopic faculty is defective'. Others have labelled it 'short-sighted' or 'myopic'. However, the heart of the matter is whether, and to what extent, intertemporal preferences revealed from *own* consumption behaviour should govern social intertemporal choices. One resolution of this dilemma, given that we wish to base social decision making on individuals' preferences, stems from the view that individuals play a 'dual role', as a consumer and as a citizen (see Sen[19]). The *social* time preference rate could therefore be elicited from the explicit framing of social intertemporal choices, which is the approach chosen in this study.

Given the somewhat abstract choices involved, one could not expect a very high response rate in the random sample; 45% is in fact higher than expected. More surprising was the lower response rate of 37% in the planners' sample. However, we have a clear impression from sources in the Department and the Directorate that this was due to an aversion towards the actual framing of the choices. (They were characterized as 'out of touch with real life', 'irrelevant', etc.) Hence, we have no reason to believe that the non-respondents had systematically different views, but rather that they simply did not wish to participate in such an 'academic exercise'.

In the random sample, the implied rates may have been exaggerated if, as seems possible, some respondents forgot to think of the reference time as being in one year but rather than at present. For the readers who believe that, the implied rates would be 4/5 of the values given under L5 and H5, and 19/20 of the L20 and H20 values. But there are other effects which work in the adverse direction: people with low rates (high education) are overrepresented in the sample and people with high rates (60 years and older) are underrepresented.

Of the five previous studies referred to earlier, that which adopted the most similar approach is the study by Cropper *et al.*[13] A comparison shows that our results come very close to theirs. With $T = 5$, the Cropper study showed a median of 16.8% for life saving while we obtained 17.3%. We cannot make the same comparison for $T = 20$, because the Cropper study did not use this time horizon. They used $T = 10$ and $T = 25$. However, if we assume simply that the discount rate declines linearly between $T = 10$ and $T = 25$, then the weighted average of the median discount rate which corresponds with $T = 20$ becomes 8.8%. (If the discount rate follows a *convex* pattern, the rate corresponding with $T = 20$ would be somewhat lower than this.) Still, this rate is only slightly lower than our 9.4%. These results from the two studies are strikingly similar.

With respect to the suggestion in Cropper *et al.* that 'the annual discount rate follows a negatively sloped, convex pattern', our data confirm the negative slope, but with only two time horizons, cannot provide evidence on convexity. The longer the time horizon, the lower are the implied rates. This finding has important implications if one wishes to base the social discount rate on individuals preferences, in that one cannot use a rate elicited from a short time horizon for projects with a long time horizon.

The same argument holds for health improving projects, if the suggested rate has been elicited from trade-offs between life saving and one believes that time preferences for health improvements (morbidity) are different from those for life saving (mortality). An important finding in this paper, which has not been the subject of previous empirical studies, is that the time preference rates for health improvement are higher than for life saving. In other words, people were willing to forgo more future health improvements in order to improve the health of those currently in need than they were to forgo future lives saved in order to save lives now. A possible explanation is that respondents are more concerned with waiting lists

and 'current unmet health needs' due to limited hospital resources, than they are for the unknown lives that can be saved in a year's time.

The high implied rates may reflect an obligation to do something for those who need health care now rather than spending resources on programmes which prevent currently healthy people from getting sick in the future. This can be compared with the view expressed by the late Norwegian Health Director, that reallocating health care resources from cure to prevention implies that 'one takes from the sick and gives to the healthy' (Dagbladet, 27.3.92).

There are at least two explanations for the lower time preference rates observed among health planners. In the role of 'guardians of society', they are likely to be more concerned with the health of future generations, and second, because the vast majority had college or university degrees (bearing in mind the negative relationship between time preferences and educational level).

Among health planners, the standard deviations are not much less than in the random sample (Table 2). Given the assumption that health planners should be more capable of understanding the questions and, not least, the policy implications of their answers, this may suggest that the variance reflects heterogeneous preferences rather than 'casual responding', and hence low reliability, in the random sample. (For L20 and H20 the ratio of the variance in the two samples were 1.10 (F-value). Using the Levene test,[20] the observed significance levels were too high to reject the hypothesis that the variances in the samples are equal.)

Although we would hold that the results of this study seem reliable, we do acknowledge the need for more empirical research in the area. In future research, different time horizons should be used to test the possible bias towards expressing indifference values equal to T*1,000 (where T is the time horizon). Further, specific 'numbered' choices could be used rather than asking the open-ended questions involved here. In a recent pilot study of a random sample of Aberdonians, Cairns[21] used time horizons of 7 and 25 years, and discrete choices. Following a fairly similar format to the one used in the study referred to here, Cairns obtained median implied time preference rates of 16.5% with T = 7 and 10.1% with T = 25. Together with the Cropper et al. study, this study by Cairns provides supportive evidence for suggesting that the median citizen's time pre-

ference rate for health is close to 10% when the time horizons are 20–25 years.

One explanation for the high implicit time preference rates for health could be that the 'caring externality' for other people's health declines rapidly over time, analogously to what it apparently does over space. Although most people are aware that a given health care budget could save more human lives in poor countries than it would domestically, rich countries still prioritise the health of their own citizens at the expense of that of people in poorer countries. Implicitly they thereby adopt a fairly high—what Parfit[1] termed —'Spatial Discount Rate'. Given these spatial limitations on our caring externalities, one should perhaps not be too surprised by the magnitude of the social time preferences for health expressed in this study.

Our results may be particularly thought provoking to those who hold that a rate close to zero should be applied in health care programme evaluations. The moral principle that all generations should be given equal weight (Parfit[1]) is, of course, an attractive principle, but no more so than that all human beings of the same generation around the world should have equal weights. If we disregard the high implicit 'spatial discount rate' for health, the implication would be that most of our health care resources would be transferred to poor countries. It would be rather easy to believe that it is morally worse to discount the health of people in poor countries than to discount the health of future generations in one's own country.

It is, of course, a normative issue whether to base the discount rate to be used in the public sector on the intertemporal preferences citizens have for the good this sector is to provide. However, it is also normative to base health care priority setting in general on citizens' preferences. Yet most economists still have a strong affection for this normative stance.

ACKNOWLEDGEMENTS

APPENDIX

The context for the first of the four choices in the questionnaire

Part I

The health service is being allocated additional . expenditure which is to be spent on a programme which *saves human lives*. Assume now that the only differences between the programmes are *when* human lives are saved. (The programmes save lives of the same age. Thus, the same people cannot be saved at different points in time. We do not know who will be saved.)

Choice situation 1
—life saving in 1 year or in 5 years:
Programme A will save 1000 lives *in 1 year*.
Programme B will save 1000 lives *in 5 years*.
Which of the two programmes would you choose?

(Please tick)

Programme A

Programme B

A and B are equally good

IF YOU CHOSE A:
Imagine that *more than 1000 lives* could be saved in 5 years. Roughly how many lives do you think Programme B would have to save in order for *A* and *B* to be considered *equally good*?

Answer (please complete): _____ LIVES

[If you find it difficult to give a finite number, would you please indicate the range within which the number of lives saved by Programme B *in 5 years* would have to lie for *A and B* to be considered *equally good*:

BETWEEN _____ AND _____ LIVES]

Table A1. Implied time preference rates (%) in the two samples when extreme values are *not* adjusted and those assumed to have misinterpreted are included

| Choice	L5		L20		H5		H20	
Sample	rand.	plan.	rand.	plan.	rand.	plan.	rand.	plan.
No adjustments of extreme values:								
N	209	69	206	66	206	68	206	65
Mean	23.1	13.6	9.2	6.7	24.1	17.8	9.5	8.0
No adjustments of extreme values and inclusion of misinterpretations:								
N	244	73	241	70	241	72	241	69
Mean	25.5	15.1	10.2	7.2	26.4	19.0	10.4	8.4
Median	27.5	10.1	12.1	6.2	27.5	15.7	12.1	8.5

REFERENCES

1. Parfit, D. *Reasons and persons*. Oxford: Clarendon Press, 1984.
2. Goodin, R. Discounting discounting. *Journal of Public Policy*. 1982; 2(1): 53–72.
3. Parsonage, M. and Neuburger, H. Discounting and health benefits. *Health Economics*. 1992; 1(1): 71–76.
4. Drummond, M., Brandt, A., Luce, B. and Rovira, J. Standardizing methodologies for economic evaluation in health care: Practice, problems, and potential. *International Journal of Technology Assessment in Health Care*. 1993; 9(1): 26–36.
5. Gerard, K. *A review of cost-utility studies: Assessing their policy-making relevance*. Discussion Paper 11/91, Health Economics Research Unit, Aberdeen, 1991.
6. Fuchs, V. Time preference and health: an exploratory study. In Fuchs, V. (ed.) *Economic aspects of health*, University of Chicago Press, 1982.
7. Moore, M. and Viscusi, W. K. Discounting environmental health risks: New evidence and policy implications. *Journal of Environmental Economics and Management*. 1990; 18S: 51–62.
8. Viscusi, W. K. and Moore, M. Rates of time preference and valuations of the duration of life. *Journal of Public Economics*. 1989; 38: 297–317.

9. Lipscomb, J. Time preference for health in cost effectiveness analysis. *Medical Care.* 1989; **27S**: 233–253.

10. Cairns, J. Health, wealth and time preference. *Project Appraisal.* 1991; **7**(1): 31–40.

11. Williams, A. Ethics and efficiency in the provision of health care. In Bell, J. and Mendus, S. *Philosophy and medical welfare.* Cambridge University Press, 1988.

12. Horowitz, J. and Carson, R. Discounting statistical lives. *Journal of Risk and Uncertainty.* 1990; **3**: 403–413.

13. Cropper, M., Aydede, S. and Portney, P. Rates of time preference for saving lives. *American Economic Review.* 1992; **82**(2): 469–472.

14. Loewenstein, G. Anticipation and the valuation of delayed consumption. *Economic Journal.* 1987; **97**(Sept.): 668–684.

15. Olsen, J. A. On what basis should health be dis-counted? *Journal of Health Economics.* 1993; **12**(1): 39–53.

16. Norušis, Marija J. *SPSS/PC+ Statistics™4.0.* Chicago: SPSS Inc, 1990.

17. Pfaffenberger, R. and Patterson, J. *Statistical methods for business and economics.* Illinois: Irwin, 1987.

18. Pigou, A. *The economics of welfare* (fourth edition). London: Macmillan, 1932.

19. Sen, A. Approaches to the choice of discount rates for social benefit-cost analysis. In Lind, R. (ed.), *Discounting for time and risk in energy policy.* Baltimore: John Hopkins University Press, 1982.

20. Norušis, Marija J. *The SPSS GUIDE to data analysis for release 4.* Chicago: SPSS Inc, 1990.

21. Cairns, J. *Discounting health benefits.* Paper for the 2nd European conference in Health Economics. Paris: December, 1992.

Paper 6

ECONOMIC EVALUATION
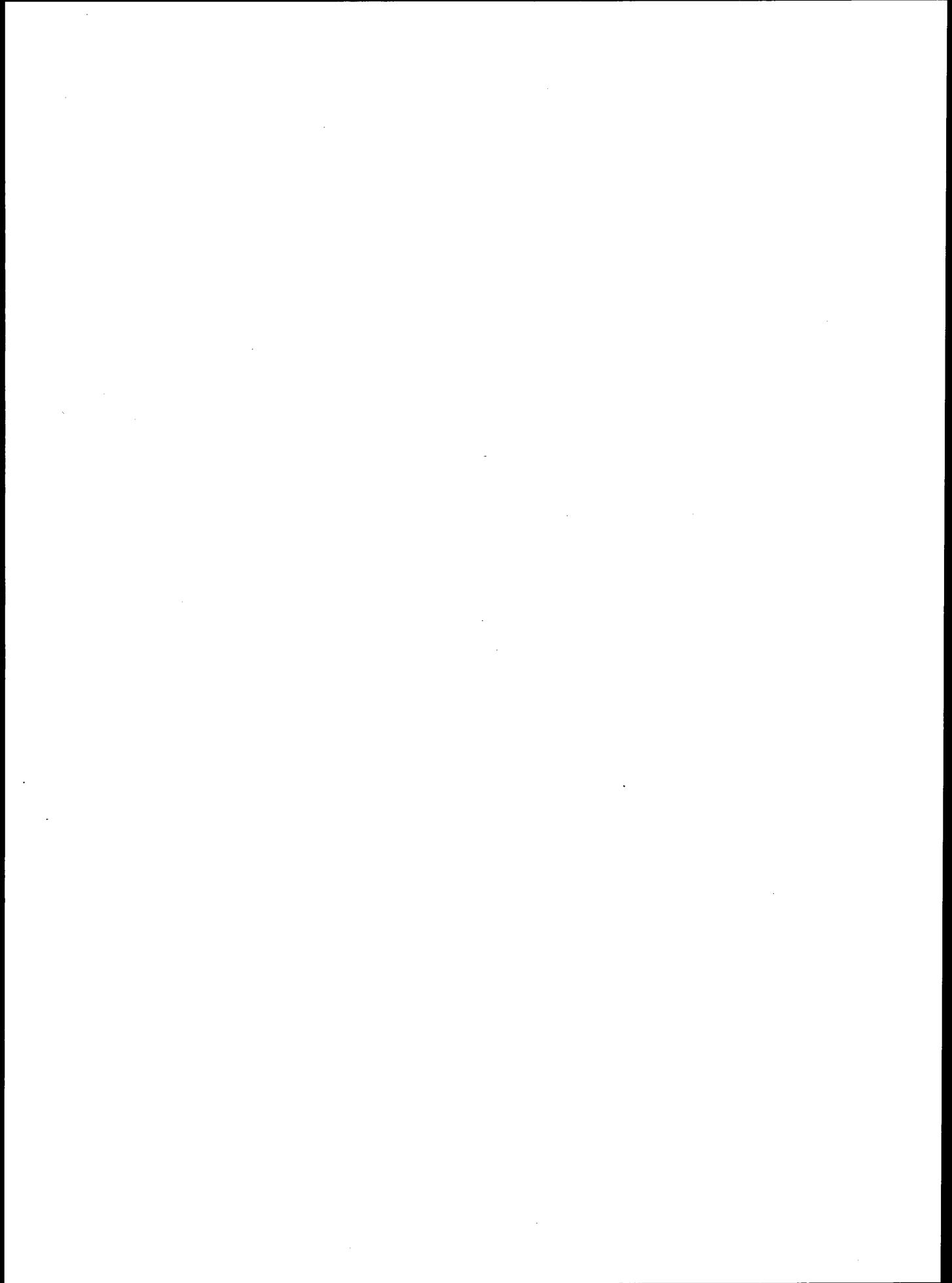
PERSONS VS YEARS: TWO WAYS OF ELICITING IMPLICIT WEIGHTS

JAN ABEL OLSEN

Department of Economics and Institute of Community Medicine, University of Tromsø, Norway

SUMMARY

The paper presents two methods for eliciting individuals' implicit weights to be attached to health benefits which extend over time. The first is a person-trade-off type which implies that social weights are attached to years with improved health when these are distributed between different groups of people. The second method is based on time-trade-off which yields private weights to succeeding years with improved health for oneself. Emphasis is placed on methodological issues, but some preliminary results obtained from small samples are also presented.

In the literature on quality-adjusted-life-years (QALYs) and in applied cost-utility analysis, it is normally assumed that 'a QALY is a QALY is a QALY' independent of age, sex, income, race, etc.. This facet of QALYs as being of equal social value is normally considered positively. However, there is more ambiguity regarding a facet of QALYs which is the subject of this paper: that a QALY has the same constant social value independent of *how many* QALYs a person will get. In the context of priority setting, this means that society would be indifferent as to how a given number of QALYs are distributed between persons: a programme which provides a large number of QALYs to each of a few would be considered equally good as one which provides a smaller number to each of many. Although there is a recognition in the literature of this ignorance of possible preferences for a more equitable distribution of QALYs,[1,2] to the knowledge of this author, no methodology has been presented for estimating the social weights to be applied on increasing QALYs.

It is also recognised that individuals do not ascribe a constant value to each succeeding year for oneself in an inferior health state.[3] Further, there are several studies which have found that the implicit health state value derived from trade-offs between duration and health quality varies with the duration of the described health state,[4-6] i.e. there is a non-constant proportional trade-off. When the proportion one is willing to give up is higher the longer the duration, this can be explained by a positive discount rate. However, again to the knowledge of this author, no methodology for eliciting the implied rate inherent in individuals' future streams of health has been presented in the literature.

The aim of this paper is to present two methods for eliciting individuals' implicit weights to be attached to health benefits which extend over time. The first method is a person-trade-off (PTO) type which implies that social weights are attached to years with improved health when these are distributed between different groups of people. The second method is based on time-trade-off (TTO) which yields private weights to succeeding years with improved health for oneself. In this paper, emphasis is placed on methodological issues. However, some preliminary results obtained from

Address for correspondence: Jan Abel Olsen, Department of Economics, University of Tromsø, S. Tollbugt 7, 9037 Tromsø, Norway.

small samples are also presented. Before going into these methods, the next section gives some theoretical background by starting with the QALY algorithm and introducing three weighting issues.

THE QALY-ALGORITHM WITH EXTENSIONS

Although health economists would be familiar with the calculation of QALYs, in this context, a distinction is made between what will be termed 'unweighted QALYs' as opposed to 'weighted QALYs'. The formula for calculation of 'unweighted QALYs' gained from a health quality improving intervention is simply the product of the health improvement and life expectancy:

$$QALYs = \Delta HS * T, \qquad (1)$$

where

$$\Delta HS = HS^1 - HS^0$$
$$0 < HS^1 \leqslant 1$$
$$0 < HS^0 < HS^1$$

HS^1 = the health state value *with* treatment,
HS^0 = the health state value *without* treatment,
T = years of life expectancy

Assuming that: 1) ΔHS correctly reflects individuals' valuation of the given health state increment, 2) this health increment remains constant, and 3) T is the expected duration with improved health as suggested by, for example, life expectancy tables, then the above QALY calculation represents the health effect, or outcome, of a given intervention. (Interventions often yield a combination of health improvement and life extension.[7]) Note that the only preferences inherent in the above formula are individuals' valuations of being in the two health states: without treatment and with treatment.

Certainly, people have preferences related to the duration of a health improvement, which suggest that succeeding years do *not* have a constant value. First, it is widely accepted that individuals have positive time preferences implying that future health benefits should be discounted.[8] Second, for some chronic dysfunctional states there are reasons to believe in diminishing marginal utility of succeeding years.[3] Third, as noted

in the introduction, when comparing the social value of health benefits between patient groups, and, cet par, numbers of years with improved health differ, equitable preferences imply that succeeding years in good health would be assigned diminishing social weights.

The formula (1) can now be extended to take account of these additional preferences. The first step is to take account of time preferences and include a discount factor, δ. Assuming a constant time preference rate, i, the present value of QALYs gained becomes:

$$QALYs_{PV} = \Delta HS [1 + \delta + \delta^2 + .. + \delta^{T-1}], \qquad (2)$$

where

$$\delta = \frac{1}{1+i}$$

While (2) is the present value of a stream with *constant* health state value, the second step is to allow for time period variations:

$$QALYs_{PV}{}^{HS} = [\Delta HS_1 + \Delta HS_2 \, \delta + \Delta HS_3 \, \delta^2 + ... + \Delta HS_T \, \delta^{T-1}] \qquad (3)$$

The ΔHS's indicates that the health profile over time is something separate from time preferences. If the health improvement diminishes over time at a constant rate, (3) changes to (3') where β becomes a factor less than unity. (If the change in health state increases at a constant rate with T, then (3') holds with $\beta > 1$.)

$$QALYs_{PV}{}^{HS} = \Delta HS \sum_{k=0}^{T-1} (\beta\delta)^k \qquad (3')$$

The third—and, in this paper, last—extension is to take account of distributive preferences. Generally, these preferences are equitable, when, cet par, a given total amount of QALYs distributed as a few to the many is preferred to one with many to the few. This implies that the social value of the k-th year to one person is lower than the $(k-1)$-st year to another person. Analogous to β and δ, we have a distribution factor, α, which is less than unity when preferences are equitable, and which, for simplicity reasons is assumed to be constant. By placing this factor *inside* the brackets of equation (4), it suggests that equity preferences refer to diminishing social valuation of a person's succeeding years with improved health. Thus, (4)

expresses the social value of QALYs, i.e. their present value (subscript $_{PV}$) and adjusted for changing health state values over time (superscript ^{HS}) and adjusted for equitable preferences (subscript $_E$):

$$QALYs_{PV}{}^{HS}{}_E = \Delta HS \sum_{k=0}^{T-1} (\alpha\beta\delta)^k \qquad (4)$$

Hence, if an increase in the number of QALYs has a diminishing social value, this means that the product of the three factors $\alpha\beta\delta$ is less than unity. This, what might be called, 'social adjustment factor' has an implicit 'social weight rate', r:

$$\alpha\beta\delta = \frac{1}{1+r} \qquad (5)$$

The challenge then is to find a method which could estimate this rate at which people would 'devalue' succeeding QALYs to one patient (group) relative to those received by another (group).

ELICITING SOCIAL WEIGHTS

A person-trade-off method

The essence of the person-trade-off technique is to present the two variables: number of persons benefiting as one, and health outcome (or severity level) as the other.[9] Respondents are asked to state which combinations of the two variables are considered equally good.

In this context one approach could be to compare the saving of a large number of lives each with few remaining life years with one which saves a small number of lives but each with a greater number of remaining life years. However, it is hard to imagine that people of the same age, and equal in all other respects, could have different expected remaining life years. It would involve the comparison of different *age groups* rather than different outcomes. The experience of the author with different pilot questionnaires suggests that respondents find it more meaningful to compare patient groups whose differences in health gains are couched in terms of the duration of a health *quality* improvement (ΔHS).

The suggested method is to have respondents state the number of persons **p** each of whom gains a long duration **T** of the health quality improvement which is considered *equally as good as* a given number of persons **P** each gaining a shorter duration t of the same health quality improvement. This preferred number **p** is taken as the 'indifference value':

$$U(p, T) = U(P, t) \qquad (6)$$

To isolate the effects of different outcomes between the two patient groups, a *constant social value of increasing numbers of persons* is assumed, which is also the assumption in the CEA/CUA methodology. Hence:

$$p*U(T) = P*U(t) \qquad (7)$$

In order to estimate the implicit social weight of increasing years, (7) is considered as two streams of benefits whose present values, each weighted by the number of recipients, are the same, and hence, changed to (8):

$$p*A_T = P*A_t, \qquad (8)$$

where:

$$A_j = \sum_{k=0}^{j-1} \left(\frac{1}{1+r}\right)^k = \frac{(1+r)^{j+1} - 1}{r(1+r)^j},$$

$$j = t, T, \quad r > 0,$$

A_j is the present value of a unitary benefit stream over the given period,

r is the implicit annual 'social weight rate', as in (5) above.

Equation (8) simply states that the present value of having **P** persons gain one unit of health over t years equals that of having **p** persons gain the same unit of health over **T** years. With **P**, t and **T** given, the rate **r** can be derived implicitly from the indifference value **p**. Going back to equation (5), this implicit rate **r** is the aggregated 'social weight rate'.

Some results from a pilot study

The focus of the study was on health enhancement with different durations. 'Imagine a chronic state of dysfunction and periodical pain. Treatment does not affect the length of life, but the health status is much improved'. The values P = 100 and T = 20 were used. In order to test for

generalisability, two different versions of the questionnaire were used, with t = 5 and t = 10. In the first sample of 90 undergraduate students in economics, the two versions were randomly distributed. The second sample was a class of 44 doctors attending a course in health administration. For this sample, only the questionnaire with t = 5 was used. Some brief background information was given, focusing on the importance of the issue in the context of using cost-effectiveness analysis in health care. Furthermore the author visualised the question using a two-axis figure with persons and years, and indicating the two equal areas when pT = Pt.

The questionnaire which was used presented discrete options with two fixed end points: one in which the total number of years would be the same (pT = Pt); and the other in which the number of persons would be the same (p = P). A guidance was provided which emphasised our interpretations of the respondents' preferred option (for pT = Pt; '... implies that you would **not** take account of how health gains are distributed', for p = P; '... implies that you would **not** take account of the length of time people benefit', and for p < P; '... you think that both factors should matter').

In the merged sample, 77% chose one of the internal options, revealing that they would make trade-offs between the number of persons and outcome in terms of the duration of health improvement. The minority was split into two equal sized groups, opting for either pT = Pt (i.e. total number of years should be the same) or p = P (i.e. number of persons should be the same). A comparison of the answers to the two versions of the questionnaire (t = 10 vs t = 5) indicates that the *smaller* the relative difference in effects (T − t = 10 vs 15), the greater is the proportion of respondents taking account of persons only, i.e. opting for p = P (17.4% vs 8.3%). Although based on a small sample, this finding appeals to intuition.

For the t = 10-questionnaire, the listed p-options were [50, 60, 70, 80, 90, 100]; when t = 5, the listed p-options were [25, 30, 40, 50, 70, 90, 100]. Table 1 shows the implicit rates computed from the stated indifference values. [In computing respondents' implied rates, those who had p = 100 (i.e. persons only matter) as their indifference value, would have had an infinite rate. For those respondents, the implicit rates of the highest *internal* option (p = 90) were used.]

Note that economics students and doctors hold

Table 1. Implicit 'social weight rates' (%)

Sample t	Econ. stud. 10	5	Doctors 5	Sum 5,10
Median	**14.9**	**13.0**	**13.0**	**13.0**
Mean	14.6	16.1	16.6	15.7
Std. err.	1.2	2.3	2.8	1.2
Valid N	46	44	40	130

similar views. Although the difference between the rates implied from the two versions of the questionnaire are small, one should be cautious in making inferences. There appears to be a bias towards the mid-point of the discrete options, so different intervals could yield different implicit rates. (Assuming mid-point bias, had all the intervals between the p-options been the same for the t = 5-questionnaire, then that would have increased the implicit rate.)

The focus of the approach used in this pilot study is one of social choice, in which the respondent is asked to focus on *other patient groups*. The implicit rates elicited from these trade-offs suggest how the respondent would value additional units of good health to her/his fellow citizen, when making comparisons of the social value of unequal amounts of health gains between patient groups. When making such person-trade-offs, we would expect distributional concerns to carry more weight, and hence that the implicit rates would be higher than they would if the choice had been related to the respondent's own streams of health.

The three most widely used methods for valuing health states [time trade-off (TTO), standard gamble (SG) and category rating (CR)] have one important common characteristic: respondents are asked to imagine *themselves* in the described health states. Using CR, a recent Finnish study[10] concluded that 'the duration seems to have no effects at all on the CR values'. However, the TTO appears to represent a more appropriate approach because it explicitly asks for which combinations of life years and health states the respondent would consider as equally good for her-/himself.

ELICITING PRIVATE WEIGHTS

Applying the time-trade-off method

The alternative method of deriving the implicit weight to attach to increased number of years,

when viewed as an individual good, is to use the TTO-technique but with two different time horizons as the base-line. One could describe an inferior health state which lasts for a period of T years, and ask how many years, T', (T' < T) in perfect health would be considered equally good. A shorter period of t years could then be used, thereafter asking in a similar fashion how many years, t', (t' < t) would be considered equally good. Now, if T'/T = t'/t, this could imply a constant individual value of increasing number of years, and no discounting. [Assuming that the TTO framing ignores distributional preferences, i.e. that $\alpha = 1$, this essentially means that $\beta\delta = 1$ in eq. (5).]

If T'/T < t'/t, this could still be interpreted as a constant value of the given health state, but now with a positive implicit discount rate. The constant value of the health state can be estimated using the following formula:

$$\frac{A_{T'}}{A_T} = \frac{A_{t'}}{A_t} \qquad (9)$$

where the A's represent the present values of the different health streams, as in equation (8). In other words the ratio of the present values is the same and reflects the actual health state value. The rate which satisfies the equality in (9) is then the individual's own rate for discounting future years. Hence this approach would give the implicit discount rate [the i of eq. (2)] *simultanously* with the constant health state value.

An example

A group of 10 MPH (master of public health)-students participated in this exercise. The questionnaire used the health state description: 'confined to wheelchair and able to move only with support, mild distress' (VI B in the Rosser matrix[11]). They were asked to make two TTO's with different durations in that health state; T = 20 and t = 5.

The figures of Table 2 should be interpreted as no more than numerical examples. However, the difference between the mean and the median figures shows an important characteristic of the formula in (9); when the proportion of remaining life years one is willing to trade off ((t − t')/t and (T − T')/T) increases as the reference time moves from t to T, the higher is the implicit rate. In other words, when the proportion one is willing to give up is higher the longer the duration, the higher becomes the implicit discount rate.

DISCUSSION

In this paper an analytical distinction has been made between 'unweighted' and 'weighted' QALYs'. Unweighted QALYs refer to the *production* of health. An intervention which gives 20 extra expected years in a particular health state is twice the health production of one which gives 10 extra expected years in the same state. However, that does not imply that people would give the former intervention twice the *social value*. Hence, it is an important distinction not to mix up health outcome with the social valuation of that outcome.

My focus has been on how health outcomes which differ in terms of the duration during which people benefit can be weighted. I have argued that when respondents state their preferences on this issue, they may be influenced by three concerns: 1) time preferences, 2) changing values of the described health state in succeeding years, and 3) equity preferences. Because the implicit rate may be influenced by such diverse preferences, it is hard to be firm with respect to which term to use: is it a time preference rate, an equity rate, or the rate at which one would devalue the health state per se?

In the context of person-trade-off, the term 'social weight rate' was used as an aggregate concept. When using the time-trade-off method, the conventional term 'discount rate' was used, although this might include diminishing health state value as well as time preferences. One way of disentangling the equity preference from the aggregated 'social weight rate', as implied from

Table 2. Implicit discount rates for own health

Time horizon	T = 20	t = 5		
Median	T' = 15.5	t' = 4.0:	i ≈ 2%	HS ≈ 0.81
Mean	T' = 12.9	t' = 3.98:	i ≈ 10%	HS ≈ 0.84

given to the issue of using social weights according to the characteristics of the patient group, i.e. that 'a QALY is no longer a QALY'. Williams[21] has shown that individuals weight the importance of being in good health according to life stages, while Charny et al.[22] argued for the introduction of a third dimension to QALYs which would depend on the characteristics of the patient group. Due to the immense ethical problems related to the weighting of QALYs depending on who you are, I share Williams'[23] scepticism with respect to the implementation of this type of weight.

The focus of this paper has been on a different weighting issue, not depending on *who* you are, but depending on *how many* QALYs you get. Certainly, more empirical research based on the suggested methods is needed before one can place numerical values on the parameters which have been introduced.

ACKNOWLEDGEMENTS

REFERENCES

1. Blades, C. A., Culyer, A. J. and Walker, A. Health service efficiency: appraising the appraisers—A critical review of economic appraisal in practice. *Social Science and Medicine* 1987; 25(5): 461–472.

2. Loomes, G. and McKenzie, L. The use of QALYs in health care decision making. *Social Science and Medicine* 1989; 28(4): 299–308.

3. Sutherland, H. J., Llewellyn-Thomas, H., Boyd, N. F. and Till, J. E. Attitudes toward quality of survival: The concept of 'Maximal Endurable Time', *Medical Decision Making* 1982; 2(3): 299–309.

4. McNeil, B. J., Weichselbaum, R. and Pauker S. G. Speech and survival. Trade-offs between quality and quantity of life in laryngeal cancer. *The New England Journal of Medicine* 1981; 305: 982–987.

5. Sackett, D. L. and Torrance, G. W. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 1978; 31: 697–704.

6. Vesth Pedersen, A. *QALYer som nyttemål: Et forsøg på empirisk testing av forutsætningerne for en sådan tolkning.* Paper presented to the 14th Nordic HESG-meeting. Tromsø: 18–20 August 1993.

7. Williams, A. Economics of coronary artery bypass grafting. *British Medical Journal* 1985; 291: 326–329.

8. Drummond, M. F., Brandt, A., Luce, B. and Rovira, J. Standardising methodologies for economic evaluation in health care: Practice, problems and potential. *International Journal of Technology Assessment in Health Care* 1993; 9(1): 26–36.

9. Nord, E. Methods for quality adjustment of life years. *Social Science and Medicine* 1992; 34(5): 559–569.

10. Ohinmaa, A. and Sintonen, H. *The effect of duration on the value given to the EuroQol states.* Paper presented to the 14th Nordic HESG-meeting. Tromsø: 18–20 August 1993.

11. Kind, P., Rosser, R. and Williams, A. Valuation of quality of life: some psychometric evidence. In Jones-Lee, M. W. (ed.) *The value of life and safety.* Amsterdam: Elsevier, 1982.

12. Cairns, J. *Discounting health benefits.* Paper for the 2nd European conference in Health Economics. Paris: December 1992.

13. Cairns, J. *Valuing future benefits.* Paper for the 14th Nordic HESG-meeting. Tromsø: 18–20 August 1993.

14. Cropper, M., Aydede, S. and Portney, P. Rates of time preference for saving lives. *American Economic Review* 1992; 82(2): 469–472.

15. Olsen, J. A. Time preferences for health gains: An empirical investigation. *Health Economics* 1993; 2: 257–265.

16. Nord, E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy* 1993; 24: 227–38.

17. Moore, M. J. and Viscusi, W. K. Discounting environmental health risks: New evidence and policy implications. *Journal of Environmental Economics and Management* 1990; 18 S: 51–62.

18. Harris, J. QALYfying the value of life, *Journal of Medical Ethics* 1987; 13: 117–123.

19. Nord, E. The relevance of health state after treatment in prioritising between different patients. *Journal of Medical Ethics* 1993; 19(1): 37–42.

20. Parfit, D. *Reasons and persons.* Oxford: Clarendon Press, 1984.

21. Williams, A. Ethics and efficiency in the provision of health care. In Bell, J. M. and Mendus, S. (eds.) *Philosophy and medical welfare,* Cambridge University Press, 1988.

22. Charny, M. C., Lewis, P. A. and Farrow, S. C.

Choosing who shall not be treated in the NHS. *Social Science and Medicine* 1989; 28(12): 1331–1338.

23. Williams, A. Cost-effectiveness analysis: is it ethical? *Journal of Medical Ethics* 1992; 18(1): 7–11.

Paper 7
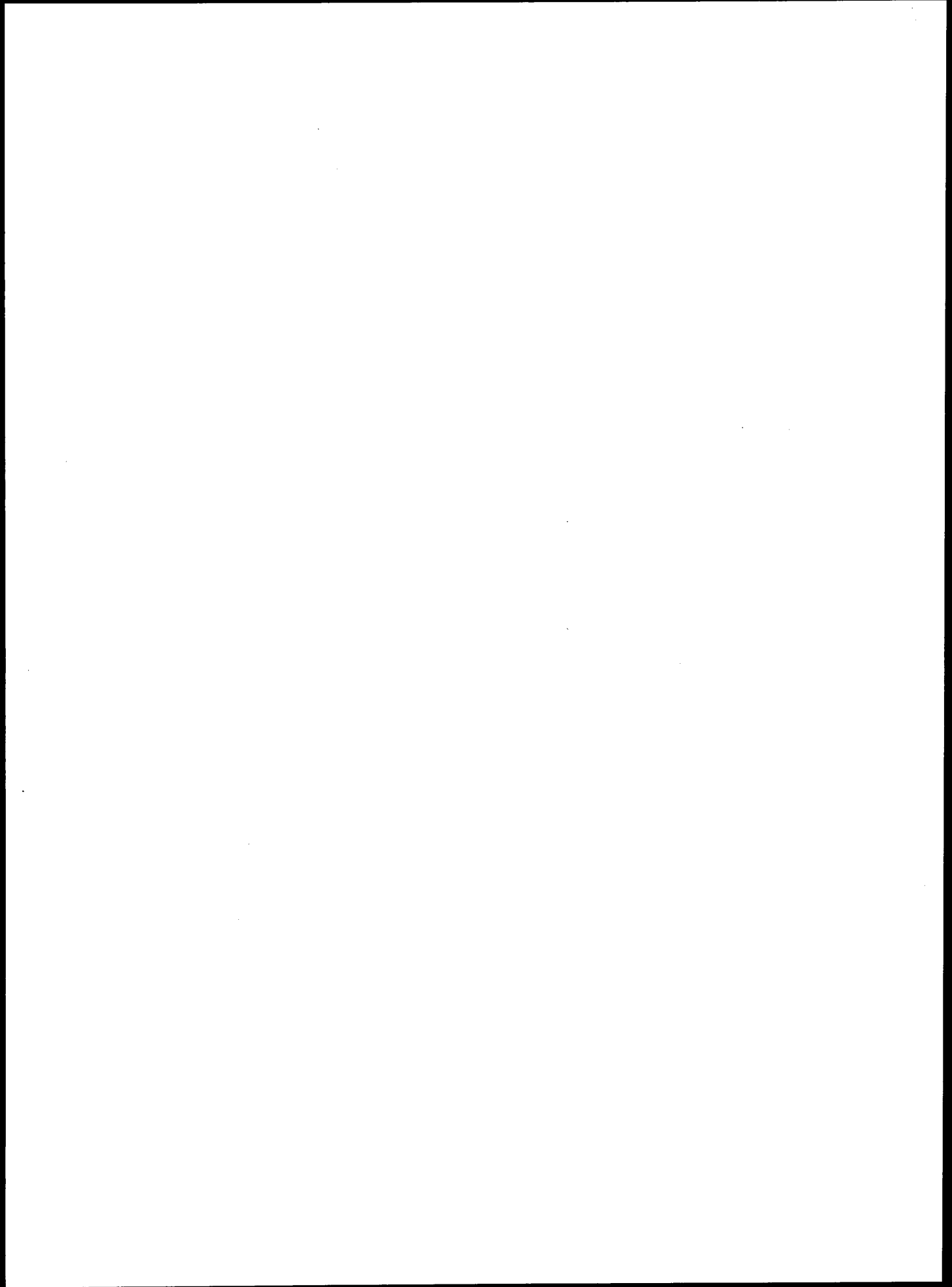
Willingness to pay
for public sector health care programmes

Jan Abel Olsen
University of Tromsø
Department of Economics, NFH
S. Tollbugt 7
9037 Tromsø
NORWAY

and

Cam Donaldson
Health Economics Research Unit
University of Aberdeen
Foresterhill
Aberdeen AB9 2ZD
SCOTLAND

Abstract:

The paper reports from a study which asked 150 interviewees their willingness to pay (WTP) in increased earmarked taxation for three different health care programmes; a helicopter ambulance service, more heart operations and more hip replacements. Reasons behind the stated WTP were asked for. OLS-regression analyses were used to analyse factors associated with WTP for each of the three programmes, and factors associated with the relative WTP for one programme compared with the total of the three. Comparisons were made of WTP for these programmes and the health outcome in terms of quality adjusted life years (QALYs).

1: Introduction

In cost-utility analysis of health care programmes, outcome is valued in terms of QALYs (quality adjusted life years) gained. The only characteristic that is accounted for in such measures is health enhancement, and this health enhancement has the same value no matter the characteristics of the programme which produces it. Yet, health care programmes appear to have utility-bearing attributes beyond health gain. Mooney (1991) argues that, beyond the benefit derived from improved health outcome, there are additional 'non-health benefits' from health care use, e.g. information. Eddy (1991) argues that health outcome is valued differently depending on the 'the nature of the services' which produce them. Hence, the 'Rule of Rescue' can be accounted for by weighting the QALY-outcome from rescue-type health care higher relative to the same QALY-outcome from other programmes. Furthermore, there is a large literature on equity concerns with regard to health care (Donaldson and Gerard 1993). The implication is that a health care programme which improves equity has an additional valuable attribute.

One problem with using QALYs as a benefit measure is that, in their current form, they disregard these additional characteristics of health care which people may value. This convention in the health economics literature goes against the long-established Lancastrian approach to consumer theory which holds that, in general, a good will possess more than one characteristic (Lancaster 1966). Two of the key problems in valuing health care programmes lie in identifying, and including, valuable additional characteristics of health care beyond those which are picked up by the health outcome measure.

The aim of this paper is to investigate whether the use of a different valuation method, based on willingness to pay (WTP), could prove fruitful in these tasks. In particular WTP values were elicited from members of the community with regard to three different public sector health care programmes. Although the main focus of the paper is the relationships between respondents WTP and their demographic characteristics and attitudes, QALYs gained from each programme are also compared with the WTP values.

In the following section, a summary of the state of the art of WTP as applied to health care is presented. Differences between previous studies and the one presented in this paper are highlighted. In the third section, the chosen method for the study is outlined. The key results are then presented. The discussion in section five focuses, amongst other things, on the issue of WTP per QALY gained for each of the health care programmes. It is suggested that the higher the stated WTP per QALY in one programme compared with another, the more valued is the health outcome and/or the more influenced is the value of this health care programme by characteristics other than its health enhancing capabilities.

2: Some received wisdom

Although quite a few studies have been carried out in the neighbouring sub-disciplines of environmental and transport economics (Johansson 1987, Jones-Lee 1989), there are still a relatively small number of references to WTP for health or health care programmes. Since Acton (1973), studies have been sporadic, but the research effort is now taking off again. According to a count by Donaldson (1993), there are 18 studies of WTP for health or health care published in English language sources. Advances have been made in these studies, but there are still some problems with them. There are three such problems, in particular, which the study reported in this paper attempts to overcome.

2.1 Problems with previous studies

Focus on own health improvement only
Most studies concentrate on the degree to which health care will improve respondent's (or her/his family's) *own* health (Thompson 1986, Johannesson 1992). In the cited cases it is for complete elimination of athritis and for a reduction in lipid levels respectively. The implicit assumption is that there are no other benefits from health care.

Out of pocket payments for individual goods
As in the above cited studies, it is common for preferences to be elicited from those who already are, or could be, users of the service being evaluated. This approach might be

relevant in settings in which users pay for the full cost of health care (through charges or private insurance), and in which health care is viewed as an individual good. However, the approach becomes less relevant in settings in which the wider community pays, and in which health care is considered to have utility-bearing characeristics in the form of "caring externalities" (Culyer 1971). Assuming that the mode of payment may affect one's stated valuation (Mitchell and Carson 1988), out-of-pocket-framed WTP questions may not be applicable to tax-funded health services.

Studies have been partial, involving no comparisons between programmes
In none of the 18 studies referred to above were WTP-values elicited for more than one type of health care. The problem with doing such partial WTP studies is that, in a priority setting context, one would not have any idea of how the particular programme performs relative to other health care programmes on which the resources could be spent, i.e. the true opportunity cost is not taken into account. Furthermore, partial studies may also create an 'embedding problem' (see e.g. Arrow et al 1993), i.e. that respondents have a tendency to overestimate their valuation of the particular programme, because it is seen as representing a wider health sector for which one has a general positive attitude.

2.2 Overcoming these problems

The ways in which the study reported in this paper attempts to overcome the above problems are as follows:

Wider descriptions of the programmes were presented to respondents. These focused on:
- the actual health care programme to which resources could be allocated,
- a statement of patients' initial health problems, and health states after treatment,
- duration of the improvement,
- characteristics of those benefiting; in terms of age, sex and where they live.

Respondents are asked for their *willingness to contribute in terms of extra earmarked taxation* per annum. As such, we have a community programme focus, which is most relevant in societies were the health service is tax financed. *Reasons for respondents' WTP*

4

were asked for, in order to facilitate an assessment of whether wider community benefits were taken into account.

Three alternative programmes were presented, in order to make the respondent focus on their valuation of each particular programme rather than viewing an individual programme as one which represents the health care sector as a whole. Comparisons can then be made between the social valuations of alternative ways of allocating health care resources. More important, however, is the possibility of using this approach to gain increased understanding of which attributes of the different health care programmes are valued.

3: The study

3.1 Background to the study

The original intention was to compare the value of spending resources on a helicopter ambulance service with the values of spending these resources on other health care programmes. A thorough cost effectiveness study had made available unique outcome data on the helicopter ambulance (Kristiansen 1992). The focus was on the *marginal effects* in terms of increased survival probability and improved quality of life that could be ascribed to the shorter transport time with the helicopter compared with the alternative ground ambulance. The mean time saving was 69 minutes for those who were transported. The target region for this helicopter is Troms county (excluding the cities of Tromsø and Harstad) which is one of the three counties in the region of Northern Norway.

The helicopter ambulance service was considered to be a particularly interesting candidate on which to carry out a WTP study. The service carries other assumed utility-bearing characteristics: 1) it is a "rescue-service" which may reduce anxiety among people in rural areas; 2) although few people use or benefit directly from the service, it may have a considerable "option value" derived from knowing that the service is available should use in the future be required; and 3) it contributes to reducing regional inequalities in access.

Two alternative health care programmes were presented, each having the same annual costs [10 mill. Norwegian kroner (£ 1 = NOK 10)] as the helicopter service. The criteria for choosing the alternatives were: first, it was likely that people were familiar with them, so brief descriptions were sufficient; and second, in comparing alternatives with a rescue-type programme such as the helicopter, we wanted one programme which produces health gains in the form of both life extension and improved quality of life, and one programme which produces improved quality of life only. The two programmes chosen were 80 more coronary artery bypass operations and 250 more hip replacements.

Rather than describing the programmes in individual probability terms, as recently advocated by Gafni (1990), they were framed as being to the benefit of the collective community, i.e. for people living in the county or region. However, the probabilities were described implicitly. For the helicopter, respondents were informed how many people used the service and how many live in the target region. For the other two programmes, the relative increase in provision was emphasised. Since most people would have a fairly clear idea of how many live in Northern Norway, it was decided not to include an additional number on the cards describing the latter two programmes.

3.2 Study design

The study was carried out in October 1992. A member from each of 150 households in Troms county was interviewed; 75 in Tromsø, 25 in a smaller city and 50 in the rural area. People were selected in the following way: in the two cities, the telephone book was used to randomly select 15 and 5 streets respectively, in each of which five people were interviewed. In the rural area 5 municipalities in different parts of the county were chosen, in each of which ten people were interviewed. There, the interviewers started from two different locations, one in the municipality centre and one in a sparsely populated area.

From the starting address in each street/municipality the interviewers went along to each house (by ascending number) until they achieved the required quota of respondents. In total there were 135 who would not participate. The usual excuse was that they were busy, but some also expressed general disquiet at being interviewed. Only one of the refusals

expressed specific aversion to this type of health care evaluation. Six trained interviewers were used, and the average interview time was 35 minutes.

The introductory information for respondents focused on the proposed method of measuring individuals' valuations of health care programmes, i.e. in terms of *how much one would be willing to contribute through increased taxation which would be earmarked to health care programmes*. Respondents were also informed of the three programmes to be addressed in the interview. As the interview proceeded, the respondents were given a yellow card with a brief description of the helicopter ambulance service, and then a white card containing the same description with the payment options listed (see appendix). This is a common way of eliciting WTP values (Mitchell and Carson 1984). The alternative of providing individuals with one WTP figure and asking for a "yes"/"no" response to whether they would be willing to pay that amount, as proposed by Arrow et al (1993), requires a larger sample size than that used in this study.

After having circled the amount which was the maximum they would be willing to pay each year, respondents were asked, in open ended questions, to explain the reasons why they were willing to pay. Then the interviewer read out some precoded reasons which were thought to influence valuations. A visual scale on degree of importance for being willing to pay was handed to the respondents. They were then asked to rate the reasons on this scale, which was presented with *equal spacing* between the degrees of importance. To each of the four points on the scale, we attached the following values: 0 for "No", 1 for "Little", 2 for "Some" and 3 for "Much" importance. The same procedure was used for the heart programme and after that the hip programme. For each programme a partial valuation was emphasized, WTP questions being accompanied by the statement "... if this were the *only* programme to be implemented?".

Before giving their valuations respondents were asked to compare the yellow cards and prioritise between them. Those who were not willing to pay anything in extra taxation, because of tax aversion, were asked whether they would be willing to give a voluntary donation to the different programmes or pay through a private insurance premium. If so, they were asked WTP questions using their chosen mode of payment. It was explained

that the study was concerned more with their valuation of a service than how one would wish to pay. At the end of the exercise, the interviewer read out respondents stated WTP figures, and allowed them to change the figures if they so desired. Finally demographic details were asked for.

4: Results

4.1 How much were people willing to pay?

Of the 150 completed interviews, seven were excluded because respondents would neither express any WTP figure nor would they prioritise. Some of them expressed reservations about the whole exercise.[1]

Table 1: *Willingness to pay per year (in NOK)*

N = 143	Helicopter ambulance	80 more heart operations	250 more hip operations
Mean	316	306	232
St err	25	25	21
Median	200	200	150

There were no statistically significant differences between the mean WTP values for helicopter and hearts, but the value for hips was different from both helicopter and hearts (using Wilcoxon signed rank test, $p < 0.0001$). Because we have fixed end points, and that the intervals between the discrete payment options increased (from NOK 25 to 200) as the options themselves increased, normal distributions were not expected. Hence, for the following analysis, the three dependent variables, i.e. the WTP for each programme, were log-transformed.

---

[1] Nine respondents were not willing to pay for any of the three programmes in any of the three modes of payments, but they *were* willing to prioritise. They are included.

4.2 Which factors influenced willingness to pay?

Ordinary least squares regression analysis was used to test for factors associated with WTP. Table 2 gives a specification of the independent variables.

Table 2: *Variable specification*

SEX	= sex; 1 for male, 0 for female
EDUC	= education level; 1 for A-level, college or university, 0 otherwise
MUNICIP	= urban/rural split; 1 for the two cities, 0 otherwise
OWNHE	= own health, 1 for good/very good, 0 for fair/poor
INCPERS	= income in NOK 100,000s, adjusted for number of persons in the household (OECD weights: 1 for first adult, 0.7 for additional adults, 0.5 for each child). The income was taken as the midpoint of the income interval (NOK 50,000 from 0 to 500,000+)
AGE	= age, in years
AGESQ	= AGE squared
TIMEINT	= length of interview, in minutes (range 20 - 50)

The last two variables of Table 2 require further explanations. The AGESQ variable was introduced because, for hearts and hips, the age variable was expected to have a curvilinear form with respect to WTP. The variable TIMEINT is introduced as an indicator of a possible "income effect". Given the chosen sampling procedure, one identical questionnaire with the same ordering of the programmes was preferred. Although the partial valuation was emphasised, we would hypothesise that the longer the duration of the interview, the more time would the respondent have to sit back, forget the previous WTP figure and accept the partiality of the next valuation. In other words, the faster one gives a WTP figure for the second and third programme, the more likely it is that one's overall budget constraint will come into play. Answers will then be influenced by an income effect.

Hypotheses with respect to the influence of these independent variables on WTP were as follows (the superscripts refer to the independent variable, and the subscripts indicate the programme; 1 = helicopter, 2 = hearts and 3 = hips):

H^S_2 men are willing to pay more for hearts.

H^S_3 women are willing to pay more for hips.

H^E_{1-3} education level matters.

H^M_1 people in the rural area are willing to pay more for the helicopter.

H^O_{2-3} people in fair/poor health are willing to pay more for hearts and hips.

H^I_{1-3} the higher the income the higher the WTP.

H^A_{2-3} at the lower age levels; the higher the age, the higher the WTP for hearts and hips.

H^{AQ}_{1-3} generally, the WTP decreases at old ages.

H^T_{2-3} the longer the length of the interview, the higher the WTP for hearts and hips.

The same regression model was used to analyse WTP for each programme. This permits comparisons of coefficients across programmes.[2] Table 3 shows the results from the multiple regression analysis. There is no evidence to support H^S_2. For hips, women were willing to pay significantly more, which supports H^S_3. The education level had a significant (negative) impact on WTP for the helicopter, providing support for H^E_1. There is not sufficient support for H^E_{2-3}, H^M_1, H^O_{2-3} or H^I_{1-3} .

When age was entered linearly, its coefficients were negative for all three programmes ($p < 0.01$). Although, when entered quadratically, the coefficents of AGE and AGESQ were not significant any more, the important point is that, for hearts and hips, the sign for AGE switched from negative to positive (indicating H^A_{2-3}). The negative coefficients of AGESQ indicate that the WTP falls at older ages (H^{AQ}_{1-3}).[3]

---

[2] It was hypothesised that people with children are willing to pay more for the helicopter. However, when a dummy for child(ren) under 16 in the household was introduced as an additional variable, it had no influence on WTP. Again, because the same model was preferred for all three programmes, this independent variable is not included.

[3] When AGE was excluded, AGESQ came out as significantly negative in all three programmes ($p < .001$).

Table 3: *Factors influencing WTP (n = 143)*

Independent Variable	Dependent Variable		
	HELICOPTER Log WTP B (SE B)	HEARTS Log WTP B (SE B)	HIPS Log WTP B (SE B)
Constant	6.1543 *** (1.4672)	3.0776 (1.7140)	1.0710 (1.5995)
SEX	-.8623 ** (.2922)	-.4894 (.3413)	-1.2900 *** (.3185)
AGE	-.0178 (.0567)	.0333 (.0663)	.0703 (.0618)
AGESQ	-.0005 (.0005)	-.0008 (.0006)	-.0011 * (.0006)
INCPERS	.3620 (.2756)	-.0724 (.3220)	.2732 (.3004)
MUNICIP	-.2923 (.3328)	-.5933 (.3888)	-.2364 (.3629)
OWNHE	-.1807 (.3083)	-.4433 (.3601)	.3327 (.3361)
EDUC	-.7735 * (.3201)	-.5279 (.3740)	-.1388 (.3490)
TIMEINT	.0301 (.0216)	.0835 *** (.0252)	.0801 *** (.0235)
Adjusted R Sq	.241	.144	.205
Signif F	.0000	.0003	.0000

B = coeffisient, SE B = standard error of the coefficient

*   = $p < .05$
**   = $p < .01$
***   = $p < .001$
two-tailed
one-tailed if according with a hypothesis: SEX Hips, AGESQ Hips,
TIMEINT Hearts and Hips

As to the importance of the length of the interview, there is strong evidence in support of H^T_{2-3}, although the coefficients are small. Interestingly, there is no significant impact of TIMEINT on willingness to pay for the first programme, i.e. helicopter, which also supports our assumption of an income effect.[4]

While in Table 3 each of the three programmes is considered in isolation, the next model analyses what determines the *relative preference* of one programme. As the dependent variables, the share of the WTP for the programme concerned over the total WTP for the three programmes was used, or more precisely; *the share of the sum of the partial WTP*, i.e. a type of a "budget share". Assuming partial valuations, this model will indicate how the relative WTP depends upon the background variables.[5] The independent variables used in Table 4 are the same as those in Table 3 except that AGESQ is excluded.

Note that, for each of the independent variables in Table 4, the sum of the coefficients over the three programmes is zero. This is explained by the fact that the sum of the budget shares is always unity.

In Table 4, it can be seen that men have a significantly higher *relative* WTP for hearts, which now lends support to H^S_2. A strong relative preference for hips is associated with women, good own health, high income, old age and, perhaps, higher education. Furthermore, for the helicopter ambulance, the young have the strongest relative preference.

---

[4] When running identical regressions after excluding those nine respondents who did not have a positive WTP for any of the three programmes, the same pattern emerged with respect to the impact of TIMEINT ($p < .01$ for hearts and hips, with B close to zero for helicopter).

[5] In computing Table 4, those nine respondents who did not have a positive WTP for any of the three programmes were omitted, simply because it is meaningless to analyse budget shares when the budget is zero.

Table 4: *Factors influencing WTP of each programme over the total WTP (n = 134)*

	Dependent Variable		
	HELICOPTER share WTP	HEARTS share WTP	HIPS share WTP
Independent Variable	B (SE B)	B (SE B)	B (SE B)
Constant	.8696 *** (.1206)	.1987 (.1010)	-.0684 (.1022)
SEX	.0236 (.0336)	.0546 * (.0281)	-.0782 ** (.0285)
AGE	-.0045 ** (.0014)	.0021 * (.0012)	.0023 * (.0012)
INCPERS	.0043 (.0298)	-.0606 * (.0249)	.0562 * (.0252)
MUNICIP	.0155 (.0379)	-.0301 (.0318)	.0146 (.0321)
OWNHE	-.0243 (.0359)	-.0390 (.0300)	.0633 * (.0304)
EDUC	-.0543 (.0370)	-.0068 (.0310)	.0612 (.0313)
TIMEINT	-.0079 ** (.0026)	.0044 * (.0022)	.0035 (.0022)
Adjusted R Sq	.109	.123	.138
Signif F	.0029	.0012	.0005

B = coeffisient, SE B = standard error of the coefficient

*   = $p < .05$
**   = $p < .01$
***   = $p < .001$
two-tailed
one-tailed if according with a hypothesis: SEX Hearts and Hips, AGE Hearts and Hips

4.3 Why were people willing to pay?

The open ended reasons

For each programme, after the respondent had stated their WTP, they were asked an open ended question about the reasons for their positive WTP. The stated reasons were grouped, allowing each respondent to give more than one reason. The most frequently stated reasons (with numbers stating the reason in brackets) for the helicopter were: "time saving" (49), "life saving" (38), "good for the rural areas" (28), "accidents" (20), "reassuring that it exists" (20) and "I/my family might benefit from it" (16).

Assuming that these reasons reflect characteristics which are positively associated with the level of WTP, they were entered as dummies in a regression analysis together with INCPERS (i.e. adjusting for income differences), and with the log-transformed WTP figure as the dependent variable. The dummies were given values of 1 if a respondent mentioned the reason and 0 otherwise. However, these reasons did not explain anything, except for "time saving" (B=.53 and p<.01).

The open ended reasons for WTP *for hearts* were grouped as follows: "I/my family might benefit from it" (36), "life saving" (19), "heart queues" (15), "human right to have an operation" (12), "improved quality of life" (11) and "more operations needed" (9). *For hips*, the open ended reasons for WTP were: "improved quality of life" (62), "efficient for society" (28), "I/my family might benefit from it" (17), "women" (13), "caring attitude" (12) and "many people will benefit from the programme" (8). When entered as dummies together with INCPERS with the log-transformed WTP figure as the dependent variable, there were no significant relationships between the stated reasons and WTP for either hearts or hips.

While these reasons may add to an understanding of positive WTP, they do not explain the variance within WTP. After eliciting these open ended reasons, there was a structured precoded exercise which attempted to shed further light on what reasons influenced WTP.

14

The precoded reasons for WTP

The mean values for each of the reasons presented under each programme are given in Table 5.[6] There are open spaces in Table 5 because not all reasons were presented for all programmes.

Table 5: *Mean scores on the precoded reasons for WTP*
Scale on degree of importance [0 - 3]

	Helicopter	Hearts	Hips
You/your familiy could need it	2.15	2.58	2.24
Other people could need it	2.78	2.67	2.50
Reassuring to know that it exists	2.73		
More equal access to health care regionally	2.61		
Don't want a health service to close down	2.29		
"Heart queues" and waiting lists		2.57	
Some will be able to return to work		2.14	
Elderly will be able to manage themselves and will need less help from others			2.77
Elderly people deserves better health			2.46
N	124	118	118
missing	2	3	2
missing because unwilling to pay	17	22	23

In Table 5, the first two reasons for people's valuations can be interpreted as indicators for the importance of selfishness and altruism respectively. The non-parametric Wilcoxon signed ranks test showed that ratings for altruism are significantly different from ratings for selfishness for helicopter ($p<.0001$) and for hips ($p=.0209$), but not for hearts.

The "option value" of the helicopter ambulance can also be considered as a selfish concern. The reason "Don't want a health service to close down" can be interpreted as an indicator of "possession preference" (i.e. not wanting to give something up once it already

---

[6] Within each programme, roughly 25% of the respondents had consistently given the highest score to all the listed reasons. Since this group appears to be less reflective on this exercise, there is an argument for excluding them in Table 5. However, although the mean values increase towards the upper end of the scale, their inclusion does not affect the ordinal ranking of the reasons under each programme.

exists).[7] The selfish reason is strongest for the heart programme, but there is also much concern for "heart queues". The "efficiency argument" that some patients will return to work got the lowest score. However, for the hip programme, the fact that "Elderly will be able to manage themselves and will need less help from others" gained the highest score.

For each of the three programmes these precoded reasons from Table 5 were entered as independent variables in a regression analysis together with INCPERS, and with the log-transformed WTP figure as the dependent variable. The major results can be summarised as follows:

For the helicopter ambulance, the ratings for the "selfish" reason and the "option value" were merged into *one* "selfish" variable.[8] This came out with a positive influence on WTP (B=.29 and p<.01). The "altruistic" variable, however, was a slightly more important determinant (B=.33 and p<.05). In a separate regression analysis with the four attitude variables entered together with the background variables of Table 2,[9] the importance of altruism for WTP came out even stronger (B=.42 and p=.015).

For the heart programme, the altruistic reason came out with the highest positive sign, but it was not significant. Hence, it was merged with the reason "heart queues and waiting lists" into an altruism variable. Acknowledging that this may also reflect a selfish attitude, the selfish reason was also merged with "heart queues and waiting lists" to a selfish variable. When regressed together with the reason "some will be able to return to work" and INCPERS, the "altruistic" variable came out as the most important one for WTP

---

[7] The helicopter ambulance is an existing service. People were asked to imagine that the money for the helicopter were to be transferred to other health services, and state how much they were willing to contribute to prevent it being closed down.

[8] For each respondent, the mean value of the two reasons was used.

[9] Acknowledging that the distributions of the degree of importance attached to these reasons were far from normal (many claimed the reasons were of "Much" importance), we examined the effect of entering these variables as dummies (1 for "Much", 0 otherwise). The results were close to those reported, except that the explained variance was slightly less.

(B=.37 and p<.05), while the selfish variable had no impact (B=.10 and p=.6). In a separate regression analysis with the three attitude variables entered together with the background variables of Table 2, the importance of altruism for WTP was still apparent (B=.33 and p=.062), while the selfish one was not significant.

For the hip programme, "Elderly will be able to manage themselves and will need less help from others" was the only significant reason for WTP (B=.52 and p<.001) when regressed together with the other three attitude variables and INCPERS. The altruistic reason was not significant. The "Elderly will be able ..."-reason came out with the same B- and p-values in a separate regression analysis with the four attitude variables entered together with the background variables of Table 2.

The interpretation of the B-coefficents of these attitude variables is as follows: for each unitary level of importance (from 0 to 3), the B-value gives the predicted percentage increase in WTP.

4.4 Why were people *not* willing to pay?

Those who were not willing to pay earmarked taxes for the helicopter ambulance were asked whether they would be willing to pay in other ways. When faced with the two following programmes, this group were also asked which way they would prefer to pay. Among those with positive WTP, the vast majority were willing to pay in terms of taxes. However, the proportion differed from 73% among those willing to pay for helicopter, 83% for hearts and 88% for hips.

The number of respondents who were *unwilling* to pay in any way were 17 for the helicopter, 22 for hearts and 23 for the hip programme. The stated reasons were grouped, again allowing each respondent to give more than one reason. The most widely expressed reasons were: "the public sector has enough money to pay, so reallocation should be possible" (7 for helicopter, 9 hearts, 11 hips), and "can't afford to pay" (5 helicopter, 3 hearts and 4 hips).

For the helicopter ambulance, 11 held that "other health services are more important" while 3 meant it was "overused and inefficient". The straight selfish reason "I won't need it myself" was expressed by 4. *For the hearts programme*, 5 said it is a "status-activity which has got enough resources already", and 3 held that it is "self-inflicted life style illness". A further 3 said that it was "inefficient" without further explanations. *For the hips programme*, 4 gave the reason that "it doesn't save lives", while 2 said "no more taxes".

There appears to be two types of reason for not being willing to pay. One is the view that there are enough resources in the public sector to finance these programmes, so further payments should not be required. This view does not necessarily reflect zero *valuations* of the programmes. However, the other, more common, reasons are more indicative of zero valuations.

5: Discussion

Compared with previous WTP studies of health care, this study is extensive. Of the 18 studies reviewed by Donaldson (1993), only 8 used face-to-face interviews, and only 2 of these had a sample size of over 100 (Thompson 1986, Berger et al 1987). As with all others, these two studies focused only on health gain, the first in a rather unrealistic way (looking at WTP of rheumatoid arthritics' for the complete elimination of their condition). All studies, except that of Berger et al (1987), examined WTP in relation to one programme only. In just two studies (Berger et al 1987, Loehman et al 1979) were members of the community at large surveyed. As the result of the limitations of the above studies, broad comparisons of health care programmes by use of WTP has not been possible. The study presented here represents the first attempt to make such comparisons.

Recently an expert panel in the US gave some recommendations as to how a WTP study should be designed (Arrow et al 1993). Although their paper was published *after* our study was undertaken, we would like to draw the attention to the fact that the views expressed by that panel to a large extent correspond with our approach. In particular, a "conservative design" was used: 1) respondents were asked for their WTP rather than willingness to

18

accept; 2) respondents were reminded of their budget constraint; 3) because three programmes were compared, there is less chance of an "embedding effect"; 4) the scale of the payment card had a reasonable upper limit, the midpoint being close to the actual cost per household [and 5) WTP per annum was asked for and not WTP per month]. Given the above, the absolute values of the mean and the median WTP figures reported in our study appear reasonable.

Nevertheless, some outstanding issues remain to be discussed. These are: how such results might be used in policy making; what research is needed to further develop the method; and how WTP compares with measures of health outcome, like QALYs.

5.1 Policy implications

The partial valuation of each of the three programmes shows that the benefits exceed the costs in each case. However, for allocating a given health service budget, it is the relative benefit/cost ratios which are more important. While the helicopter is to the benefit of people in Troms county the other two programmes are for the entire region of Northern Norway, covering three times as many people. Thus, while the average costs of the helicopter per household in Troms county is NOK 200, the average costs of each of the other two programmes per household in Northern Norway is NOK 70. This implies that the benefit/cost ratio is highest for hearts and lowest for the helicopter ambulance. Using the median WTP figures (see Table 1), again in a partial sense, the indication is that a majority would favour the implementation of each programme. This gives the same ordering as with the mean WTP, i.e. the median benefit/cost ratio is highest for hearts and lowest for the helicopter. Strictly, a true comparison would require values elicited for all three programmes from a sample of people in the other two counties in Northern Norway.

5.2 Methodological issues

Publicity bias

The possibility of some biases in favour of the helicopter programme has to be acknowledged. During the months before interviewing there was much "turbulence" in the

local media - particularly in Tromsø - as to the life-saving importance of a new landing site for the helicopter ambulance.[10] This focus might have influenced the respondents to give higher valuations for the helicopter, although it was explicitly stated that this study was not concerned with the landing site. Although the mean WTP was higher in Tromsø than in the other city, in each city, the helicopter's mean value was equal to that of hearts, while the rural region valued the helicopter highest. This might indicate that the absolute WTP for the helicopter was influenced by media attention, but its relative performance in each city seems to have been unaffected. However, the figures do lend some support for the suggestion that the WTP for the first programme affects WTP for the subsequent ones.

Ordering bias

The lower valuation of hips could be influenced by an ordering effect. Although the partial nature of the valuations was explicitly stated, the interviewers got the impression that when being faced with the last of the three health care programmes, some respondents felt that they had contributed to two programmes (done their "fair share") and could not afford more (income effect), and/or they were tuned in on life saving which was not a part of the outcome of the hip programme. However, at the end of the exercise, an attempt was made to reduce the impact of an ordering effect by reading back the ranking of the programmes and their WTP figures. Respondents were then allowed to change the ranking and the figures.[11] Twenty took the opportunity and made changes.[12] However, given that such changes require extra time and a concession that one has changed one's mind during the interview, it may be that more people should have taken this opportunity.

---

[10] The doctors' preferred choice would save 30 seconds compared to the cheaper alternative, and some of them had been very active in arguing for the life-saving importance of this marginal time saving.

[11] Some interviewees felt it was difficult to be sure about how much they could be willing to pay, but held that their valuations should be interpreted as their relative ranking between the three programmes.

[12] There were 7 changes in WTP figures for the helicopter (5 up and 2 down), 11 changes on hearts (8 up and 3 down) and 10 changes on hips (8 up and 2 down).

The influence of the length of the interview (Table 3) indicates a general methodological lesson for WTP studies: that the stated WTP for programmes presented after the first one depends on how much time the interviewee takes to accept and digest the partial nature of the WTP exercise. These results suggest an income effect, though it is small. Further experimental research is needed to disentangle the effects of ordering and income.

Explaining absolute and relative WTP

Compared with the adjusted R squared obtained from WTP studies in the value of life literature (see Kidholm 1992), the model in our paper performs well.

Partial studies of WTP are generally limited in their ability to suggest anything with respect to how much the relative valuation of a particular programme is associated with which background variables. In this paper, the dependent variable, share of total WTP for the three programmes, was introduced as an indicator for the relative preference of each programme compared to the total. This approach provides new insights as compared with the partial models of Table 3. First, it supports the intuitive hypothesis that men have stronger relative preference for hearts and women for hips. Second, the relative preference for hearts and for hips increase with age. As this analysis was designed specifically for this study, it is important to test whether these results can be replicated. If so, this approach could prove fruitful for analysing the reasons behind *relative preferences* for health care programmes.

The open ended reasons for WTP for each programme may explain why people are willing to pay but not the WTP level. For the precoded reasons, the mean values in Table 5 were biased towards the upper end of the scale. However, they indicate some interesting explanations for positive WTP, such that altruistic reasons are more important than selfish for all three programmes. One might doubt that people are truthful in giving such ratings. However, the regression analyses do suggest that people were willing to back up their claimed altruism with higher WTP.

Community focus

Altruism might well have been influenced by a possible "community bias" in this study, which was particularly concerned with describing the programmes as public sector health care. Rather than consistently using the term "willing to pay", we phrased the questions in terms of "willing to contribute". This is a further way in which the design of this study complies with the recommendations of Arrow et al (1993). The intention was to emphasise that we were concerned with the valuation of health care programmes that were meant for the whole community depending on need, and not with private insurance based health care only for those who had paid. As long as health services are publicly financed, we would argue that the correct framing is the one chosen.

Following the suggestion of Gafni (1990) to frame health care programmes in terms of an individual insurance option, concern for others may not be included. The notion of probability of use of services can still be incorporated though, but by describing health care actions in terms of how many people in the community are likely to benefit from them in any period of time. Given the size of the community is known or presented to the respondents, this is equivalent to providing people with a probability, but a *community* rather than an *individual* probability. Perhaps further research is required comparing these two ways of expressing probabilities.

5.3 Comparing valuations with health outcomes

There has recently been increased focus on the issue that people value a health care programme for more than its health-enhancing capabilities (Mooney and Lange 1993). Yet, there is still little evidence and knowledge as to the relative importance of additional characteristics of health care beyond those which are picked up by the health outcome measure. One approach to eliciting their relative importance is to compare valuations from WTP with health outcomes. Assuming such outcomes can be measured in terms of QALYs gained, the implication is that the higher the WTP/QALY in one programme compared with another, the more valued is this programme for its 'non-health benefits' and/or the QALY-outcome produced by the particular programme is more highly valued.

The health outcomes as described on the three cards can be reduced to undiscounted QALYs. The outcome of the helicopter programme was 5 saved lives with a total of 150 life years (=150 QALYs[13]). The health outcome from the heart programme was estimated as 2.5 undiscounted QALYs per person * 80 = 200 QALYs, while the outcome from the hip programme was estimated to be 4.5 undiscounted QALYs per person * 250 = 1,125 QALYs (see the appendix). As a second option, using conservative estimates and halving the quality enhancement part of the QALY gain, the outcome from the heart programme becomes 160 QALYs while that from the hip programme becomes 563 QALYs. Thirdly, discounting the first set of QALY estimates at a rate of 10% gives 47 QALYs for the helicopter, 100 QALYs for hearts and 563 QALYs for hips. Hence the *ordering* of the health effectiveness between the programmes is not sensitive to discounting or low estimates. The mean WTP per QALY gained for the different programmes under these three sets of assumptions can be seen from Table 6.

Table 6: *Mean annual WTP (in NOK) per QALY gained*

	Helicopter ambulance	80 more heart operations	250 more hip operations
mean WTP	316	306	232
QALYs undiscounted	150	200	1,125
mean WTP/QALY	2	1.5	0.2
mean WTP/QALY low estim.	2	1.9	0.4
mean WTP/QALY 10% disc	6.7	3.1	0.4

Although the relatively lower valuation of hips might be influenced by the ordering and the life saving focus of the first two programmes, the first line of Table 6 indicates the preferred ranking of the three programmes when based on WTP. Remembering that the costs of the three programmes were identical (NOK 10 mill.), the second line suggests a *reverse* ranking when based on QALY maximization.

---

[13] The cost-effectiveness study of the helicopter project showed that gains in terms of life years and QALYs were about equal (Kristiansen 1992). Because of this the descriptions focused on healthy life years.

The most striking thing from the comparisons of WTP and QALYs is the much lower valuation in terms of WTP for QALYs gained from hip replacements. This might reflect the attitude that "hip replacements don't save lives", i.e. that QALYs gained from life improvements are valued lower than those gained from life extending or life saving. Further, the higher valuation of heart operations might be influenced by a perception of the effectiveness of the programme existing before the marginal 80 would be added. Also, hearts charities and media have a tendency to exaggerate efficacy. The highest valued QALYs are those from the helicopter ambulance, which might reflect a preference for the 'Rule of Rescue'. Life savings from this programme take place at a very critical risk level. Hence this WTP study lends support to the view that life saving is more valued than the same QALY outcome from life extension and life improvement (see Nord 1993).

Because the hearts and hips programmes were to the benefit of a wider community (Northern Norway with 420,000 inhabitants) than the helicopter programme (Troms county with its 145,000), this might well have influenced the respondents to value the helicopter higher. When adjusting the valuations for these differences in probabilities of own use, the ordering of hearts and helicopter changes, but still life improving QALYs are valued lower than life extending and life saving QALYs.

The type of comparisons as outlined in Table 6 could represent a fruitful approach to weighting the relative value of QALYs gained from various health care programmes. One measurement problem, though, is the evidence that WTP does not increase with the good (see Arrow et al 1993), which implies that it is the *programme per se*, rather than the number of patients treated by the programme, which is important for one's WTP. While this is a general measurement problem for WTP studies, it has particular relevance in the suggested context of using WTP to weight QALYs. We are certainly facing a problem if it is the case that respondents ascribe *diminishing* marginal WTP for additional patients treated, while the comparative QALY yardstick values additional patients at a *constant* marginal value. In other words, if the QALY denominator changes at a different speed from the WTP numerator, one should be cautious to generalise from the WTP/QALY comparisons in Table 6. Results expressed in this paper are therefore tentative. We certainly think that further experimental research on this measurement issue is required.

24

However, while acknowledging the methodological problems, we believe that the WTP approach as applied in this paper represents a constructive way of eliciting peoples' preferences for health gains produced by different programmes.

References:

Acton, J.P. (1973): Evaluating Public Programmes to Save Lives: the Care of Heart Attacks, Rand Corporation, Santa Monica, Report NO R95ORC.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H. (1993): Report of the NOAA Panel of Contingent Valuation, Federal Register, Vol 58 (10) 4601-14, Washington DC.

Berger, M.C., Blomquist, G.C., Kendel, D. and Tolley, E.S. (1987): Valuing changes in health risks: a comparison of alternative measures, Southern Economic Journal, 53: 967-984.

Culyer, A.J. (1971): The nature of the commodity health care and its efficient allocation, Oxford Economic Papers, 23: 189-211.

Donaldson C (1993): Theory and practice of willingness to pay for health care, HERU Discussion Paper 01/93, University of Aberdeen.

Donaldson, C. and Gerard, K. (1993): Economics of Health Care Financing: The Visible Hand, Macmillan, London.

Eddy, D.M. (1991): Oregon's methods: did cost-effectiveness analysis fail?, Journal of the American Medical Association, 266: 2135-2141.

Gafni, A. (1990): Willingness to pay as a measure of benefits: relevant questions in the context of public decision making about health care programs, Medical Care, 29: 1246-1252.

Johansson, P.O. (1987): The Economic Theory and Measurement of Environmental Benefits, Cambridge University Press, Cambridge.

Johannesson, M. (1992): Economic evaluation of lipid lowering - a feasibility test of the contingent valuation approach, Health Policy, 20: 309-320.

Jones Lee, M.W. (1989): The Economics of Safety and Physical Risk, Basil Blackwell, Oxford.

Kidholm, K. (1992): Oplæg til dansk betalingsviljeundersøgelse af værdien af trafiksikkerhet, Paper presented for the Nordic Health Economists' Study Group meeting, Kuopio, Finland 19.-20. August 1992.

Kristiansen, I.S. (1992): Cost effectiveness analysis of the helicopter ambulance service in Troms, Institute of Community Medicine, University of Tromsø, mineo.

Lancaster, K.J. (1966): A new approach to consumer theory, Journal of Political Economy, 74: 132-57.

Loehman, E.T., Berg, S.V., Arroyo, A.A., Hedinger, R.A., Schartz, J.M., Shaw, M.E. et al (1979): Distributional analysis of regional benefits and cost of air quality control, Journal of Environmental Economics and Management, 6: 222-243.

Mitchell, R.C. and Carson, R.T. (1988): Using surveys to value public goods: the contingent valuation method, Resources for the Future, Washington DC.

Mooney, G. (1991): Agency in health care: what's the maximand?, HERU Discussion Paper 10/91, University of Aberdeen.

Mooney, G. and Lange, M. (1993): Ante-natal screening: What contributes to benefit?, Social Science and Medicine, 37: 873-878.

Nord, E. (1993): The trade-off between severity of illness and treatment effect in cost-value analysis of health care, Health Policy 24: 227-238.

Thomson, M.S. (1986): Willingness to pay and accept risks to cure chronic disease, American Journal of Public Health, 76: 392-396.

Torrance, G.W. (1987): Utility approach to measuring health-related quality of life, Journal of Chronic Diseases, 40: 593-600.

26

Appendix

The descriptions of the health care programmes

The descriptions[14] presented on separate cards (each separate description was presented on a single card using larger typescript).

Helicopter ambulance

The helicopter ambulance is a supplement to car ambulances. It carries a doctor and provides quicker transport to the hospital.

The helicopter is primarily a benefit in connection with <u>acute illnesses and accidents</u> for the 80,000 people in Troms living outside Tromsø and Harstad.

The helicopter has roughly 260 missions each year. Studies have revealed that most people who have been flown would have had same probability of survival had a car ambulance been used. Doctors estimate that each year the helicopter will save <u>5 human lives</u>, with a total of 150 life years, that would not be saved if car ambulance had been used.

Heart operations

<u>80 more heart operations</u> can be provided each year in addition to the 300 which are currently done in Northern Norway.

Most heart patients are men aged 50-60 years. They have chest pain and breathe heavily when strained.

The operation will make 75% of the patients completely <u>free from pain</u> with less pain for the rest. Without the operation the patients are expected to live 8-10 years. With the operation they will on average <u>live 1-2 years longer</u>.

The operation mortality risk is 2% (so 1 in 50 people die whilst being operated on).

Hip replacements

<u>250 more hip replacements</u> can be provided each year in addition to the 600 which are currently done in Northern Norway.

Most hip patients are women and on average 70 years old. They have problems in walking and have severely reduced functioning. They have pain when they walk and during nights. Without the operation many will be in need of care.

After an operation the pain will disappear for most patients, and their <u>physical capabilities will be considerably improved</u>. Most patients benefit from the operation the rest of their lives (about 15 years), but the operation in itself is not life extending.

---

[14] In framing the descriptions we wish to thank E. Nord, I.S. Kristiansen and O.H. Førde, and from the University Hospital in Tromsø: Professor K. Rasmussen, Chief Consultants O. I. Solem and J. Thoner.

The payment card[15]

	NOK
Make a ✓ by each amount you	0
are sure you <u>would</u> pay	25
	50
Make a **X** by each amount you	75
are sure you <u>would not</u> pay.	100
	150
Make a ◯ around the amount which is the	200
<u>maximum</u> you would be	250
willing to pay <u>each year</u>	300
	400
	500
	600
	800
	1.000 +

The framing of the most important questions

Q4: How much is the maximum your household would be willing to contribute each year for a helicopter ambulance service?
Bear in mind that the extra taxation would reduce what you have left to spend on yourself after taxes.

Q7: How much is the maximum your household would be willing to contribute each year for the heart operation programme, if this were the <u>only</u> programme to be implemented?

Q10: How much is the maximum your household would be willing to contribute each year for the hip operation programme, if this were the <u>only</u> programme to be implemented?

After 4, 7 and 10:
I will now read some reasons why one could be willing to pay for this service. Using the scale in front of you, would you please tell to which degree the different factors influenced your valuation. Did you emphasize that?:

Q12: On second thoughts, would you change any of your numbers about the ranking or maximum willingness to pay? Remember it is not a question of how much you would be willing to pay in <u>total</u> for the three programmes.

---

[15] In framing the payment options, we acknowledge the helpful suggestions from Mike Jones-Lee.

Estimating QALYs gained from each programme

The health status measurement literature does not provide index numbers of health states which would mirror the health gain of the hearts and hip patients as described on the cards used in this study. Therefore, some data from Torrance (1987) were used and adapted.

For the 80 patients who would benefit from a heart operation, their initial health status index was set at 0.85. After treatment, an index of 1 was set to reflect the health state for the 75% who would be free from pain, and 0.9 for the remaining 25% with less pain. The quality increment is described to last 8-10 years. Using 9 years, each patient's expected gain from the health enhancement part becomes:
$[(1-0.85)*0.75 + (0.9-0.85)*0.25] * 9 = 1.125$ QALYs

Furthermore, an additional life quantity increment is described to last 1-2 years. Using 1.5 years, each patient's expected gain from the life extension part becomes:
$(1*0.75 + 0.9*0.25) * 1.5 = 1.4625$ QALYs

Adding these two parts and adjusting for the described survival rate of 0.98 gives each patient's expected QALY gain from the programme:
$(1.125 + 1.4625)$ QALYs $* 0.98 \approx$ **2.5 QALYs.**

For the 250 patients who would benefit from a hip replacement, their initial health status index was set at 0.7, and after treatment at 1. This gain is described to last for the remaining 15 years of patients' lives. Each patient's expected gain then becomes:
$(1-0.7) * 15 =$ **4.5 QALYs.**

Given that the QALY gains from the two programmes might be overestimated due to low initial health state indices and/or high indices after treatment, a simple sensitivity test is to halve the quality enhancement parts. That would give 2 QALYs per heart patient, and 2.25 QALYs per hip patient. Given the life extension part of a heart operation, a conservative estimate of the health enhancement does not affect the QALY gain very much.

Paper 8
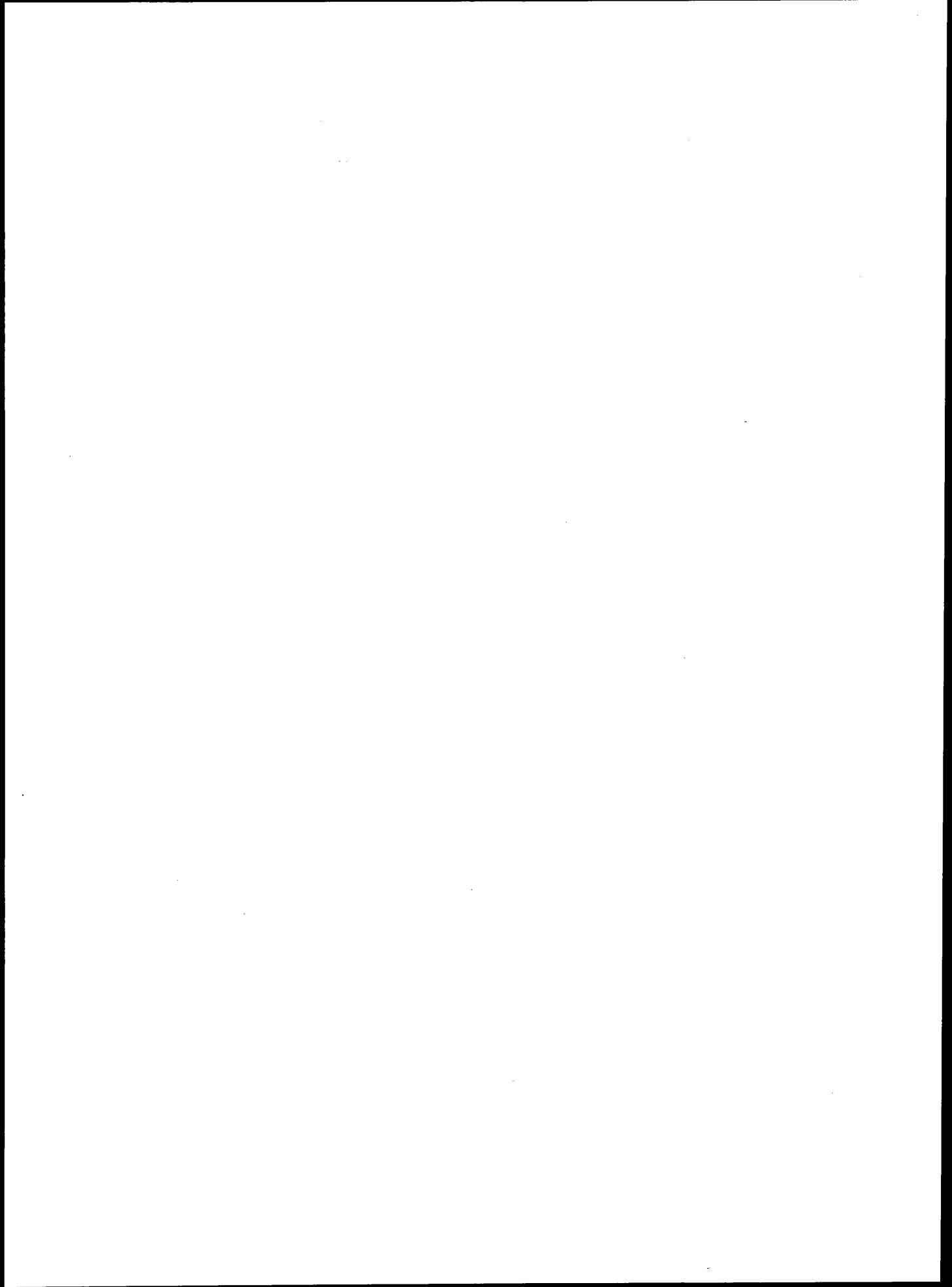
But *health* can still be a necessity ...

Jan Abel Olsen*

University of Tromsø, Department of Economics, Tromsø, Norway

1. Background

In the literature on cross-national comparisons of health care expenditures, some health economists have drawn attention to the issue of whether health care is a luxury good or not (Newhouse, 1977, 1987; Parkin et al., 1987; Gerdtham and Jönsson, 1991; Søgaard, 1991). The substance of the argument will be familiar to readers of this journal. What appears primarily to have evoked concern is the implication of finding Engel-elasticities exceeding unity, simply because of its counter-intuitive labelling of health care as a luxury good.

The point I wish to make in this note is based on the view that consumers demand health, not health care. Theoretically therefore it is the income elasticity of health that should be our concern. It is argued that depending on the magnitude of the observed income elasticity of health care, and the marginal productivity of health care, *health* might still be a necessity even if *health care* is a luxury.

2. The optimal choice between health and consumption

While acknowledging the critique in using the microeconomic concepts 'necessity' and 'luxury' on macrodata, they are, in this note, used as technical definitions related to the magnitude of the income elasticity. Furthermore, a microeconomic approach to the demand for health is applied which makes this note subject to the same criticism when studying macrodata. However, this approach could be applied under the assumption that society's preferences for health vs consumption goods represent an aggregation of individual preferences.

Correspondence to: J.A. Olsen, University of Tromsø, Department of Economics, S. Tollbugt. 7, 9037 Tromsø, Norway.
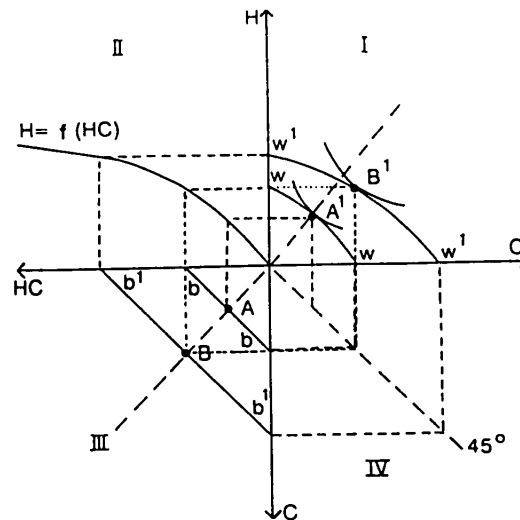
Fig. 1. Constant budget shares for health care and health is a necessity.

A very clear diagrammatical exposition of the household production theory approach to the demand for health is given by Wagstaff (1986) (or see McGuire et al., 1988). Fig. 1 is based on Wagstaff's figure 10. Rather than considering all of the inputs which affect health, quadrant II shows the health production function, $H = f(HC)$, reflecting the narrower view that only health care, HC, affects health, H, (see e.g. Evans, 1984). The production function has the standard properties of positive but decreasing marginal productivity, which are later demonstrated to be of crucial importance in this context. Quadrant III shows the budget constraint; the given income, Y, is to be spent on health care or consumption goods, C. Quadrant I indicates the available combinations of health and consumption goods ('the welfare possibility frontier'), which depends on the financial and health technology constraints and society's indifference curves representing the trade-offs between health and consumption. Note that the optimal choice between health and consumption goods is determined in quadrant I, but that the corresponding optimal choice of *health care* becomes visible in quadrant III.

Consider now the effects of an increase in income. Assume that the initially optimal combination of HC and C, at point A on the budget line b–b in quadrant III, reflects the optimal choice between H and C at point A^1 on the welfare frontier w–w in quadrant I. Taking a special case where the income elasticity of HC is unity, implying constant budget shares, an increase in income causes a movement along a ray from the origin passing through A to point B on the new budget line b^1–b^1. However, due to the shape of the production function, B reflects an optimal choice between H and C in quadrant I; B^1 on w^1–w^1, which lies to the *right of* a ray from the origin
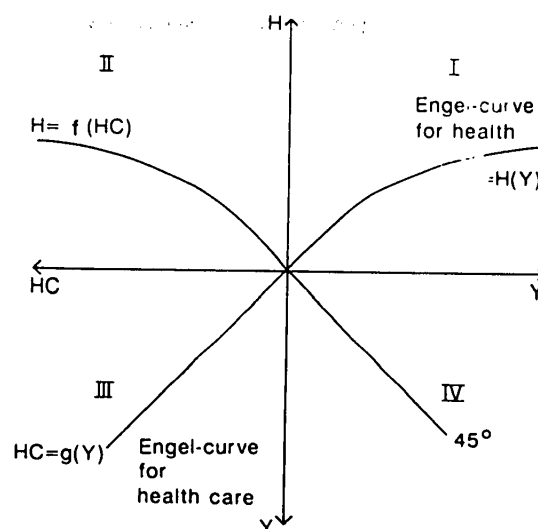
Fig. 2. Engel-curves for health and for health care.

through the initial optimal point A^1. Thus, the relative increase in demand for H is less than that for C, i.e. health is a necessary good. More generally, this implies that, to some extent, the budget share for HC may increase (i.e. HC is a luxury good), but health is still a necessity.

3. Engel-curves and -elasticities

In Fig. 2, C is replaced by Y, so that Engel-curves can be drawn. From an estimated Engel-curve for HC in the quadrant III, one can infer an Engel-curve for H in quadrant I, provided that the health production function in quadrant II can be estimated. In the special case that the income elasticity of HC equals unity (the straight line in quadrant III), then the income elasticity of H equals the output elasticity of health care, i.e. health becomes a necessity.

The generality of the point can be demonstrated by the use of algebra. Consider the health production function (1) and the health care demand function (2):

$$H = f(HC) \quad f' > 0 \quad f'' < 0 \tag{1}$$

$$HC = g(Y) \quad g' > 0. \tag{2}$$

Substituting (2) into (1) gives the health demand function:

$$H(Y) = f[g(Y)]. \tag{3}$$

The income elasticity of demand for health is:

$$E_{H,Y} = H'(Y) \frac{Y}{H} \tag{4}$$

$$= f'(HC) \frac{HC}{H} * g'(Y) \frac{Y}{HC}.$$

Hence, the income elasticity of demand for health is the product of the output elasticty of the health production function and the income elasticity of demand for health care:

$$E_{H,Y} = E_{H,HC} * E_{HC,Y}. \tag{5}$$

The value of $E_{HC,Y}$ is assumed to be given by empirical estimates. The second derivative of the health production function in (1) implies that $E_{H,HC}$ < 1. Thus, if the output elasticity is sufficiently small to make the product of the right hand side of (5) less than unity, we can conclude that although the income elasticity of health care is observed to exceed unity, the income elasticity of health per se may still be below unity. For example, using the mean real GDP elasticity of health care expenditure for OECD-countries in the period 1975–84 (Culyer, 1990), $E_{HC,Y} = 1.3$. Then $E_{H,HC}$ has to exceed 0.77 if health is to be labelled a 'luxury good'.

4. Discussion

Most economists and laypeople appear to think of health and health care as necessities. Yet the evidence that the richer a society becomes, the larger the fraction of its income is allocated to health care, implies that health care is a luxury in technical terms. I suspect that it is the problem of separating this technical definition from the connotation of indulgence, which is attached to the term 'luxury' in everyday speech, that causes disquiet. This note may help them to reconcile their view with observed expenditure patterns; as long as we cannot estimate the shape of the health production function, on the basis of data which suggest that health care is a luxury good, one can never infer that health is a luxury good as well.

The argument in this note could be reversed to explain why health care is a luxury good. Provided that health is a 'normal good' $(E_{H,Y} > 0)$, then, because of the decreasing productivity of health care, the higher the existing expenditure on health care, the higher becomes the increase required to provide a desired increment in health. An additional point[1] which follows

---

[1] Thanks to David Parkin who made me aware of this point.

from the model is that if health is a normal good $(E_{H,Y} > 0)$, then the empirical observation that health care is a normal good $(E_{HC,Y} > 0)$ implies that the marginal productivity of health care on health must be positive as well, i.e. no 'flat-of-the-curve-medicine' or beyond.

References

Culyer, A.J., 1990, Cost containment in Europe, in: Health care systems in transition, OECD Social Policy Studies No. 7.

Evans, R.G., 1984, Strained mercy: the economics of Canadian health care (Butterworths, Toronto).

Gerdtham, U.G. and B. Jönsson, 1991, A note on conversion factor instability in international comparisons of health care expenditures, Journal of Health Economics 10, 227–234.

McGuire, A., J. Henderson and G. Mooney, 1988, The economics of health care (Routledge and Kegan Paul Ltd., London).

Newhouse, J.P., 1977, Medical-care expenditure: a cross-national survey, Journal of Human Resources 12, 115–125.

Newhouse, J.P., 1987, Cross national differences in health spending – what do they mean?, Journal of Health Economics 6, 159–162.

Parkin, D., A. McGuire and B. Yule, 1987, Aggregate health expenditures and national income: is health care a luxury good?, Journal of Health Economics 6, 116–123.

Søgaard, J., 1991, The aggregate income elasticity in health care expenditure, Paper presented to the Nordic HESG meeting, Copenhagen August, 19–21.

Wagstaff, A., 1986, The demand for health: theory and applications, Journal of Epidemiology and Community Health 40, 1–11.
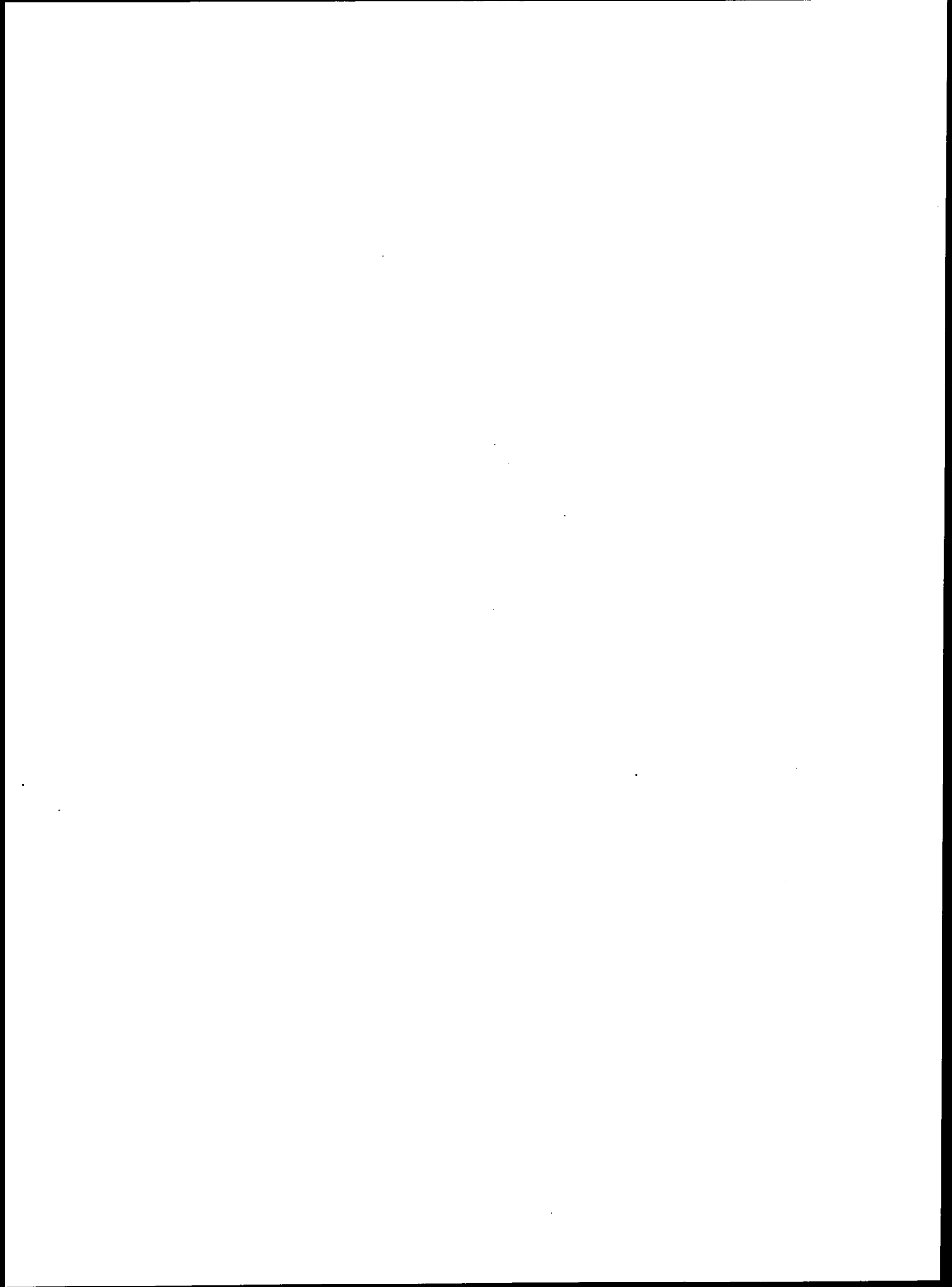
Appendices
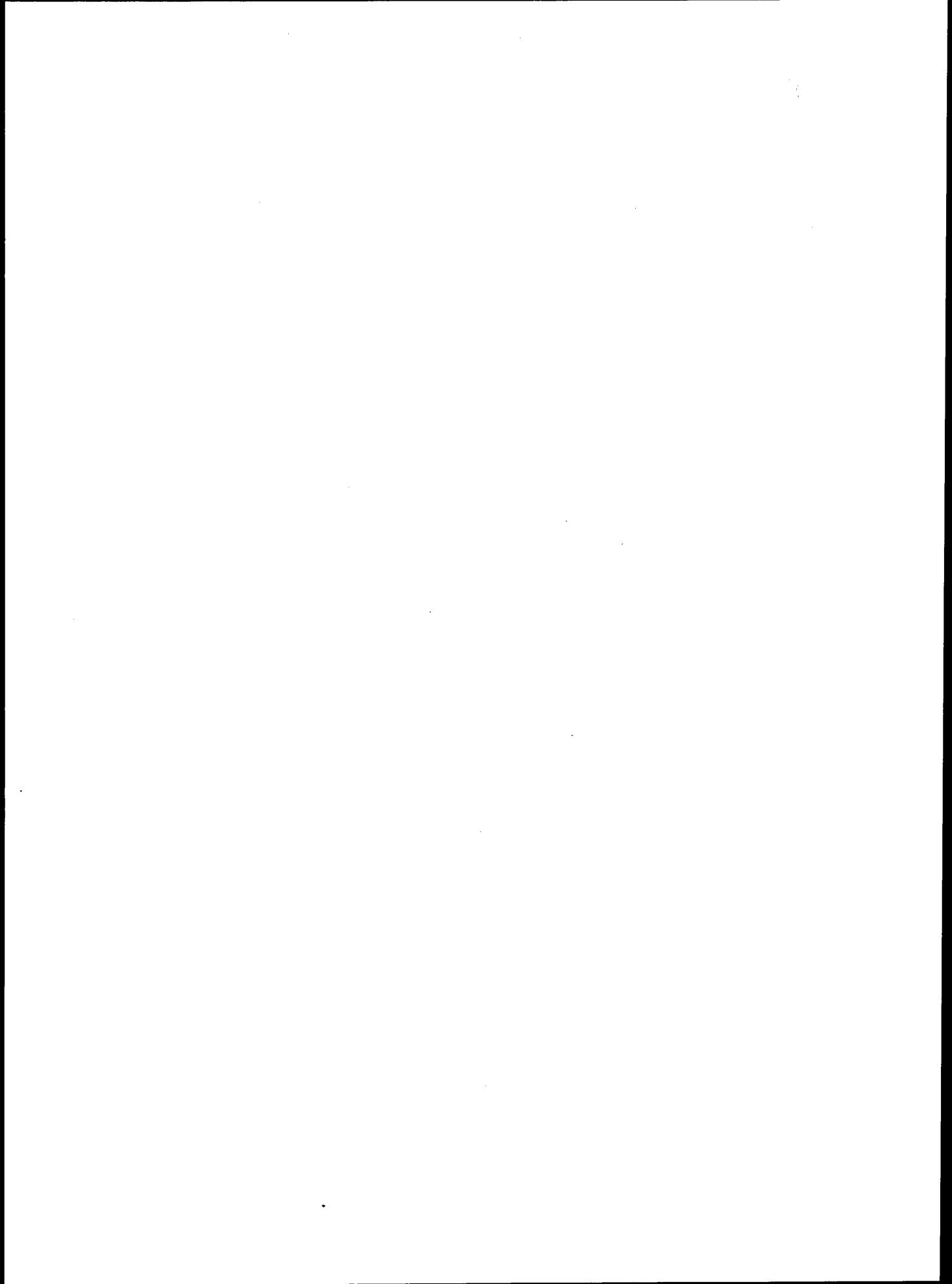
Appendix 1

An English translation of Paper 1

VALUATION OF HUMAN LIFE
By Jan Abel Olsen [1] [2]

This article discusses different methods for valuing human life: human capital, willingness to pay for risk reduction and implicit valuation. We are concerned with how the methods correspond to three alternative criteria for allocating life saving resources, namely equity, utilitarian and consumer sovereignty. First a critical presentation of the literature is given and then some possible causes are indicated for the observed discrepancy in the value of human life in different areas. The article discusses which allocation criteria should be followed and thereby which method(s) is (are) most relevant for use by public decision makers.

1. Introduction

When economists discuss the subject of economic valuation of human life it, not surprisingly, evokes an immediate negative reaction. Even if one values a person, is it really necessary to give this person a price tag? Our noble intentions in dealing with this topic must therefore be clarified. We argue that this apparently perverse topic actually revolves around the allocation of resources in order to reduce one's risk of death. This can happen through explicit prioritization. This stresses the ultimate importance of this topic, inviting economists to dabble in an unusually humane area.

When we talk here about human life, we are concerned with a more statistical concept of life than individual named persons. We are also concerned with reducing the probability of *premature death*. It is not only J. M. Keynes who came to the realization that "in the long run we are all dead".

With the extension of human life as a positive argument in the utility function, there has been increasing interest in the use of life years and "quality adjusted life years" (see, *inter alia*, Williams 1985). We shall not go into this here, and neither do we

---

2

pretend to derive a final value of human life. Instead we deal with the principle of valuing human life.

Two questions form the basis of the problem: how many resources should be used for life saving, and according to which criteria ought these resources be allocated? The question regarding the size of the budget depends on our actual valuation of human life. If the overriding goal of society were to save as many lives as possible, then all resources would be spent on the saving and prolongation of life. When such an allocation does not occur, then this expresses the fact that citizens prefer a shorter life with education, culture and material goods to a long life at the "existential minimum". The size of the "life saving budget" thus reflects a trade off between two types of goods: a low risk of death and other goods. In the utility function, material goods such as tobacco and motorcycles are traded off against life prolonging or death risk reducing goods such as vegetables or jogging shoes. The amount of resources which ought to be used for life saving thus depends upon these trade offs. Naturally, this is the reason why there is no definitive final answer as to the size of the optimal budget.

In principle, there are three criteria for the allocation of resources to life saving. Firstly, the "equity criterion" which states that access to life saving programmes should be *independent of income*. This criterion corresponds to the basic egalitarian ideology behind public health care and also an accepted ethical norm that income should not determine one's risk of death. Secondly we have the "utilitarian criterion" which states that *as many human lives as possible* should be saved. This is a classic economic maximization problem, namely achieving the maximum benefit possible for a given level of resource expenditure. A utilitarian belief - "as much as possible to as many as possible" - follows from this.

The third criterion is that peoples *preferences* for risk reduction in different sectors should determine the allocation between these sectors, "the consumer sovereignty criterion". This criterion builds upon a basic normative law in welfare theory, namely that the consumer always knows best. Consequently, we should take account of the strength of the expressed preferences of consumers for risk reduction in different sectors. If these preferences are measured through willingness to pay, then this will conflict with the equity criterion. If the consumers are disinformed about the objective risks of death and/or have

preferences regarding the *manner of death* they wish to avoid, then this third criterion will conflict with the second.

The purpose of the consumer sovereignty criterion is to realize potential Pareto improvements. It is based on individuals' own trade-offs given the existing income distribution. Thus there is a clear correspondence between the trade-offs of each individual and the principle adopted by society. The equity and utilitarian criteria also deal with the *question of distribution*. They are coherent with the distributional principle which says that allocation should be *independent* of ability to pay.

The aim of this article is to discuss different methods for valuing human life, with explicit reference to the three aforementioned criteria. After a critical presentation of the literature, some causes of the observed differences in the valuation of human life in different areas are briefly discussed. After this, the three allocation criteria are compared and the relationship between consumer sovereignty and the other two criteria is discussed.

2. Methods for valuing human life

There are three main methods: human capital, willingness to pay and implicit valuation. The human capital approach uses the wealth creation of individuals in valuing their lives. Willingness to pay for risk reduction measures how individuals value a reduced probability of death. Implicit valuation measures the value of human life on the basis of how many resources society allocates in practice to schemes whose purpose is to save lives; "life saving schemes".

2.1 Human capital

Human capital refers to individuals' productive capacity which is determined by education, ability and skills. Consequently, productive capacity, and thus the contribution to the economy, varies between individuals.

Up until the beginning of the 1970's, most attempts to value human life were based on the human capital approach (Dublin and Lotka 1946, Reynolds 1956, Weisbrod 1961, Dawson 1967, Rice and Cooper 1967, Hanlon 1969). Using the premise that *gross earnings* show the value of an individual's marginal product, society's economic valuation of human life is measured as the present value of future expected income adjusted for expected productivity growth. As both earnings and the number of remaining years in the

active labour force vary, it follows that young individuals with high earnings have greater human capital than older ones with low earnings.

In the often quoted study of Rice and Cooper (1967), the American population is divided into socio-demographic groups according to age, education, sex and race. White men between 30 and 34 years of age with at least 16 years education had the highest human capital ($ 223 471) The lowest human capital ($ 396) was calculated for a coloured man over 85 years of age. Differences between the sexes deserves mentioning. According to their tables, a new born boy would have almost twice the human capital of a new born girl. Women had lower human capital than men right up until they neared retirement age. However, as the marginal product of housewives was calculated as corresponding to the average wage of a domestic worker, women had a higher human capital than men once they passed the age of 60 because of the service they provide to their pensioned husbands(!):

"As women age and leave the labour force, they are still credited with a dollar value if they stay home and keep house, which prevents their earnings from falling as rapidly as those of men who have left the labour force." (Dorothy Rice and Barbara Cooper 1967)

The clear advantage of the human capital approach is that it is simple. It possess, however, an even more apparent weakness; it does not measure what it should, namely how much we value *a life*, rather it gives the value of being *alive*. Consequently, according to Rhoads (1980), almost all academic economists have abandoned this approach as a method for valuing human life. The method lives on however. The most recent contributors seem to have a more careful attitude to the approach, legitimizing it by giving the results a different interpretation; instead of measuring the value of a human life, human capital measures the *economic costs of illness* (Hodgson and Meiners 1982, Scitovsky 1982).

This explicit change of emphasis towards the costs of illness corresponds to the original intention behind the method. According to Robinson (1986), the reason for economists' use of the human capital method was to show that it was economically prudent to allocate resources to the health sector. It is therefore not surprising that human capital based cost-benefit calculations are popular among personnel in the health sector for "proving" to politicians how wise it is for society to allocate resources to their field. Several contributions in this tradition have originated from authors connected to the

American Health Department (Brody 1975, Cooper and Rice 1976, Hodgson and Meiners 1979, 1982).

The intention is praiseworthy enough. It is also not uncommon to attempt to change a normative statement, "we think more money *should* be allocated to the health service", to a positive one, "human capital calculations show that this *is* prudent". The following quotation can be directly interpreted as suggesting that the human capital method is an unbiased economic calculation:

"If one wants to know what the economic burden of illness was last year, what resources will be saved by preventative measures that reduce the incidence of disease, or what the economic impact of improved survival rates will be, the human capital method provides an appropriate, although partial measure." (Hodgson and Meiners 1982)

By partial, these authors mean that the human capital method does not pretend to measure anything other than strictly economic aspects. It measures the *economic benefit* of making people capable of returning to the productive sector - rather like a calculation regarding the repair/maintenance of a factor of production.

The policy implication is that *if the benefit*, measured by the present value of future income, *is greater than the costs* of saving or treating a group of people, then it is economically prudent to use resources to save the members of this group.

2.1.1 Some reservations

There are two classes of reservations about the human capital approach: methodology and ethics. A methodological reservation regards the use of *gross earnings as a measure of ones human capital* - capital which has economic value because of its productive capacity. The defining characteristic of "human capital" is that it exists within a human being. It is determined by ones ability and skills, and is enhanced through formal education.

A high wage *can* therefore reflect a large amount of human capital, but can also be due to other factors. The wage depends upon ones *relative bargaining power* in the labour market; strong unions achieve higher wages than weak ones. The wage also mirrors *discrimination* which is not connected to differences in productivity; women and men with the same human capital often receive different wages.

Another point is that the wage does not necessarily reflect the human capital which *the worker herself* possesses, but often how much fixed capital the worker is combined

with in the production process. The marginal product of labour will be high in a capital intensive sector which may well lead to higher wages than in more labour intensive sectors. It is therefore meaningless to use such wage differences as an indication of inequality in individuals' productive capacity. (However, to the extent that a high wage is caused by the fact that there is a *limit* to the number of people who can perform a particular job, the wage will reflect a large amount of *human capital* among those who possess the required skills).

The most serious reservations revolve around *ethics*. What in fact does a Rice and Cooper type tabulation *actually* indicate? It gives the economic benefit of treating/saving individuals from different socio-demographic groups. The consequent economic interpretation is that, assuming identical costs per saved life, it is economically most expedient to save those with the highest level of human capital. *It does not take much to go from this intended unbiased calculation to an interpretation as a priority criterion*, where people are ranked in terms of the present value of their future incomes. At the back of the queue we would find women, people with little education and the elderly. As Jack Wiseman has indicated:

"The young (...), the basically fit, and those with the highest expected earnings, would provide the highest rate of return and would therefore be given access to medical resources on the most favourable terms. The old, in contrast, constitute a liability... Indeed, if growth is the sole aim of our policy maker, there might be a strong case for providing only one medical service for those who can no longer work: euthanasia." (Wiseman 1963)

Prioritization in accordance with patients' "economic worth" can be observed in the Norwegian health service. It must be possible to view the introduction of occupational health care as being based on the desire that employees should keep themselves healthy and productive, rather being an intended fringe benefit. The idea of using social insurance funds for increasing treatment capacity also follows human capital reasoning; if the value of the loss of production during the time which the patient is away from work, sick and awaiting treatment, is greater than the cost of treatment then it is economically beneficial to increase the capacity for treatment.

2.2 Willingness to pay for risk reduction

The first attempts in this tradition attempted to calculate *individuals' valuation of their own lives*, V, by using the following simple formula:

$$V = \frac{dv}{dr}$$

(2.1)

where dv is the minimal compensation the individual would require for accepting a change in risk of dr. In other words, by dividing the minimal compensation by the change in risk, a final price of the individual's own life emerges. If one, for example, accepts 20,000 NOK extra for taking a job which involves an increased probability of death of 0.01, the implicit value of ones own life will be 20,000/0.01 = 2 million NOK.

An obvious weakness with this approach was the assumption of *linearity between the probability and willingness to pay* for all values of each. Thus, if an individual faces the "choice" between living and certain death, the method assumes that we can calculate her value of her own life from her earlier willingness to pay for reduced risk! However, a final price does not exist. In such a hypothetical situation the person would demand an infinite level of pecuniary compensation for being deprived of the right to life saving treatment. It is therefore not possible to realize Pareto improvements.

In Thomas Schelling's pioneering article from 1968, the focus was turned from the individual's valuation of her own life to her valuation of a reduced probability of death: "It is not the worth of human life that I shall discuss, but of 'life saving', of preventing death. And it is not a particular death, but a statistical death."

Furthermore, he was concerned with individuals' preferences:
"It is worth while to remind ourselves that the people whose lives may be saved should have something to say about the enterprise."

This perspective is taken up by Mishan (1971) and Jones-Lee (1974, 1976). Mishan's point of departure was that projects with large social benefits could involve the loss of some lives. How should these lives be valued in a cost-benefit analysis? His point was that it is not the lives of known persons which are being "sacrificed", but that

8

all involved are "making a sacrifice" through an increased risk of death. Consequently, it is the increased risk of death which should be compensated:

"The relevant sum to be subtracted from the benefit side are no longer those which compensate each person for their certain death but are those sums which compensate each person in the community for the *additional risk* to which he is to be exposed" (Mishan 1971).

Let us imagine that a scheme to be carried out will have as a consequence that one person from a particular population will die. *No one knows in advance who this will be.* It is assumed that people require compensation for accepting this increased risk. Of course, the amount needed will vary across individuals. If we then sum the compensation required by each member of the population, we will obtain the *value of a human life.* This can be expressed as

$$V = \sum_{i=1}^{N} dv_i$$

(2.2)

where V is the value of a human life and dv_i is the compensation required by individual i. As we have assumed that one individual is sacrificed, the increase in risk is, by definition, given by dr = 1/N. Notice that this value, V, is not derived from a calculation of one particular individual's willingness to pay to avoid certain death as in (2.1). What we have here is the compensation required by a particular population for sacrificing one of its members.

Whilst Mishan is concerned with the minimal compensation necessary for individuals to accept an increased risk of death, Jones-Lee writes about willingness to pay for risk reduction. Even though Mishan considers an increased risk and Jones-Lee a decreased one, the two approaches are alike in principle. Both have a two good - wealth (material goods) and security (low risk of death) - utility function as a basis. The consumer faces a trade off between wealth and security, where increased security implies a reduction in wealth. According to Jones-Lee, the expected utility of an individual is given by:

$$(2.3) \quad E(U) = (1 - p^*)L(W^*) + p^* D(W^*)$$

where p^* is a given probability of death, W^* is a given level of wealth (income), $L(W^*)$ is the utility if one survives whilst $D(W^*)$ is the utility of being dead(!). (Dardis (1980) operates with a simpler version of this utility function by assuming that the decision maker attaches a zero ex ante utility to the state of being dead; in other words, that one does not take into account the utility of leaving bequests).

If one can buy a reduced risk of death of $p < p^*$, then this will require the individual to refrain from some consumption. The maximal amount which one is willing to pay, V, will by definition be that amount which gives the same expected utility as in (2.3):

$$(2.4) \quad (1-p)L(W^* - V) + pD(W^* - V) = (1-p^*)L(W^*) + p^* D(W^*)$$

There are two central connections between p and V: (1) above a certain maximum risk threshold, p', there is no amount of compensation one could be offered to be enticed to accept the increased risk; (2) the maximum one is willing to pay for a given reduction in risk will decrease as the level of risk decreases i.e. one's willingness to pay for a risk reduction from 0.2 to 0.1 is smaller than a reduction from 0.5 to 0.4. Intuitively, it seems reasonable to assume that the higher one's level of risk, the more one would be willing to pay for a given reduction in risk. These characteristics can be illustrated by a convex indifference curve in a diagram with two goods: the probability of surviving, (1 - p), and compensation V.

For any given change in risk, the value of a human life will be given by the *average* of the consumers' marginal rates of substitution between wealth and the probability of surviving, corresponding to V in (2.2). For a more precise description, see Bergstrom (1982) and Dehez and Dreze (1982).

A common element of the central contributions of Schelling, Mishan and Jones-Lee is that all treat the valuation of human life as an ex ante reduction in risk, which implies that we are *dealing with statistical rather than known lives*. Broome attempted to refute this "conventional wisdom" in what has become a much discussed article (Broome 1978). It is worth mentioning as it triggered many responses which contained the necessary specifications.

V

1   p   (1 - p)
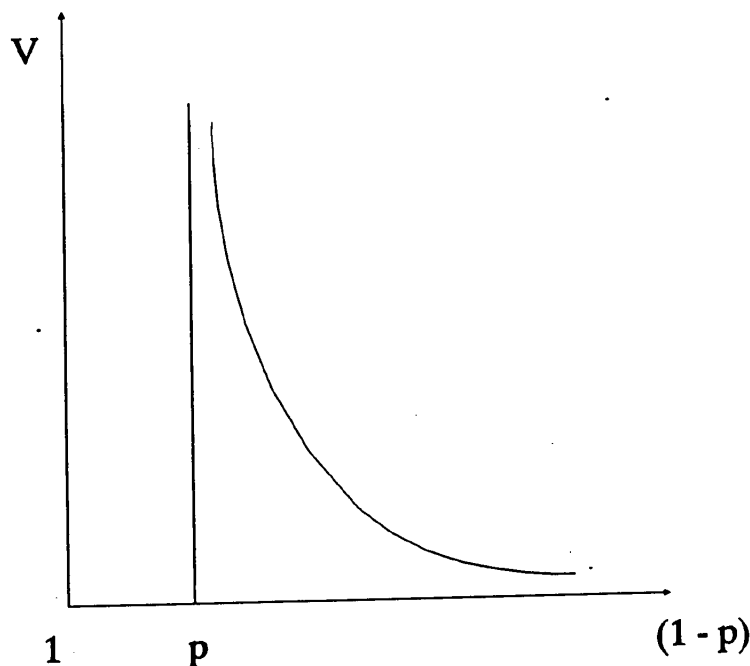
The essence of Broome's criticism is that even if *individuals* are willing to accept risk, a *public decision making authority* cannot value a lost statistical life in a cost-benefit calculation because this person, no matter how 'statistical', will attach an infinite price to his life, which implies that a cost-benefit analysis of a project which involves the loss of at least one life will always be negative. Broome claimed that the reason why the project may be undertaken is that *those making the choices are not fully informed* - they do not know who will die.

The point that lost statistical lives will eventually be lost known lives, which cannot be compensated, confuses the relationship between ex ante and ex post. As indicated by Jones-Lee (1979), the value judgements of a social decision ought to reflect the preferences of the individual *at the time that the decision is taken*. From the information available at this time, the decision makers *are* fully informed.

Broome has also been criticized because he does not present alternative methods or suggestions for improving existing ways of valuing human life. Mishan described him as follows: "He is a knight errant in search of a holy grail. I wish him luck in his wanderings - better luck, at any rate, than he had in his 1978 essay." (Mishan 1981)

McGuire and Mooney (1985) claim firstly that Broome misunderstands the economic *concept of value*. Life has a high user value but no worth in exchange. The risk of losing life can, on the other hand, be exchanged for material goods - wealth which can of course only be enjoyed if one survives. Secondly, Broome's attempt to cast doubt upon ex ante valuations implies that he has forgotten that economics is about *choices* between alternatives with different *expected* outcomes.

Broome seems to have a rather conspiratorial view of cost-benefit analysis as a decision tool for a state seeking to legitimize the sacrifice of some citizens' lives in order that projects can be carried out. Another view - "neutral" or naive - is that the implicit valuation of human life which appears by summing minimal compensations reflects citizens' actual behaviour - a genuine respect for their risk preferences.

2.2.1 Empirical studies

From observed behaviour, we know that individuals are willing to pay for a reduced risk of death, and that they usually require compensation to accept an increased risk of death. Several empirical studies have been carried out to calculate how people value different types of such changes in risk. These studies are of three kinds: "safety-related consumption", questionnaire and compensation for risky jobs.

The first method is concerned with measuring how people deal with risk via their market behaviour. Melinek (1974) studied how pedestrians in London, who ran across the road instead of using the subway, traded off time saving against increased risk. They had a value of £ 87 000 in 1972. Blomquist (1979) examined the use of car safety belts. The "cost" was that it takes eight seconds to fasten a seat belt. With a time cost equal to half of gross earnings, the value of an American driver was calculated to be $ 370 000 in 1978.

A *questionnaire* was used by Jones-Lee (1976) to examine peoples' willingness to pay for increased air safety (which surprisingly turned out to be £ 2.5 million). Dardis (1980) asked people about their willingness to pay for smoke alarms and calculated Americans' valuation of life to be $ 101 000 to $ 137 000 in 1979. The first large questionnaire-based study, based on a representative national sample, was carried out by Jones-Lee et al. (1985), asking about peoples' willingness to pay for transport safety. From the average willingness to pay, a statistical life was valued at £ 1.5 million.

That people must be compensated for undertaking risky jobs was indicated by Adam Smith, who claimed that, between two jobs of otherwise identical conditions, the one with the highest risk must offer the highest wage in order to attract workers. Empirical studies from the USA by Thaler and Rosen (1976) and Viscusi (1978) support this hypothesis. Further, it will be the case that workers with low risk aversion will be over-represented in dangerous jobs, because they are willing to accept such jobs with less compensation than other workers. A study by Robinson (1986b) confirms that comparable workers with dangerous jobs earn more than those with safe jobs. He also noted, however, that dangerous jobs were also worse jobs than safe ones in terms of working conditions, promotion prospects and status. It is the second string in the labour market - those with few alternatives - which take the dangerous jobs, from which it is difficult to "advance".

Not surprisingly, risky white collar jobs offer significantly higher compensation. The study by Marin and Psacharopoulos (1982) finds an implicit value of non-manual workers which is 3.5 times larger than that of manual workers in England.

The willingness to pay for life assurance has been discussed as a method for calculating one's valuation of one's own life; but this does not express the attitude to the risk of losing one's own life, rather it expresses the attitude to the risk of the economic loss which will be suffered by one's family. (However, several criminal cases show that there *can* be a connection between the amount of life assurance and the risk of own death. A large life assurance increases the incentive for a person to kill their boring provider. An optimal life assurance must therefore be based on a trade off between the desire that those left behind should be given economic security and the possibility that this person will prefer a high pension to a living provider.)

2.2.2 Discussion

Even if the willingness to pay method achieved the status as the conventional methodology among academic economists a decade ago (Jones-Lee 1982), it still has several problematic facets: problems of perception and measurement, and the distributional aspect.

The assumption that the consumer always knows what is best for her is problematical in the case of consumption goods which change the risk of death. Consumers usually purchase goods with guaranteed qualities. A guaranteed result cannot be provided

through payment for a reduced *probability* of death. Furthermore, consumers can only make meaningful choices from experience. It is extremely seldom that one experiences that "life-saving goods" (smoke alarms, car safety belts etc.) actually save one's life.

However, the most fundamental problem associated with the willingness to pay method is that people have *great difficulty in understanding its content and in behaving consistently in the face of small risks* (see, inter alia, Hammerton 1982, Thaler 1982). This problem comes to the fore in several empirical studies. An interesting observation can be made in Blomquist (1982) who, from a comparison of several empirical studies, shows how the value of human life increases as risk reduction decreases. One of the works referred to was a questionnaire study of peoples' willingness to pay for a reduced risk of nuclear accidents where the value of human life was calculated to be $ 62 000 for a risk reduction of 10^{-3}, $ 428 000 for a risk reduction of 10^{-4} and $3 576 000 for a risk reduction of 10^{-5}.

These calculations are very interesting because they indicate that the respondents have not understood the difference in size of these risk reductions, *but have taken the changes in risk to be roughly similar*. If the two final values had been calculated using the same risk reduction as the first, namely 10^{-3}, the value of human life would have been $ 62 000, $ 42 800 and $ 35 760 respectively.

The same problem of perception was found among the respondents of Jones-Lee et al. (1985) where almost half expressed the same willingness to pay for different reductions in risk. People not only have problems behaving consistently with respect to risk *reductions* but also to the *level* of risk. Smith and Desvousges (1987) varied the initial level of risk ("baseline risk") between different groups of respondents and found that the valuation of risk reduction was almost constant across large differences in the risk level. This corresponds poorly to the assumption that willingness to pay decreases with the risk level.

The perception problem touches upon the relationship between objective and subjective risk. Objective risk is the true or statistically correct risk, whilst subjective risk refers to that which the consumer experiences and behaves according to. The subjective risk can either be larger than the objective one and thus contain an "anxiety element", or it can be less than the objective risk as the consumer believes that she is less liable than average to the objective risk.

For Mishan, problems of perception and the concept of objective risk seem uninteresting. Since consumers by definition behave according to a subjective risk, it is their expressed willingness to pay for reducing this subjective risk which is important for cost-benefit analysis. In other words, it does not matter whether they pay for reality or an illusion, as long as they actually express a willingness to pay!

The perception problem cannot be rejected on the basis of such a strict welfare-theoretical interpretation. As Mooney (1978) correctly indicates, the use of the subjective perception of risk depends firstly on the consumer *being* best and secondly *believing* she is best at making the necessary evaluations; furthermore, that she *can* and, as importantly, *wants* to do this. If these four requirements are not satisfied, then the assumption that consumers are the correct people to evaluate risk is doubtful.

Another methodological problem with hypothetical questions is that their framing influences the respondent (Tversky and Kahneman 1981, Loomes 1988). Rhoads (1980) refers to a study by Jan Acton of the willingness to pay for a given reduction in the probability of death from heart failure, in which two different questions were posed. The question which emphasized the benefit to the particular respondent gave almost twice the willingness to pay as the question which emphasized the benefit to society!

The willingness to pay for risk reducing schemes has in principle two *distributional aspects* depending on whether the scheme benefits a cross section of the population or a specific income group. If it benefits a cross section, then the total willingness to pay will be larger the greater is income inequality, because risk reduction is usually assumed to be an income elastic good with an Engel elasticity greater than one. A society with unequal income distribution will thus demand more risk reduction than one with equal income distribution. If the scheme benefits a particular income group, the willingness to pay will depend on their *ability* to pay.

2.3 Implicit valuation

The implicit method is based on public resource allocation and calculates society's valuation of human life in different sectors depending on the amount of resources which are actually allocated to projects whose exclusive purpose is to save human lives. Dividing the resource expenditure by the number of lives saved yields the minimum implicit price of human life. If, for example, the authorities choose to spend 10 million

NOK on a project which is expected to save 5 people, then the implicit valuation of human life is *at least* 2 million NOK. If, on the other hand, the project is not carried out "because it is too expensive", then human life is implicitly valued below 2 million NOK. We cannot therefore derive a critical value; only the *minimum value* if the project is introduced and the *maximum value* if it is not.

It is important here to distinguish between known and statistical lives. When known, named people are in danger then *economic* arguments are seldom used for not providing life saving services. Another reason that large amounts of resources are used to save known lives is that *without* this expenditure this particular individual will die *with certainty*. The purpose of other life saving schemes is to reduce the *probability* of death.

Society's valuation of statistical lives can be calculated in several ways, but the main approach studies public resource allocation and legislation. We have a series of laws and regulations, especially in the traffic sector, whose intention is to reduce accidents. This often involves costs, both of the safety measure itself and of the monitoring needed to ensure that the law is followed. In addition there are the direct costs paid by consumers, e.g. safety belts. If we divide the total costs of the scheme by the calculated number of human lives saved, then we reach an implicit price of human life.

A much cited example is the introduction in England of a regulation specifying that tractors had to have closed in cabs (Mooney 1977). The total increase in the tractor costs of English farmers was £ 20 million, but this would save 200 human lives. Thus, the implicit value of a farmer turned out to be £ 100 000. The overview by Graham and Vaupel (1981) of 57 studies of implicit valuation of human life in the USA showed enormous differences both within and between sectors.

An alternative approach is to look at legal judgements surrounding accidents. Society's valuation of human life is viewed here as being reflected by the compensation received by the victim's family. In Great Britain the principle of full compensation ex post applies, which means that compensation is the amount which the surviving family would have had if the provider had lived the rest of his working life. In the USA there is a tradition that the person responsible should be punished, which means that surviving members should have more than full compensation.

It is not just life, but also parts of the body which have an implicit monetary value. The invalidity table of the Norwegian Social Insurance System is interesting reading in

this respect. Here the degree of invalidity is given according to types of disease and loss of limbs. For example, total scalping implies a 15-25% degree of invalidity, whilst the loss of both ears amounts to 25% invalidity. (Van Gogh became 15% invalid when he cut off one of his ears).

2.3.1 Discussion

The human capital and willingness to pay methods are, in principle, cost-benefit analyses with different ways of valuing human life. A human capital based calculation takes peoples production value on the benefit side, whereas the benefit side of a willingness to pay calculation would reflect peoples' willingness to pay for a reduced risk of being among the unlucky ones. The implicit method, on the other hand, does not construct any benefit side at all; it is a cost-effectiveness analysis.

The method has some practical advantages in that it makes visible the consequences of decisions which are made/not made. It thus has important policy implications. The method calculates society's valuation of human life based on the amount of resources which are actually used in various risk reducing schemes. The most important contribution of the implicit method is that it shows society in which sectors it should prioritize resources for life saving schemes and in which sectors life saving schemes should possibly be reduced (given the assumption that the utilitarian criterion holds). Independent of the amount of resources to be allocated to life saving schemes, their distribution should occur such that the marginal costs of saving a life are equal in all sectors.

2.4   Summary

The *human capital method* does not correspond to any of the three allocation criteria. The value of a person's life depends, among other things, on her income and is thus not in accordance with the equity criterion. The method does not pretend to save as many lives as possible using a given budget, but rather it calculates to what extent it is worth saving a person with given characteristics. The method also does not heed individuals' preferences.

The *willingness to pay method* does not meet the equity criterion because willingness to pay depends upon *ability* to pay. The method is not concerned with saving as many lives as possible with a given budget, but it is genuinely concerned with following the third allocation criterion, namely that individuals' preferences should count.

The *implicit method* does not take account of income at all and therefore corresponds to the equity criterion. The method calculates the value of human life within different sectors depending on previous political decisions and will not say anything about future allocation as such. However, the method can be used to calculate the costs per saved life across different projects and thus will show where it is cheapest to save lives. Consequently, the implication of the method corresponds to the criterion of saving as many lives as possible with a given budget. The method does not take account of the preferences of the individual.

3.   Some reasons for differences in the value of human life

The welfare-economic interpretation of the observed differences in the valuation of human life will be that willingness to pay depends as much on the risk *level* as the size of the *change* in risk, and also that the consumers have preferences over *causes* of death they wish to avoid, and that they have different *abilities* to pay for risk reducing products. Let us firstly look at the *distribution of income*. Usher (1985) assumes that safety is a "personal good" with an increasing marginal cost curve for the individual; the larger the degree of risk aversion, the higher the marginal costs. Consequently, given the same degree of risk aversion, the rich will demand more risk reduction than the poor - safe cars are often expensive cars. A high implicit price on life can thus be an expression of the fact that the scheme reduces the risk of death for the rich.

Mooney (1977) refers to a study which found that between 80% and 95% wanted to die suddenly and seemed to want this at any cost. The respondents of Jones-Lee et al. (1985) were asked which cause of death (traffic accidents, heart complaints or cancer) they would most like to be reduced. A huge 76% wanted to reduce the number of deaths from cancer. Thus, we prefer to die quickly and without suffering rather than enduring a slow, painful death. In relation to the consumer sovereignty criterion, it would therefore be rational to use more resources per saved life on those causes of death we would most like to avoid.

There also seems to be an *anxiety element* in handing control of one's life over to others, even though the risk of dying is smaller than when we (think we) are in control ourselves; fear of flying is much more widespread than fear of car travel! The anxiety element is also linked to the dimension of accidents - people seem to have a large degree

of "catastrophe aversion". The willingness to pay for a marginal reduction of the risk of a catastrophe gives a higher value of life than is indicated by the willingness to pay in sectors with a greater probability of small accidents.

While we are conscious of the cause of death and the worry element which reflect *rational preferences*, the perception problem is an expression of the fact that either one does not know that one is in a risky situation, or that one does not manage to relate to the consequences of small changes in risk. To the extent that this is why the subjective risk deviates from the true level, ones behaviour reflects *"irrational preferences"*.

Thus far we have discussed psychological and economic variables - rational preferences, the perception problem and income distribution - which explain the discrepancies which appear from using the willingness to pay method. It may well be useful to seek explanations of inefficient resource allocations by the public sector in political science. There is a greater return for politicians who highlight, through their favourite causes, that they are saving a few sufferers rather than saying "an alternative allocation would have been able to save more people". Another explanation is the media. Headlines such as "He died in heart queue" have undoubtedly contributed to the increase in resources for heart operations.

A third set of explanations deals with *ethics*. This would be able to contribute to explaining firstly a higher valuation of the young vis-a-vis older people, and secondly a higher valuation of known lives versus statistical ones. In a recently published study (Lewis and Charny 1989) based on a random sample of the population in Cardiff, the following question was posed: "Which of two patients would you treat when they are different in age only and you cannot treat both?". When the choice was between a five year old and a seventy year old, 94% chose the five year old, 5% would not answer and only 1% prioritized the seventy year old. Even more noticeably, 70% thought that it was an easy choice to prioritize the five year old. We should note that this prioritization of children over the elderly does *not* have a theoretical basis but reflects a widespread social ethic that childrens' lives are more highly valued than those of the elderly.

As mentioned, a prevailing assumption in the literature on the valuation of life is that we are dealing with statistical lives - not identifiable individuals. Identifiable individuals are favoured by our "social conscience" for prioritizing people who are heading for certain death without treatment over reducing the risk of everyone else dying. This affects

our view of the relative allocation between schemes which cure and those which prevent, leading to a higher willingness to pay for schemes which are designed to cure.

4. Which allocation criteria ought to be followed?

In the introduction we established three allocation criteria to which the discussion of the methods for valuing human life is related. In this section we shall see to what extent these criteria conflict and, if they do, which should be given the greatest weight. There is no conflict between the equity criterion and the utilitarian criterion, because one's income does not determine whether one should be saved. The consumer sovereignty criterion is more problematical, however, as it is on conflict with both of the other criteria.

4.1 Equity and/or consumer sovereignty?

As mentioned earlier, an individual's willingness to pay expresses her ability to pay and her preferences, which vary according to her income and the degree of risk aversion. A high willingness to pay for risk reduction is due to high income and/or a large degree of risk aversion. Low willingness to pay is caused by a correspondingly low income and/or low risk aversion. If, as Mishan suggests, the allocation of public funds to life saving were to take account only of the willingness to pay method, then in reality this means that more money ought to be used to save a rich person than a poor one. As Usher (1985) indicates, such an allocation rule will *increase* the differences in living standards. The study of Marin and Psacharopoulos (1982) illustrates this point as they find that the value of a non-manual worker is 3.5 times larger than that of a manual worker. The cost-benefit implication is that, at identical cost per saved life, it will be economically profitable to prioritize non-manual workers.

With the generally accepted attitude that one's ability to pay should not affect one's access to the health sector, it is difficult to find acceptance for the idea that the ability to pay should count in *other* sectors which deal with life and death. The facet which makes the willingness to pay approach problematic to use is that it is methodologically difficult to *extract* that part of the willingness to pay signals which are due to unequal abilities to pay from those which are due to different preferences.

It is obvious that consumers have *different preferences about risk*. In the literature one distinguishes between "risk lover" and "risk averter". Risk "stimulates" risk lovers,

who are willing to pay in order that they may take on a specified degree of risk. There are no economic arguments against allowing "risk-stimulating consumption" as long as this type of behaviour does not have external effects. It can seem unreasonable, however, that risk averse people will have to pay taxes in order to reduce others' risk of death from this risk-stimulating consumption.

As Dehez and Dreze (1982) indicate, public projects which are intended to increase citizens' probability of survival have the same characteristics as a *collective good*. A collective good may well be limited in supply or linked to a particular locality so that one can only obtain utility from a publicly financed risk reduction through a particular type of consumption or in a specified area of residence. One's probability of survival is also greatly influenced by the consumption of *private goods*.

Mishan (1971) divides risk into *voluntary* and *involuntary* components. Voluntary risk is linked to the consumption of a good with which the consumer is familiar and therefore accepts this by virtue of her consumption. The risk will be accounted for in her demand, for example smoking and mountain climbing. The demand for the good will thus increase if its associated risk can be reduced. Involuntary risk on the other hand is forced upon the consumer.

Let us consider three types of risk, where the first two will be involuntary whilst the third is voluntary. First we have a *collective risk*, r_k, which is lower the greater the amount of resources society allocates to the risk reducing collective good. Such goods can be vaccinations and safety standards in general. Second we have a *local collective risk*, r_l, which varies across regions. Traffic safety and access to health services are such goods with pronounced regional variation. Third we have a voluntary risk which is linked to variations in *individuals' behaviour*, r_p. The sum of these three risks will be one's total risk of death, r. It is reasonable to assume that the three probabilities are *independent* of each other and that their sum will usually be significantly below 1.

$$r_k + r_l + r_p = r$$

This categorization makes two dilemmas visible. The first concerns one's attitude to voluntary risk. The second is concerned with regional variations in risk and is discussed in the next section. For a given budget, more people will be saved the lower is

the value of r_p. One could therefore argue that public resources be used to reduce r_k and r_l whilst resources to reduce r_p should be financed through forced insurance and taxes on risky consumption. For example, smokers pay a tax to finance a health service which cures them.

In summary, there are two main problems with the consumer sovereignty criterion in relation to the equity criterion. The first is that ability to pay is unequally distributed and the second is that the degree of risk aversion varies.

4.2   Utilitarianism and/or consumer sovereignty

The third main problem with the consumer sovereignty criterion has already been mentioned: the *problem of perception*. A paternalistic solution is at hand, namely to transfer the decision making authority to a body which will behave rationally in the face of the objective risks. On this subject, Mooney writes: "Provided - and it is a major proviso - that consumers are prepared to accept that there are certain areas where 'acting in their own interests' it is better to hand over certain decision-making, or in this particular case knowledge on which to base decision-making to the state (since by so doing they are likely to increase their **ex post** utility), then there is nothing to prevent the state in stepping in to correct the misestimates of probabilities by individuals." (Mooney 1977)

In an area in which the equity criterion is so generally accepted, it may be fruitful to take the approach of Usher (1985) and others and adopt Rawls' theory of justice. The point of departure is how distributive rules in a society would be formulated by its future members "behind a veil of ignorance", in other words if they do not know which place in society they themselves will have and have an equal probability of landing in anybody's shoes. The question invites a maximin justice criterion, i.e. one which maximizes the utility of the least favoured member of society.

Linnerooth (1982) interprets Rawls' maximin justice criterion to imply a bias towards those who always have the largest risk of death, in other words known rather than statistical lives. The implication is that more resources will be used to save a known than a statistical life. An allocation in accordance with Rawlsian justice will thus not correspond to a utilitarian allocation.

This interpretation is not convincing. It seems remarkable to assume that the future members of society, who in principle will have the role both of statistical and

known lives, will establish distributive rules which favour known lives at a time at which, by definition, these known lives are statistical. Because of this it is difficult to imagine that the distributive rules for the allocation of resources to life saving and death risk reduction will not be close to the rule which sets equal costs per saved life, as it is this which gives the future society member the longest possible expected life.

Let us imagine that the population recognizes the problems involved in signalling their preferences through a series of partial willingness to pay measures and instead are faced with the following question presented by for example the Social, Transport and Communication, and Municipal Affairs ministers in combination: Do you want us to allocate our total life saving resources so that as many lives as possible can be saved, or do you have an aversion against particular causes of death which means that more money should be allocated to these areas with the consequence that fewer lives can be saved and hence your total risk of death increases?

Again it seems difficult to imagine that the allocation will not be close to one which involves equal costs per saved life, in other words the utilitarian criterion. Notice that there is nothing to prevent one from simultaneously expressing a preference for using more money to save known lives in present danger. Within a given budget, such a priori-tization of known over statistical lives naturally implies that fewer lives can be saved. Weale (1979) argues against this skewed allocation in favour of known lives: "There is nothing right in itself or good on balance about discriminating in favour of known lives in present danger, if greater efficiency can be achieved otherwise" (Weale 1979).

Weale is of course correct if efficiency is defined as an allocation in which it is impossible to save more lives without additional resources. He seems to stretch the point too far however. If our "social conscience", or what Culyer (1976) refers to as a "caring externality", prefers to save known lives in present danger at a higher marginal cost than statistical lives then this is *both* correct in itself and good on balance.

Finally there is a challenge and a dilemma: the *challenge* lies in the timing and level of decision making. The actual allocation to specific programmes has to be undertaken before victims become victims i.e. *ex ante* before statistical lives become known individuals. It is not only society which expresses a lower willingness to pay to rescue statistical lives, but also those who will be victims of a particular illness express a lower willingness to pay for its cure *before* they know that they will be among the

victims. Then their private willingness to pay for a short extension of life will reflect the fact that they have very little else on which to spend their money. Society, however, could have used the same resources to rescue more lives. Linnerooth (1982) makes a good point: "the important point is that from the utilitarian's perspective, in contrast to the welfare economist's, the differential treatment of risk populations cannot be justified by reference to an individual's *own* preferences for reducing high risks to himself, but can only be justified by reference to a desire on the part of others to respond disproportionately toward the more threatening situation." (Linnerooth 1982)

"Timing" has to do with the decision-making level in that "tragic choices" ought to be taken at a high level *before* health service personnel stand face to face with the patient. According to Calabresi and Bobbitt (1978), tragic choices are decisions about who shall receive access to limited life saving resources.

The *dilemma* regards which *type* of equity should matter. So far we have treated equality synonymously with equal cost per saved life independent of income. An alternative is *equality of involuntary risk*. In the previous section we indicated that r_i will vary. To the extent that there exist regional variations in reducing the costs of traffic accidents, for example, - topographical differences which make it more difficult to straighten dangerous bends in the road in some regions - it will be more expensive to save a statistical life in such regions than in others. If the allocation criterion is to achieve the same cost per saved life, then people in these regions will be exposed to a greater risk of death. The utilitarian criterion says that r_i should vary but a risk equity norm requires the same exposure to involuntary risk.

Keeney (1982) is concerned with the fact that there is a social preference for risk spreading, as it seems intuitively more just that risk be spread rather than concentrated. This justice norm or preference for risk spreading reflects the individual's assessment of the degree of risk to which she is exposed herself, namely decreasing marginal utility of successive risk reductions.

In summary, we have argued for the idea that it can be rational for individuals to prefer a paternalistic solution which allocates according to the utilitarian criterion. Still, there is nothing to prevent "moderated paternalism" by taking account of the preferences of the population about the cause of death and their social conscience for prioritizing

24

known lives in present danger. Such preferences can be elicited by methods other than measuring willingness to pay.

5. Summary

The annoying thing with economists is that we insist on making "tragic choices" explicit. Doctors can isolate themselves from tragic decisions by only thinking about the patient in front of him rather than the whole queue. Politicians prefer to avoid making tragic decisions for fear that the next election could be tragic for themselves!

Rhoads (1980) expresses the view that it is demoralizing when society publicly puts a value on human life and argues that the topic and evaluations should be discussed less openly. Is it better from a moral viewpoint to set a value covertly? That a topic is sensitive is not an argument against taking it up:

"The claim that something is inherently complicated does not excuse us from explicitly stating the principles which should be brought to bear upon the decision." (Marin 1983)

The intention of this article has been to discuss the different methods for valuing human life with explicit reference to three criteria for the allocation of a "life-saving budget". The theoretical discussion often seems to overlook the fact that there are explicit distributional goals attached to this budget.

The question itself should not be unethical as long as we are not considering a price on *known lives*. Yet the *human capital method* seems to have dubious ethical implications. The method does not correspond to any of the three allocation criteria which we mentioned in the introduction and should therefore have been dispensed with as a method of calculating the value of life. The method still creeps in, being used opportunistically by health care personnel who are concerned with proving how beneficial their activities are to society.

The *willingness to pay method* calculates the value of human life based on how people actually relate to the risk of losing it. The method is thus in the same vein as the consumer sovereignty criterion. However, willingness to pay reflects also *ability* to pay and thus strongly contradicts the equity criterion. Additionally, this method does not correspond to the utilitarian criterion because it does not pretend to calculate a final or average value of human life which could be used as a basis for budget allocation.

"This is simply because the crucial dv/dr ratio varies, *inter alia*, with the level of risk, r, the magnitude of dr, with the interpretation placed on dr and also, of course, with the specific sort of death envisaged." (Mishan 1985)

As the "value of human life", V, which is obtained from dv/dr will vary from zero to infinity, it follows that *V must be calculated for each individual project.* The implication of Mishan (1985) is that this requires partial cost-benefit analyses of the projects which influence peoples' risk of death, in order to determine to what extent the willingness to pay for saving a life is larger than the costs.

Given the serious methodological problems connected with measuring consumers' true preferences, such measurements would have to be comprehensive indeed. I seriously doubt that the economic benefits of such studies outweigh the total costs! This is one of the reasons why I have a more pragmatic view of the question of using an average value of human life. This will correspond to both the equity and utilitarian criteria.

The *implicit method* is not really a method for the valuation of human life but rather a method which calculates the *costs per saved life* in different sectors. The method shows where it is cheap and where it is expensive to save an extra life. It can therefore indicate possible improvements if the aim is to save as many as possible within a given budget. Its obvious advantage lies in the policy plane in that it "will allow decision-makers to be more consistent in their decision-making and at the same time more efficient" (Mooney 1986). Can we require much more?

We can require that the method should also say something about the size of the budget, but this method does not. However, if society, on ethical grounds, does *not* wish to derive an average value of its citizens (a reasonable assumption?), then the indications from the implicit method will be a very good alternative.
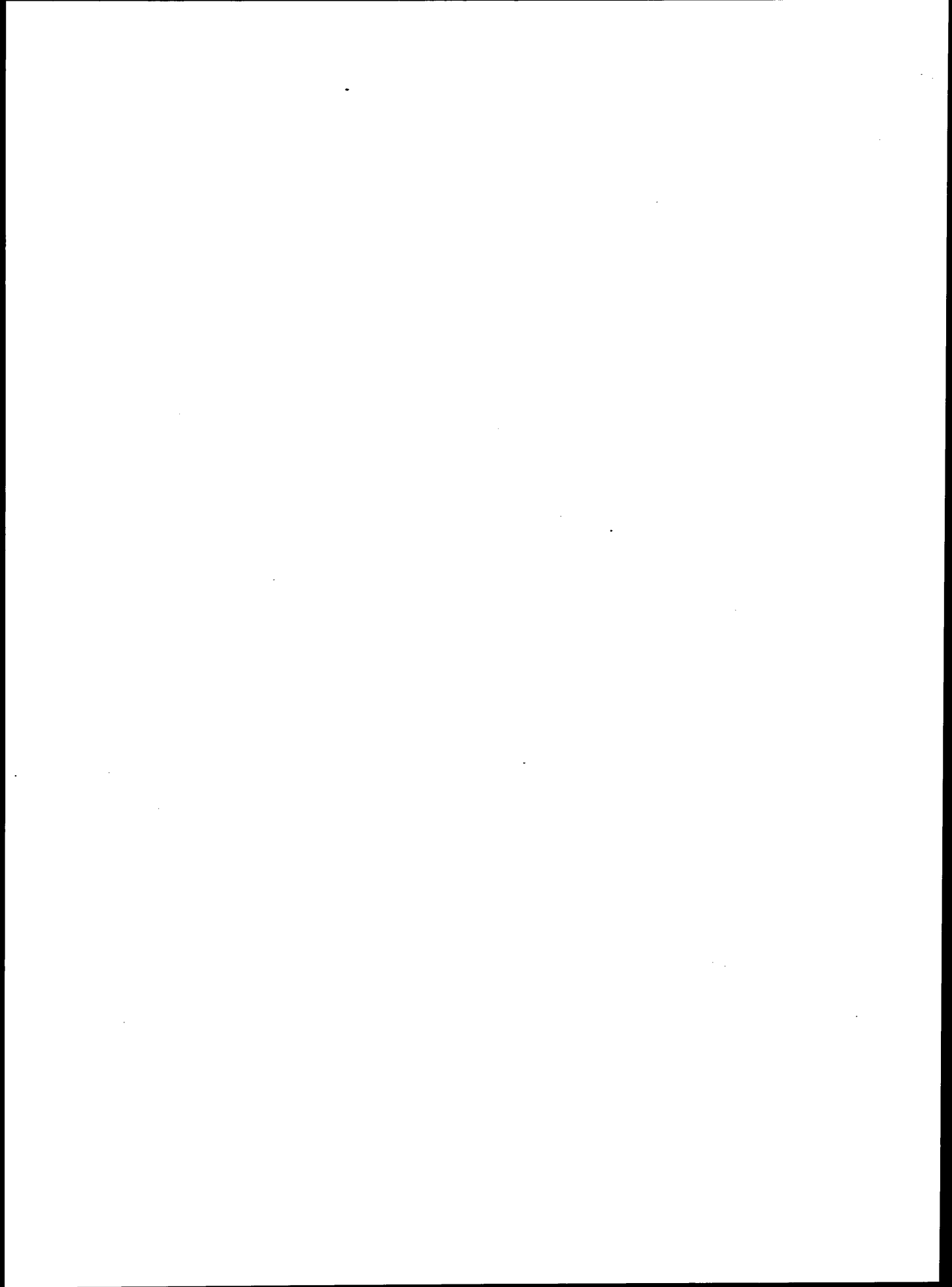
Appendix 2

The questionnaire used for Paper 5

(self administered)

Jan Abel Olsen, UiTø
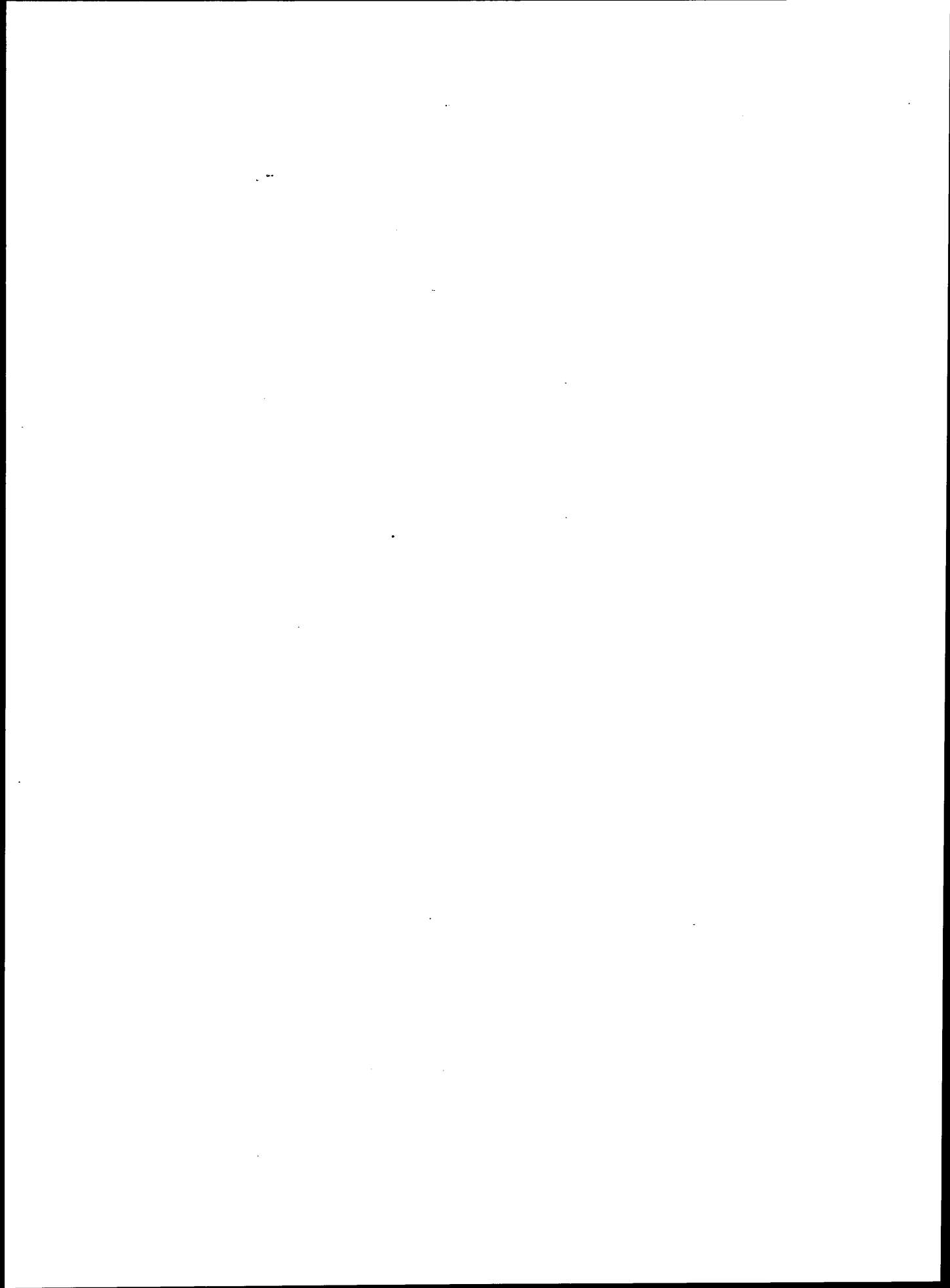
Attitudes to health benefits over time

There is a wide range of health care programmes which can save
lives or improve health. But the health service is faced with
constrained resources and as a consequence cannot implement
all programmes. Different health care programmes have to be
"set against each other" in the competition for resources, in
order to find those which are to be given priority.

The difference between health care programmes is often a
question of <u>when</u> the outcomes arise. In such cases we shall
have to compare health benefits at different points in time.
Although there is uncertainty related to the outcome of long
term programmes, normally one would operate with an expected
value. The purpose of this study is to elicit how <u>you</u> think
society should make trade-offs between health gains which
occur at different points in time, when making priorities
between different programmes.

The questionnaire is in two parts, with a total of four choice
situations. In each situation there are two alternative
programmes. Assume that all programmes that are to be assessed
<u>cost the same</u>, and that all costs occur in the <u>present</u> budget
period. You can therefore disregard cost variations.

You are asked to imagine yourself in the role of a decision
maker in the health service, where you are to decide what
makes the two programmes equally good. There are no right or
wrong answers, but try to think through the problems before
you answer.

Part I

The health service is being allocated additional expenditure which is to be spent on a programme which <u>saves human lives</u>. Assume now that the only differences between the programmes are <u>when</u> human lives are saved. (The programmes save lives of the same age. Thus, the same people cannot be saved at different points in time. We do not know who will be saved.)

Choice situation 1
- life saving in 1 year or in 5 years:

<u>Programme A</u> will save 1,000 lives <u>in 1 year</u>.

<u>Programme B</u> will save 1,000 lives <u>in 5 years</u>.

Which of the two programmes would you choose? (Please tick)

Programme A

Programme B

A and B are equally good

IF YOU CHOSE A:

Imagine that <u>more than 1,000 lives</u> could be saved in 5 years. Roughly how many lives do you think Programme B would have to save in order for <u>A and B</u> to be considered <u>equally good</u> ?

Answer (please complete): _____ LIVES

[If you find it difficult to give a finite number, would you please indicate the range within which the number of lives saved by Programme B <u>in 5 years</u> would have to lie for <u>A and B</u> to be considered <u>equally good</u>:

BETWEEN _____ AND _____ LIVES]

Choice situation 2

- life saving in 1 year or in 20 years:

Programme A will save 1,000 lives in 1 year.

Programme C will save 1,000 lives in 20 years.

Which of the two programmes would you choose? (Please tick)

Programme A

Programme C

A and C are equally good

IF YOU CHOSE A:

Imagine that more than 1,000 lives could be saved in 20 years.
Roughly how many lives do you think Programme C would have to
save in order for A and C to be considered equally good ?

Answer (please complete): _____ LIVES

[If you find it difficult to give a finite number, would you
please indicate the range within which the number of lives
saved by Programme C in 20 years would have to lie for A and C
to be considered equally good

BETWEEN _____ AND _____ LIVES]

Part II

The health service is then being allocated additional expenditure which is to be spent on a programme which <u>improve the health</u> of people in a chronic state of dysfunction and distress. Assume now that the only differences between the programmes are <u>when</u> the health improvements take place.
(The programmes improve the health of people of the same age. Thus, we are not talking about alternative timing for improvement of the same peoples health. We do not know who these people are.)

Choice situation 3
- health improvements in 1 year or in 5 years:

<u>Programme X</u> provides full health for 1,000 people <u>in 1 year</u>.

<u>Programme Y</u> provides full health for 1,000 people <u>in 5 years</u>.

Which of the two programmes would you choose? (fill in)

Programme X

Programme Y

X and Y are equally good

IF YOU CHOSE X:
Imagine that <u>more than 1,000</u> people can have their health improved in 5 years. Roughly how many peoples' health do you think Programme Y has to improve in order for <u>X and Y</u> to be considered <u>equally good</u> ?

Answer (fill in): _____ PEOPLE

[If you find it difficult to give a finite number, would you please indicate a range in which the number of people that Programme Y must make healthy <u>in 5 years</u> has to lie for <u>X and Y</u> to be considered <u>equally good</u>:

BETWEEN _____ AND _____ PEOPLE]

Choice situation 4

- health improvements in 1 year or in 20 years:

Programme X provides full health for 1,000 people <u>in 1 year</u>.

Programme Z provides full health for 1,000 people <u>in 20 years</u>.

Which of the two programmes would you choose? (fill in)

Programme X	
Programme Z	
X and Z are equally good	

IF YOU CHOSE X:

Imagine that <u>more than 1,000</u> people can have their health improved in 20 years. Roughly how many peoples' health do you think Programme Z has to improve in order for <u>X and Z</u> to be considered <u>equally good</u> ?

Answer (fill in): _____ PEOPLE

[If you find it difficult to give a finite number, would you please indicate a range in which the number of people that Programme Z must make healthy <u>in 20 years</u> has to lie for <u>X and Z</u> to be considered <u>equally good</u>:

BETWEEN _____ AND _____ PEOPLE]

So that we can analyse the responses we get from different people, would you please give some few informations about yourself. Your answers will be confidential.

Age: ☐ years

Sex (fill in):

Female ☐

Male ☐

What is your highest education (fill in):

O level . . . . ☐

vocational training . . . . ☐

A level . . . . ☐

1-2 years further education. ☐

university . . . . ☐

Thank you very much!

Appendix 3

The questionnaire used for Paper 6

(self administered)

Priority setting in health care:

how should we trade off numbers of people against the number of years people benefit after treatment?

When comparing different health care programmes, higher priority would be given to the programme which attains most of something. The two most commonly used measures for this are:
1) the number of persons who are treated; and
2) the total number of years of improved health for all patients.
Years are considered to be a better measure than persons, because they take account of how long time people benefit from improved health. On the other hand, it is argued that we are not indifferent with respect to how the total number of years of improved health are "distributed" across patient groups (e.g. many would prefer 2 persons to have good health for 10 years each rather than one person have good health for 20 years, even if the total number of years is the same).

On the back of this sheet you are asked to compare one health care programme with another. The programmes yield the same health improvement for each individual, but the durations differ. Imagine a chronic state of dysfunction and periodical pain. Treatment does not affect the length of life, but the health status is much improved.

You are asked to imagine yourself in the role of a health planner, where you are to decide what makes the two programmes equally good to implement. Assume that the programmes cost the same and that the patient groups are similar in all respects except for the duration of the benefit and the number of persons treated.

There are no right or wrong answers. Different people hold different views. It is these differing views that this study is examining.

Circle the number of persons, for each of whom health has to
be improved for 20 years for you to consider it **equally good
as** improving the health of 100 persons each for 5 years.
[If you find it hard to decide on one number, you may circle
around two succeeding numbers.]

100 persons get improved health for **5** years each	=	**25** persons get improved health **20** years
		30 "
		40 "
		50 "
		70 "
		90 "
		100 "

Guidance:

*If you think that the total number of years of improved health
is the only thing which should matter, you will choose the
number 25. This implies that you would **not** take account of how
the health gains are distributed.*
*If you think that the number of persons treated is the only
thing which should matter, you will choose the number 100.
This implies that you would **not** take account of the length of
time people benefit from health improvements.*
*If you think that both factors should matter, you will choose
a number in between.*

Thank you very much!

Finally, about yourself:
Year of birth: _____ Sex: please circle   M   or   F

Appendix 4

The questionnaire used for Paper 7

(face to face interview)

Willingness to pay
for health care programmes

I Introductory information

The purpose of this study is to try to get better knowledge about people's **valuation** of different types of health care and how they would prioritise between them. This is important to know so that the health service can put its money into programmes which the community values most highly.

Our way of measuring your valuation of health care programmes is by asking you how much you would be willing to **contribute** through increased **taxation which would be earmarked** to new health care programmes. For existing health care programmes, it is a matter of how much you would be willing to contribute to prevent the service closing down.

Although this study is in a sense **experimental**, it is nevertheless a fact that it is our taxes which pay for the health service. So your willingness to contribute to a particular health care programme will in this study be used as the measure for your valuation of it. There are no right or wrong answers.

TAKE A BREAK AND REPEAT SLOWLY IF NECESSARY

We shall talk about three different health care programmes: helicopter ambulance, heart operations and hip operations. During the interview you will get some cards with more detailed descriptions of these services.

The first service we will talk about is the helicopter ambulance service. At the Regional Hospital in Tromsø a helicopter ambulance has been in operation since 1987. It is meant to **supplement** car ambulance and doctor boats. The service is mainly to the benefit for people living outside Tromsø and Harstad.

1: Before I contacted you, did you know about the helicopter ambulance service at the Regional Hospital in Tromsø?

Answer: No ☐

Yes ☐   **How did you know?**

A) Read or heard about it

B) Read or heard about actual rescues

C) Know someone who has used the service

D) other: _____

II Willingness to pay for the helicopter

I would now like to ask you how much you **value** this helicopter ambulance service. First of all, would you like to read this? I will read it through with you.

GIVE THE YELLOW **CARD A** AND **READ** IT THROUGH WITH HIM/HER. Take time to digest all of the information.

Try to imagine that the money for the helicopter were to be transferred to **other** health services. It would then be of interest to measure how much people could have been willing to contribute to **prevent** it being closed down. As long as there are no plans for closing down the helicopter service, this question might sound somewhat hypothetical. But our purpose is to measure people's valuation of this service.

2: Would you be willing to pay anything in extra taxation for a helicopter ambulance service?

Answer: Yes ☐   GO TO Q4

No ☐   GO TO Q3

Don't know ☐   Does that imply that it wouldn't mean anything to you if it were closed down? PRESS FOR a Yes or No in Q2

3: Would you explain why you are not willing to pay anything in extra taxation for the helicopter ambulance service?
NOTE DOWN THEIR RESPONSE

PROBE:

                                               Yes   No

A) Because you don't believe in
its effectivenes

B) Because you don't like the service
(noise, risky, etc)

C) Don't want to pay more taxes   Yes, GO TO Q3.1

If **No** on 3C, GO TO PART III

3.1: Would you be willing to give a
voluntary donation to the service?   Yes   GO TO Q3.3

                                                 No   GO TO Q3.2

3.2: Would you be willing to pay for a
private insurance premium, which gave
the same service in case you or your   Yes   GO TO Q3.3
family needed it?

                                                 No   GO TO PART III

3.3: In this study we are more interested in
your valuation of the service than how you   Yes   GO TO Q4.
prefer to pay for it. Would you go along with
the interview if thinking in terms of your   No   GO TO PART III
preferred way of paying?

4: How much is the maximum your household would be willing to contribute each year for a helicopter ambulance service?
Bear in mind that the extra taxation would reduce what you
have left to spend on yourself after taxes.

GIVE THE WHITE HELICOPTER CARD AND LET HER/HIM CROSS OUT

4

CHECK, **so your household would be willing to pay maximum NOK**
_____ **per year for the helicopter ambulance service in Troms**
(or the sum lies within the range _____ - _____ **NOK) ?**

5: Could you please explain the reasons why you would be
willing to pay for the helicopter ambulance service?
GIVE THE RESPONDENT TIME TO THINK IT THROUGH, AND NOTE DOWN
THEIR RESPONSE WITHOUT PROMPTING.

THEN, GIVE THE CARD WITH THE SCALE ON DEGREE OF IMPORTANCE.

I will now read some reasons why one could be willing to pay
for this service. Using the scale in front of you, would you
please tell to which degree the different factors influenced
your valuation:

	Degree of importance			
Did you emphasize that:	much	some	littl	no
A) you/your family could need it sometime				
B) It is reassuring to know that a helicopter exists				
C) Other people in the community may need it				
D) It contributes to a geographically more equal access to health services				
E) The service already exists, and you do not want health services to close down				
F) other _____				

_____

III Willingness to pay for two other programmes

The helicopter ambulance service costs 10 mill. NOK per year. Of course there are other health care programmes on which this money could be spent. We shall consider two alternative programmes, each of which costs the same as the helicopter service. These programmes differ from the helicopter service in that they are for people in Northern Norway and not only for Troms.

Hearts

The **first** alternative is increasing **heart** operations.
GIVE THE YELLOW **CARD B** AND **READ** IT THROUGH WITH HIM/HER.

Take time to digest all of the information.

Take a look at the card and compare it with that for the helicopter ambulance service.
6: Which of the two would you consider to be most important, that is; which should be prioritised first?

Answer: First priority _____

{IF THE RESPONDENT HAD TAX AVERSION FOR THE HELICOPTER, 6b:
How would you prefer to pay for the increased number of heart operations?
Answer: Voluntary donation ____
        Private insurance ____
        Taxation ____ }

7: How much is the maximum your household would be willing to contribute each year for the heart operation programme, if this were the only programme to be implemented?

GIVE THE WHITE HEART CARD AND LET HIM/HER CROSS OUT

CHECK, so your household would be willing to pay maximum NOK
_____ per year for the increased number of heart operations in
Northern Norway (or the sum lies within the range
_____ - _____ NOK)?

**8: Could you please explain the reasons why you would be
willing (or would not be willing) to pay for more heart
operations?**
GIVE THE RESPONDENT TIME TO THINK IT THROUGH, AND NOTE DOWN
THEIR RESPONSE WITHOUT PROMPTING.

(IF NOT WILLING TO PAY, GO TO NEXT PAGE)
Would you please find the blue card with the scale on degree
of importance? I will now read some reasons why one could be
willing to contribute to this programme. Using the scale in
front of you, would you please say to which degree the
different factors influenced your valuation:

	Degree of importance			
	much	some	littl	no
Did you emphasize that: A) you/your family could need it sometime				
B) Other people in the community may need it				
C) Because there are "heart-queues" and people are waiting to be treated				
D) Because some patients will be able to return to work				
E) other _____				

_____

TAKE A BREAK TO SIGNAL A NEW SUBJECT

Hips

The **second alternative** is increasing **hip** operations.
GIVE THE YELLOW **CARD C** AND **READ** IT THROUGH WITH HIM/HER

Take time to digest all of the information.

Take a look at the card and compare it with the other two
yellow cards.
**9: How would you rank the hip operation programme compared
with the other two; that is, how would you prioritise them?**

Answer: First priority   _____

        Second priority _____

        Third priority _____

{IF THE RESPONDENT HAD TAX AVERSION FOR THE HELICOPTER, 9b:
**How would you prefer to pay for the increased number of heart
operations?**
Answer: Voluntary donation   ____

        Private insurance   ____

        Taxation   ____ }

**10: How much is the maximum your household would be willing to
contribute each year for the hip operation programme, if this
were the only programme to be implemented?**

GIVE THE WHITE HIP CARD AND LET THE RESPONDENT CROSS OUT.

CHECK, **so your household would be willing to pay maximum NOK
_____ per year for the increased number of hip operations in
Northern Norway, (or the sum lies within the range
_____ - _____ NOK)?**

8

11: Could you please explain the reasons why you would be willing (or would _not_ be willing) to pay for increased hip operations?

GIVE THE RESPONDENT TIME TO THINK IT THROUGH, AND NOTE DOWN THEIR RESPONSE WITHOUT PROMPTING.

(IF NOT WILLING TO PAY, GO TO NEXT PAGE)
Would you find the blue card with the scale on degree of importance? I will now read some reasons why one could be willing to contribute to this programme. Using the scale in front of you, would you please say to which degree the different factors influenced your valuation:

	Degree of importance			
Did you emphasize that:	much	some	littl	no
A) you/your family could need it sometime				
B) Other people in the community may need it				
C) Because elderly can manage themselves and will need less help from others				
D) Because elderly patients "deserve" better health				
E) other _____				

_____

Let me see. You have made the following ranking and valuations.
FILL IN AND READ THROUGH

Q12: Second thoughts

	Ranking	WTP figures		Ranking	WTP figures
First	_____	_____		_____	_____
second	_____	_____		_____	_____
third	_____	_____		_____	_____

12: On second thoughts, would you change any of your numbers about the ranking or maximum willingness to pay? Remember it is not a question of how much you would be willing to pay in total for the three programmes.

Answer:

Change ranking: GO TO THE THIRD COLUMN ABOVE

Change valuations: GO TO THE FOURTH COLUMN ABOVE

IF THE RANKING AND WTP-FIGURES DO NOT CORRESPOND, POINT OUT THE INCONSISTENCY AND ASK: **Why?**

13: In your ranking and valuation of the programmes, were you concerned with comparing the different numbers of people benefiting?

Answer: Yes ____ GO TO NEXT PAGE AND Q14

　　　　No ____ **What was then important to you?**

14: I will now ask you to make one pairwise comparison.

IF **HEART** PREFERRED TO HIP:
You have ranked the heart programme before the hip programme.
That could be understood as meaning that you think the hip
programme does not treat enough people compared with the
number treated from the heart programme. Is that correct?
**14a: Now compare these two cards again, roughly how many
people would you require to go through the hip programme in
order to make the two programmes equally good?**

Answer: _____

IF **HIP** PREFERRED TO HEART:
You have ranked the hip programme before the heart programme.
That could be understood as meaning that you think the heart
programme does not treat enough people compared with the hip
programme. Is that correct?
**14b: Now compare these two cards again, roughly how many
people would you require to go through the heart programme in
order to make the two programmes equally good?**

Answer: _____

IV Demographic variables

So that we can analyse the responses we get from different
people, I would now like to ask you a few questions about
yourself and your household. **We don't ask your name, so it
will not be possible to find out what individual persons have
answered.**

15: What is your highest level of education?

O level

Vocational training

A level

College/university degree

Health care education

Other _____

16: What is your main activity?

In paid employment

Housework

Unemployed

Student

Retired/pensioner

Other _____

**17: I'd like to know roughly what the total income of your
household was last year. (Pensions, unemployment benefits and
study loans count as income). Please have a look at this card.
Could you please tell me the letter which corresponds with the
range in which your household's income lies?**

12

MAKE A CIRCLE AROUND ONE LETTER

under 50.000	A
50.000 - 100.000	B
100.000 - 150.000	C
150.000 - 200.000	D
200.000 - 250.000	E
250.000 - 300.000	F
300.000 - 350.000	G
350.000 - 400.000	H
400.000 - 450.000	I
450.000 - 500.000	J
500.000 or more	K

18: Are you the person in the household that would earn the largest share of its total income?

Answer: Yes _____

No _____

About equal _____

19: Respondent's sex Male _____   Female _____

20: In which year were you born? Answer: <u>19</u>_____

21: Taking your age into account, do you think your health status is: good/very good _____

fair _____

or poor? _____

22: How many persons are there in your household?

Answer: _____

23: How many of these are children under 16?

Answer: _____

Thank you very much!

24: Did you enjoy the interview?

Yes: _____   No: _____

Why:

25: Did you find it difficult to answer?

Yes: _____   No: _____

Why:

26: Do you think such questions on priority setting in the health service are so difficult that they should be left for experts in the health service to decide?

Yes: _____   No: _____

Why:

Interviewers own additional information/comments:

How long did the interview last? _____ minutes
Was the respondent influenced by other people hanging around?
Yes ___   No ___

Other comments on how the interview went:

CARD **A**

Helicopter ambulance

The helicopter ambulance is a supplement to car ambulances. It carries a doctor and provides quicker transport to the hospital.

The helicopter is primarily a benefit in connection with <u>acute illnesses and accidents</u> for the 80,000 people in Troms living outside Tromsø and Harstad.

The helicopter has roughly 260 missions each year. Studies have revealed that most people who have been flown would have had same probability of survival had a car ambulance been used. Doctors estimate that each year the helicopter will save <u>5 human lives</u>, with a total of 150 life years, that would not be saved if car ambulance had been used.

The <u>helicopter ambulance</u> is a supplement to car
ambulances. It carries a doctor and provides quicker
transport to the hospital.

The helicopter is primarily a benefit in connection
with <u>acute illnesses and accidents</u> for the 80,000
people in Troms living outside Tromsø and Harstad.

The helicopter has roughly 260 missions each year.
Studies have revealed that most people who have been
flown would have had same probability of survival
had a car ambulance been used. Doctors estimate that
each year the helicopter will save <u>5 human lives</u>,
with a total of 150 life years, that would not be
saved if car ambulance had been used.

	NOK
Put a ✓ by each amount you	0
are sure you <u>would</u> pay	25
	50
Put a **X** by each amount you	75
are sure you <u>would not</u> pay.	100
	150
Put a ◯ around the amount which is the	200
<u>maximum</u> you would be	250
willing to pay <u>each year</u>	300
	400
	500
	600
	800
	1.000 +

CARD **B**

Heart operations

<u>80 more heart operations</u> can be provided each year in addition to the 300 which are currently done in Northern Norway.

Most heart patients are men aged 50-60 years. They have chest pain and breathe heavily when strained.

The operation will make 75% of the patients completely <u>free from pain</u> with less pain for the rest. Without the operation the patients are expected to live 8-10 years. With the operation they will on average <u>live 1-2 years longer</u>.

The operation mortality risk is 2% (so 1 in 50 people die whilst being operated on).

Heart operations

<u>80 more heart operations</u> can be provided each year in addition to the 300 which are currently done in Northern Norway.

Most heart patients are men aged 50-60 years. They have chest pain and breathe heavily when strained.

The operation will make 75% of the patients completely <u>free from pain</u> with less pain for the rest. Without the operation the patients are expected to live 8-10 years. With the operation they will on average <u>live 1-2 years longer</u>.

The operation mortality risk is 2% (so 1 in 50 people die whilst being operated on).

	NOK
Put a ✓ by each amount you	0
are sure you <u>would</u> pay	25
	50
Put a **X** by each amount you	75
are sure you <u>would not</u> pay.	100
	150
Put a ◯ around the amount which is the	200
<u>maximum</u> you would be	250
willing to pay <u>each year</u>	300
	400
	500
	600
	800
	1.000 +

CARD **C**

Hip replacements

<u>250 more hip replacements</u> can be provided each year in addition to the 600 which are currently done in Northern Norway.

Most hip patients are women and on average 70 years old. They have problems in walking and have severely reduced functioning. They have pain when they walk and during nights. Without the operation many will be in need of care.

After an operation the pain will disappear for most patients, and their <u>physical capabilities will be considerably improved</u>. Most patients benefit from the operation the rest of their lives (about 15 years), but the operation in itself is not life extending.

Hip replacements

250 more hip replacements can be provided each year
in addition to the 600 which are currently done in
Northern Norway.

Most hip patients are women and on average 70 years
old. They have problems in walking and have severely
reduced functioning. They have pain when they walk
and during nights. Without the operation many will
be in need of care.

After an operation the pain will disappear for most
patients, and their physical capabilities will be
considerably improved. Most patients benefit from
the operation the rest of their lives (about 15
years), but the operation in itself is not life
extending.

	NOK
Put a ✓ by each amount you	0
are sure you would pay	25
	50
Put a X by each amount you	75
are sure you would not pay.	100
	150
Put a () around the amount which is the	200
maximum you would be	250
willing to pay each year	300
	400
	500
	600
	800
	1.000 +

Degree of importance:

Much   some   little   No

|————————|————————|————————|
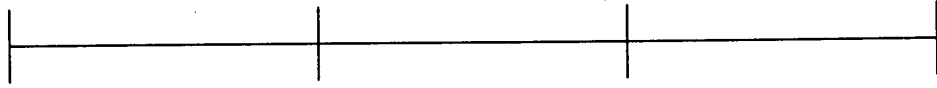
The household's total gross income last year:

less than 50,000	A
50,000 - 100,000	B
100,000 - 150,000	C
150,000 - 200,000	D
200,000 - 250,000	E
250,000 - 300,000	F
300,000 - 350,000	G
350,000 - 400,000	H
400,000 - 450,000	I
450,000 - 500,000	J
500,000 or more	K