Production Gains from Health Care: What Should be Included in Cost-effectiveness Analyses?

Professor Jan Abel Olsen
Institute of Community Medicine, University of Tromsø, Norway

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

(Published in: Social Science & Medicine, vol 49, 1999)
CENTRE PROFILE

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

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

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

PUBLICATIONS

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

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

The Co-ordinator  
Centre for Health Program Evaluation  
PO Box 477  
West Heidelberg  Vic 3081, Australia  

Telephone + 61 3 9496 4433/4434  
Facsimile + 61 3 9496 4424  
E-mail CHPE@BusEco.monash.edu.au
The Health Economics Unit of the CHPE receives core funding from the National Health and Medical Research Council and Monash University.

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

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

The research described in this paper is made possible through the support of these bodies.
Recent literature has been concerned with the correct measurement of the ‘indirect costs and benefits’ of health care as well as the issue of including these items in economic evaluations. This article considers the question of which ‘indirect benefits’ to include in CEA/CUA. Within the context of a collectively financed health scheme the relevant issues include not only the size of the net resource costs of providing health care but also which costs and benefits the society is prepared to consider in its assessment of health services. The strong preference for ‘equal access for equal need’ implies that some production gains may have to be disregarded in the social welfare function. We introduce the notion of socially relevant and socially irrelevant production gains. The analysis suggests that the magnitude of the socially relevant part of the production gains may vary between countries as it depends, first, upon differences in patients’ potential contributions to the rest of society (tax rates), and second, the strength of preferences for equity.
# TABLE OF CONTENTS

1 Introduction

2 Measuring Production Gains

3 Should Production Gains be Subtracted?
   3.1 Is There Double Counting?
   3.2 Can Externalities Account for Social Values?

4 How Much Production Gains Should be Subtracted?
   4.1 Potentially Relevant Production Gains
   4.2 An Equity-Efficiency Trade-Off
   4.3 Which Compensation Principle Should be Applied?

5 Concluding Remarks

References

List of Tables:

Table 1: Concepts and Terminology
Table 2: 

List of Figures:

Figure 1:
1 Introduction

The appropriate method for including the effects of health and health care on the level of society’s production has been a recurring theme in the recent literature. The saving of a productive person’s life or the curing of a person who returns to productive activities would, in principle, involve benefits to society which go beyond the positive health effects. Similarly, the production lost whilst undergoing treatment or participating in prevention programmes involves costs to society beyond the costs of health care. The best method for including such benefits and costs in economic evaluations remains unresolved (Drummond et al, 1993).

This paper seeks to make two contributions. First, it attempts to clarify the concepts and terminology. Section 2 deals with this issue and with the question of correctly estimating the magnitude of the production gains. However, it is important to distinguish between this positive issue, vis what is the magnitude of the production gain, and the normative issue of what should or should not be taken into account in cost-effectiveness analyses of health care programmes. The second aim of this paper is to analyse this latter, normative issue. We argue that technical analysis alone cannot answer this question as the answer depends upon social values, and an equity-efficiency trade-off, which may vary from country to country.
2 Measuring Production Gains

There is a lack of conceptual clarity in the literature. This is partly due to misleading definitions and partly to the different perspectives in cost of illness (COI) studies as opposed to cost effectiveness analysis (CEA) and its variant cost utility analysis (CUA). In COI the focus is the illness, and the aim is to estimate the total cost to society of the illness - sometimes referred to as the economic ‘burden of disease’ (see eg Henke and Behrens, 1986; Murray and Lopez, 1996). In CEA/CUA the focus of attention is not the illness, but the intervention. There is a terminological convention in this literature to refer to direct vs indirect costs and benefits, of which the indirect benefits are the production gains caused by the intervention.

Note that the concepts used in these two sets of analyses have different meanings; ‘indirect costs’ in COI are ‘indirect benefits’ (averted production losses) in CEA. There are other ambiguities as different authors use different terminology. Some authors seem to apply a terminology from COI in the context of CEA/CUA. In Table 1 we attempt to clarify these concepts.

The distinction between ‘direct’ and ‘indirect’ is not descriptively accurate, as it depends on the institutional viewpoint of the analyses. Direct costs include all sorts of health care costs; HC. Direct benefits are health effects or health gains, which in CEA/CUA are non-monetarized as eg lives saved, life years or QALYs (quality adjusted life years). These benefits will be referred to as health; H. Indirect costs are the non-health care costs which accompany the treatment. Often these are measured as different types of time costs, the largest item of which is the production lost whilst undergoing treatment. These costs will be called production losses; PL. Indirect benefits are, in the words used in the seminal paper by Torrance (1986) “the production gains to society because more people are well, or alive, and able to return to work”. We understand by production gains; PG, the value of the increased output which is attributable to the treatment. As one cannot add incommensurable units, the monetarized PG have to appear in the numerator of a CEA/CUA. The health care costs would be adjusted by the effects on the value of production, making what Torrance (1986) called ‘net economic costs to society’. The resulting cost-effectiveness ratio is: [(HC + (PL − PG)) / H].

In this literature there are essentially two different methods for measuring PG; termed human capital and friction cost. With gross earnings as a proxy for the value of one’s output, the present value of the future stream of earnings becomes one way of measuring the production gain to society of a person’s return to work. This estimation of production gains, as averted lost earnings, is termed the human capital approach.

In economies with high unemployment, sick employees can eventually be replaced from the unemployment pool. The (avoided) lost output in this case would depend upon the reduced productivity of each of the people affected and the length of time before productivity returned to normal. By contrast with the simple human capital approach, in these circumstances the lost output to society will be temporary, ie it is limited to a friction period. This is the underlying idea behind the friction cost method (Koopmanschap et al, 1992, 1993, 1995), which represents a technical solution responding to the exaggerated estimates of lost output when measured by the human capital technique.
Table 1: Concepts and Terminology

<table>
<thead>
<tr>
<th>COI</th>
<th>Existing Terminology</th>
<th>CEA/CUA</th>
<th>Suggested Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>direct costs</strong></td>
<td>=</td>
<td><strong>direct costs</strong></td>
<td>= health care costs, <strong>HC</strong></td>
</tr>
<tr>
<td><strong>indirect costs</strong></td>
<td>=</td>
<td>production losses, <strong>PL</strong></td>
<td></td>
</tr>
<tr>
<td>production losses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>productivity costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>COSTS</strong></td>
<td>=</td>
<td>production gains, <strong>PG</strong></td>
<td></td>
</tr>
<tr>
<td><strong>EFFECTS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>direct benefits</strong></td>
<td>=</td>
<td>health,</td>
<td></td>
</tr>
<tr>
<td>health outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(non-monetarized)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>indirect costs</strong></td>
<td>=</td>
<td>averted losses</td>
<td></td>
</tr>
<tr>
<td>production losses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>productivity costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>COSTS</strong></td>
<td>=</td>
<td>production losses, <strong>PL</strong></td>
<td></td>
</tr>
<tr>
<td><strong>EFFECTS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>indirect benefits</strong></td>
<td>=</td>
<td>production gains, <strong>PG</strong></td>
<td></td>
</tr>
<tr>
<td>averted losses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>production gains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>productivity costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(COI: production losses, <strong>PL</strong>)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 Should Production Gains Be Subtracted?

Given the magnitude of the correctly estimated production gains, whichever label there might be on the methodology used for estimating these gains, should all of it count as PG in the CEA-formula above? On this issue, Koopmanschap and colleagues seem to agree with Johannesson and Karlsson (1997) that all PG should count. The reason offered by the former group is that: "production losses due to illness influence the scarcity of resources and therefore the wealth of society. In this respect there is no difference in assessing direct and indirect costs; both should be incorporated in economic evaluations of health care programmes" (Koopmanschap et al, 1995). In another paper, an opportunity cost argument is used: “the additional resources becoming available (...) could be added to the health care budget (or added to budgets for education, housing and working conditions, which may also contribute to health status), and produce QALYs.” (Koopmanschap and Rutten, 1994).

The Washington Panel, which is the third group to have been involved in the recent discussion, argued that: '[o]nly the impact on the rest of society, not including the deceased, belongs in the numerator.' (Weinstein et al, 1997). While the Panel argued for the inclusion of the part of the production gains experienced by the patient in the denominator as part of the quality of life measurement, it regarded the inclusion of such gains in the numerator as an ‘alternative, equally valid, formulation’. This is not true. When PG is included in the numerator, its importance reflects the economists’ convention of treating all dollars equally. The large PG arising from the cure of a productive worker would then count more than the smaller PG arising from the cure of a less productive worker. However, within the CUA framework, each person’s utility is constrained between the same numerical values, so a gained healthy life year would count the same no matter how productive the patient is. In other words; income mediated effects on quality of life (in the denominator) are normalised on a utility-scale with fixed end points [0 - 1], while the income mediated effects (in the numerator) would be expressed in unweighted monetary values.

The reason why some authors would, in the words of Koopmanschap and Rutten (1993), have ‘philosophical objections’ to including PG arises from the distributional consequences of including PG. Programmes which enable productive people to return to work will have higher priority, while programmes which improve the health of people who do not contribute to society’s aggregate production will accordingly have lower priority. Some of this effect could be corrected by using an implicit wage rate for unpaid productive persons and imputing a positive value for lost leisure time. However, even if this broader approach were chosen, it remains the case that some groups are more productive than others.

A recurrent distributional criterion in many countries’ public health services is that a health gain counts the same no matter who you are; people with identical needs have the same claim - or right - to health care resources. The inescapable problem is to what extent this right to health care should increase with the value of a person’s production. Consider the situation where production gains are manifested in terms of consumption goods (C) only and where individuals consume all of their own income, Y. The production gains from curing a patient are consumed by that patient (PG = ΔY = ΔC). If production gains are taken into account, the ‘net social costs’ are (HC –
△C). In other words, the greater the personal consumption as a consequence of a cure, the more collective health care resources the remainder of society should be prepared to pay for her/his cure. Hence, the crucial matter in the context of a social valuation of production gains is to whom and in which form the gains occur. Do they end up as champagne and caviar for the cured patient, or as increased health care for others?

3.1 Is There Double Counting?

The main reservation of the Washington Panel (Gold et al, 1996) for not including production gains was the danger of double counting benefits, both in terms of health and consumption benefits. The crucial matter regarding this issue is whether or not all of the welfare gains associated with the monetarized production gains are already - or can be - fully incorporated in the non-monetarized health outcome unit. In most situations this will not happen: First, in CEAs (as opposed to CUAs) output is measured in physical units (eg fractures avoided, tumors detected), and these do not reflect the strength of preferences. Hence, there is no reference to the size of the PG in the health outcome unit used. Second, when the preference based unit of outcome has not included a dimension which explicitly describes the income changes associated with the health gain, we cannot assume that respondents have included all utility increments from increased income when expressing their health utility. This is particularly the case in economies where people have extensive sickness insurance. Third, it is rarely the case that all utility from the increased value of production is experienced by the person producing it in isolation. Other people beyond those experiencing the health effects, may benefit from the increased value of production. This can be in the form of consumption externalities, or simply in the form of net contributions in terms of taxes or donations. This third point was recently acknowledged by the Washington Panel (Weinstein et al, 1997).

The double counting argument could be considered within the framework of a simple utility function, in which utility is derived from health, h, and private consumption, c:

\[ U_i = u_i(h_i, c_i) \]  

Some minimum level of health is required to enjoy one’s consumption, as is some minimum level of consumption for maintaining this health. This minimum level of consumption could be interpreted as a subsistence level, s , per period which is needed to support life over the long run. In order to avoid double counting, this part could be subtracted from ci . To avoid the simple deletion of a utility yielding attribute, we have added it as a natural part of hi . The utility function could then be reformulated as:

\[ U_i = u_i[(h_i+s), (c_i−s)] \]

The utility from the increased consumption made available by the increased income following cure goes beyond the utility from the improved health per se. Hence, as long as individuals gain utility from consumption, and not only from consumption via health, it cannot involve double counting to include the consumption benefits.
Within the context of a Paretian social welfare function, which defines society’s welfare as a function of its members’ individual utilities, the utility an individual gain from her increased consumption leads to increased social welfare:

\[ \text{SWF} = w(U_1, U_2, \ldots, U_i, \ldots, U_n) \]  

(3.3)

From (3.1) - (3.3) and with \( \text{PG}_i = \Delta Y_i = \Delta C_i : \Delta C_i \uparrow \Rightarrow \Delta U_i \uparrow \Rightarrow \text{SW} \uparrow. \)

It follows from this SWF that when comparing two individuals with identical preferences and health, but with different consumption levels, society’s welfare would be higher by saving the life of the one whose consumption is highest. It also follows from the autonomous utility functions that the utility of the rest of society is unaffected by the saving of a member \( i \), because they do not care for her health, nor are they affected by her consumption. Hence, it appears that the simple utility and social welfare functions in (3.1) - (3.3) do not offer a description or prescription which correspond with social values.

### 3.2 Can Externalities Account for Social Values?

An extended utility function may be envisaged which includes three arguments in addition to the standard arguments of own private consumption and own health. First, there are public goods, \( G \). Second, it includes the health of one’s fellow citizens. This follows from the hypothesis that there is a caring externality for others’ health as well as an externality through reduced contagion (Culyer 1971, Evans 1984). It is assumed that this preference relates to the aggregate health of the others in society, \( \sum h_j (j \neq i) \). Furthermore, it is assumed that health is produced by health care; \( h = h(h_c) \). The final argument rests on a contention that an individual cares for the consumption opportunities of her fellow citizen \( j \) if this is under some minimum value which we will refer to as a safety net, \( c_0 \), \( (cj \leq c_0) \). While there are individual variations in the preferred level of this net, the point is that citizens care about each others’ consumption - only up to a limit. Note that when a person is able to support her own consumption at a level above the safety net, the remainder of the society is indifferent about the level of this consumption. In reality, there might be positive consumption externalities for the relatively poor and negative consumption externalities for the relatively wealthy; ie people might be egalitarian and/or envious of the wealthy. For simplicity, suppose that society has agreed on a socially acceptable minimum consumption level, \( c_0 \), which is the same for sickness benefits, unemployment benefit, and pensions. Interestingly, the justification for extending the utility function with these three arguments is simply the assumption that individuals support three basic elements of the welfare state; 1) there are public goods, 2) we care for other people’s health, and 3) every citizen should be secured by a safety net:

\[ U_i = u_i[c_i, h_i, G, \sum h_j (j \neq i), c_i | c_j \leq c_0] \]  

(3.4)

From (3.4) it follows that if the initial consumption level of member \( j \) exceeded \( c_0 \) and with \( \text{PG}_j = \Delta Y_j = \Delta C_j \), the utility of the rest of society would be unaffected by the production gain. However, the improved health which was required to bring about \( j \)'s increased consumption is valued by the rest of society through the caring externality for health per se \((\delta U_i / \delta h_j)\). Hence, they are not ignorant of, or indifferent to, their
fellow citizens’ well-being - and presence! - as was the case in the autonomous utility functions of (3.1) - (3.2). Also, with the extended utility function (3.4), the part of the production gains which ends up as own consumption above the level of the safety net is irrelevant to the rest of society. The parts of PG which are relevant are those which affect $G$ and $h$.

While the utility function described in (3.4) overcomes one problem inherent in (3.1)-(3.2), namely that individuals are not concerned with other people’s utility, it does not overcome the second problem discussed above. This is that an increase in PG that is entirely consumed by the patient will increase the patient’s utility and therefore social welfare as defined by (3.3). However, as noted above, including all of an individual’s consumption leads to the unacceptable conclusion that there should be priority treatment for those with the highest own consumption. This is an inevitable consequence of maximizing a SWF of the form given in (3.3) and can only be overcome if the SWF is modified to exclude some parts of the patients’ future consumption.
4 How Much Production Gains Should be Subtracted?

There is nothing new in disregarding some utility relevant gains in a social context. Even textbooks in welfare economics acknowledge the existence of merit bads! More importantly, it is for society to decide which gains are relevant in the context of the particular analysis. This is not to suggest that one institution (e.g., the health sector or the treasury) should ignore gains which are external to their budgets. Rather, society may decide to disregard some gains, because their inclusion would have unacceptable distributional consequences.

It is argued below that the economic impact on the rest of society represents the basis - or starting point - for judging which gains society might find ethically acceptable to include in different situations. These gains will be referred to as Potentially relevant Production Gains; PPG. This approach is similar to the Washington Panel's view expressed above that “only the impact on the rest of society (...) belongs in the numerator.” It differs from the Panel's views as the inclusion of some or all of the impact on the rest of the society may be ethically unacceptable.

4.1 Potentially Relevant Production Gains

From the self-interested perspective of other members of society the economic flows between the remainder of society and patients is complicated by the existence of transfer payments (e.g., sick pay). Hence, the analysis of PPG and transfer payments becomes inextricably intertwined. Assume that income earners have an income, $y$, which exceeds the safety net, $c_0$, and that a flat income tax rate, $t$, is imposed on $y > c_0$, i.e., tax equals: $t(y - c_0)$. The tax finances public goods, $G$, as well as transfer payments. Furthermore, assume that health care, $HC$, is financed by an earmarked income tax on $y > c_0$: $tH(y - c_0)$. Treatment may affect two types of flows between the patient and the rest of society; first, changes in the net-contribution from the patient (beyond own consumption) which go to finance public goods, health care and transfer payments to those out of work, and, second, changes in the transfer payments from the rest of society. Table 2 illustrates the economic impact under four different circumstances.

Table 2: Impacts on the Rest of Society of Curing a Patient

<table>
<thead>
<tr>
<th></th>
<th>net-contributor (returns to work)</th>
<th>net-recipient (remains out of work)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life-saving</td>
<td>$(t + tH)(y - c_0)$</td>
<td>$- c_0$</td>
</tr>
<tr>
<td>Health improving</td>
<td>$(t + tH)(y - c_0)$</td>
<td>$0$</td>
</tr>
<tr>
<td>(reduced morbidity)</td>
<td>$+ c_0$</td>
<td></td>
</tr>
</tbody>
</table>
After being saved, the person returns to work, yielding a PG = Δyj. The rest of society benefit from what \( j \) contributes in terms of taxes. These gains may justifiably be subtracted in the numerator of the CEA-formula. Hence, in this case, there is no difference between the impact on the rest of society and the PPG.

(2) Life-saving and net-recipient: PPG = 0

The idea of considering production gains to the rest of society becomes ethically difficult when analysing net-recipients. It is clear that mortality removes individuals both as producers and consumers. Saving the lives of people who do not contribute to society as producers (e.g. the disabled, pensioners) will have a PPG = 0. Yet their consumption is equal to their transfer payments set here at \( c = c_0 \). Hence, their net contribution \( (PG - c = -c_0) \), becomes negative. In other words, the costs to the rest of society of saving the life of someone who would remain out of work is the sum of the treatment costs and the present value of her/his transfer payments. However, the inclusion of ‘negative benefits’ does not correspond with widely held ethical values (Richardson, 1991).

In this case, there is certainly a discrepancy between the economic impact on the rest of society, and what may justifiably be subtracted in the numerator of the CEA-formula. The only acceptable solution to this ‘dilemma’ is to disregard the higher costs to the rest of society of saving the life of someone who would remain a net recipient, i.e. to set PPG = 0 as a response to an ethical constraint (explicit value judgment) in the analysis.

(3) Health improvements and net-contributor: PPG = \((t+tH)^* (y_j - c_0)\) + \(c_0\)

This refers to the transition from ‘sick and unproductive before cure to fit and productive after cure’. Without treatment, people would receive \( c_0 \), after treatment they pay for their own living by returning to work. [If PG < c_0 due to part time work, all gains would offset c_0: PPG = PG.] As compared with case (1), note that the gain to the rest of society from reduced morbidity is higher than from reduced mortality. The discrepancy is due to savings in transfer payments. Provided these savings are considered acceptable to take account of, the PPG remains the same as the impact on the rest of society.

Note that the PPG from morbidity is the same as the averted loss for the government. Although this is the same result as would be obtained from a simple public sector cost-saving appraisal by the treasury, this would reflect the preferences of the members of society as defined in the extended utility function (3.4).

(4) Health improvements and net-recipient: PPG = 0

Within the chosen framework, treating those who remain out of work does not affect the value of society’s production, nor would it affect the transfer payments paid by the remainder of society. Hence, their treatment has no economic impact (beyond, of course, the non-monetary benefits in the numerator).
In sum, the distinction between the economic impact on the rest of society and potentially relevant production gains follows from an analysis of which magnitudes the society wishes to take into account in its decision making, ie which are ethically justifiable to include in a social evaluation. In case (2) we argued that in most societies it would probably be ethically unacceptable to include the present value of prolonged transfer payments as a negative gain. In case (3) some societies might hold the view that it is unacceptable to let savings in transfer payments count as gains, because this would imply that the economic gains from reduced morbidity are larger than the economic gains from reduced mortality. However, the larger health gains from reduced mortality would be accounted for in a CUA, and hence outweigh, or reduce, this eventual counter intuitive favouring of reduced morbidity.

‘Only Health Care Matters’

Table 2 suggests that the economic impacts on the rest of society represent the basis for judging which gains are potentially relevant, ie which gains society might find ethically acceptable to take into account in different situations. However, parts of the gain are manifested as public goods, G, which do not affect health. Although consumption patterns in terms of both private and public goods affect peoples’ health, it remains the case that there are few, if any, substitutes to health care for curing illnesses or alleviating pain. Further, it is primarily because of the effectiveness of health care in producing health that most societies are concerned with distributing health care according to health needs. In the limiting case, further analysed below, it is possible for a society to define all non-health benefits as irrelevant in the context of an economic evaluation of health care. In this case, when only health matters, and there are no alternatives to health care for producing health, the only production gain which are relevant within the CEA/ CUA framework is that which is manifested as increased health care, ie \( \Delta HC = tH (y_j - c_0) \).

To summarise: our analysis of the economic impact on the rest of society suggests that not all flows between the patient and society are likely to be ethically acceptable to take into account. Those which appear acceptable have been referred to as Potentially relevant Production Gains; PPG. These may lead to increased public goods and/or more health care. Whether gains in terms of increased health care should be counted as more important than gains in terms of public goods depends on the society’s SWF. Hence, the magnitude of PPG would depend on which gains society may find legitimate to include in the particular context of the analysis.

Although this analysis suggests that only parts of the PG are socially relevant, it remains the case that even in the limiting case, in which only health matters, the more productive members of society will contribute more to tax and therefore more to health care. Thus even with the concept of PPG, the CEA/ CUA would still favour the high income earners. One way of approaching this dilemma is to consider it as an example of a trade-off between equity and efficiency.

4.2 An Equity-Efficiency Trade-Off

The trade-off could be represented by adapting a model used by Culyer and Wagstaff (1993). Consider two groups of patients, A and B, with equal initial health and equal
capacities to benefit. In Figure 1, the capacities to benefit are shown in quadrants II and IV, as identical production functions of health care (HC) on health (H). Quadrant III illustrates the budget constraints depicted by the \( hc \) line. Assume that society has decided that only the taxes which are returned to the health sector are relevant (\( \Delta HC \)). If A is the productive group, the (net social) costs of treating A are lower than of B. The maximum available health care for B is at point HC\( \text{Bmax} \), while the maximum available health care for A would be at HC\( \text{Amax} \), which would be equal to HC\( \text{Bmax} \) plus the additional health care resulting from the production gains, \( \Delta HC \).

Quadrant I illustrates the health frontier. By tracing around all combinations of the budget hc between A and B, the health frontier h is derived. It is important to acknowledge that when society has determined which production gains are socially relevant (and these are positive), it implies that the true social costs differ. Hence, the hc line represents the real budget constraint and h the real health frontier. Interestingly, the asymmetric health frontier h is not a consequence of differences in capacities to benefit (as in Culyer and Wagstaff, 1993) but differences in the social costs of treating members of the two groups.

On this frontier the equity-efficiency trade-off becomes apparent. Should differences in social costs be taken into account when prioritising between the two groups? For egalitarians the answer is no, and their preferred point is the equitable distribution labelled E. For utilitarians the answer is yes, and their preferred distribution is labelled U, where the total health in society is maximized. By applying the standard social welfare function which was put in a health context in the seminal paper by Wagstaff (1991), it becomes clear that the optimal point would lie somewhere between E and U, eg point W, depending on society’s degree of aversion to inequality. It follows that when society has determined which production gains they would consider as potentially relevant, which in this case is limited to the increased availability of health care, and the strength of preferences for equity is taken on board in the SWF, then the socially relevant production gains are restricted to those gains which brings society to its optimum, ie to point W. In other words, for equity reasons society has decided not to exhaust all the potential production gains which would increase health care. By implication, the gains which are not exhausted would be considered as socially irrelevant in the context of an economic evaluation.
Figure 1
4.3 Which Compensation Principle Should be Applied?

If there are equity objections to letting differences in production gains matter when prioritising between different patient groups (ie choosing a distribution different from E), one way of analyzing these objections is to ask: in which ways may losers be compensated?

i) Compensation through private consumption

If the value of the increased production, PG - no matter in which form it is manifested - more than outweigh the value of the forgone health, then potential compensation is possible. This is the heart of the argument that a cost has the same value no matter who loses, and likewise, a benefit has the same value no matter who gains. Such compensations could be made possible in a market for transferable health care vouchers, whereby high income earners could buy health care from low income earners. The simple observation that such markets are not part of any health policy, suggest that compensation through increased private consumption is not socially acceptable.

ii) Compensation through public goods

The analysis of section 3.2 suggests that compensation through public goods might be socially acceptable. If society’s aggregate utility gain from the increased availability of G more than outweigh their lost utility due to the reallocation of health care from the less productive groups to the more productive groups (assuming positive but diminishing marginal productivity of health care on health for both groups of patients), then this outcome would be approved by utilitarians.

However, the nature of public goods is such that even potential compensation is hard to imagine. When i’s utility from G is not affected by j’s consumption of G, then it does not make sense for j to compensate i by reducing his consumption of G. The only way the losers (in terms of health care) could be compensated is if their increased own utility from G more than outweigh their lost utility from the forgone health care. What society then is doing is to accept that the less productive groups receive less health care, as long as they are being compensated in terms of more socially acceptable goods (than private consumption), ie in terms of G.

Compensation Through Health Care

Although the inclusion of that part of the PG which increases the health care budget would yield more total health, it is clear from Figure 1 that those who have no such gains will lose as compared with the situation where differences in production gains were not seen as a legitimate reason for discriminating patient groups. The only way of compensating them would be if patients from group A ‘paid their way in terms of health care’, ie if the costs of treatment were less than what would be channeled back as increased health care. This means that the health frontier would be upward sloping in the equity point (see Olsen, 1994, 1997 for an elaboration of this argument).
5 Concluding Remarks

The discussion about the inclusion of production gains in CEA has largely been in terms of either/or rather than ‘to what extent’. The important contribution of the Koopmanschap school (or the Erasmus group) is their approach to a more realistic estimation of the magnitude of production gains than when estimated by the traditional human capital approach. Their view is that it is correct to account for all production gains as estimated by the friction cost method, basically for the reason that one cannot ignore costs; all costs, no matter whether they are health care costs or ‘indirect costs’ affect the wealth of society, hence, include all the (correctly measured) production gains.

The counter arguments in the literature are twofold; first, there is the more technical argument that it would involve double counting of health and consumption, and second, the ethical objections to the distributional impacts, hence: exclude the gains. Interestingly, in the recent discussion between the Erasmus group (Brouwer et al, 1997a and 1997b) and the Washington panel (Weinstein et al, 1997), the latter was very explicit in the suggestion that: ‘[o]nly the impact on the rest of society, not including the deceased, belongs in the numerator’ of a C/E ratio. In other words, the Panel hold that it is correct to take account of the parts of the production gains which go beyond own consumption.

We have argued that, given the plausible assumption that utility is derived from consumption, the double counting argument is invalid. By contrast with both the Erasmus group and the Washington panel we have argued that there is no universally correct technical solution to this issue. It is a normative issue, whose solution depends on the value judgements on which the analysis is based.

Most publicly funded health care systems are based on a value judgement of ‘equal access for equal need’, implying that a health gain has the same social value irrespective of the income levels of the beneficiaries. Individuals are counted as equals, in the sense that consumption differences are ignored. If economic evaluations are to incorporate these value judgements, a dilemma arises: the more ‘indirect benefits’ (PG) are included, the more a patient’s priority will depend upon his income, but this is exactly what the introduction of CEA/ CUA was intended to avoid. Of course, by ignoring consumption differences, the analysis also ignores the impact upon the level of society’s wealth. This is what appears to frustrate many economists. However, it is not the economist’s role to impose the maximization of income and production as the social objective, nor to impose a Paretian SWF for the health service.

Ignoring consumption differences does not imply that the analysis is restricted to the interest of other members of society, thereby excluding the patient’s interest. First, in the CEA/ CUA the patient’s interest is taken into account through the inclusion of her health gains. Second, from an ex ante perspective all future members of society are potential patients. If these future members were to choose an allocation rule for society’s health care resources, we find it exceedingly unlikely that they would choose a criterion which distributes health care in proportion to the patients’ own consumption.
To sum up the argument in the previous sections: the magnitude of production gains to be included in a cost-effectiveness analysis depends on the value judgements of the society. If society takes the view that ‘a gain is a gain is a gain’, no matter who benefit from it and in what form it is manifested, and furthermore that the provision of collective health care should increase with the magnitude of a treated person’s value of production, then it is correct to include all the PG. However, if society takes the view that the only relevant gain is that from which the rest of society benefits, then it is correct to include only parts of the PG. Lastly, if society takes the view that only health gains matter, and if only health care produces health, then only that part which ends up as increased health care is correct to include.

There is an inevitable implication in all of these cases that as soon as production gains are subtracted in the numerator, the groups of patients whose cures do not affect the value of society’s production will be given a lower priority in the receipt of health care. This dilemma was analysed in the context of an equity-efficiency trade-off along the health frontier. The extent to which society allows differences in production gains to affect priority setting between patient groups, depends on its degree of acceptance of inequality. Furthermore, it depends on the sources of compensation, and, with which goods society finds it acceptable to compensate losers; private consumption, public goods or health care.

The analysis we have suggested introduces an apparently ad hoc distinction between socially relevant and socially irrelevant production gains and has, further, blurred the distinction between costs and transfers. In both cases, the fundamental reason for this is that we have rejected the analytical tradition of treating outcomes as being distributively neutral. We have explicitly argued that the costs that the remainder of the society will bear depends upon the type and distribution of the benefits.

We have argued that the literature on ‘indirect costs and benefits’ has neglected an entire dimension of the problem by focusing exclusively upon the technical question of measuring production gains. We have suggested that there is also a normative dimension, vis determining which part of the production gain is socially relevant. Further, the socially relevant component is likely to vary from society to society as it depends, firstly, upon the tax rates, and secondly, upon the strength of the commitment to egalitarianism in the health sector.


