Equity in health: the importance of different health streams

Paul Dolan\textsuperscript{a,c,*}, Jan Abel Olsen\textsuperscript{b,c}

\textsuperscript{a} Sheffield Health Economics Group, Department of Economics, University of Sheffield, Sheffield, UK
\textsuperscript{b} Institute of Community Medicine, University of Tromsø, Tromsø, Norway
\textsuperscript{c} Health Economics Research Programme, University of Oslo, Oslo, Norway

Received 21 June 2000; received in revised form 12 March 2001; accepted 7 May 2001

Abstract

This paper develops a conceptual framework in which preferences about the distribution of future health gains depend on differences in four ‘health streams’. These are as follows: (1) the amount of health to be gained; (2) the no-treatment profiles; (3) the amount of health experienced thus far; and (4) the amount of health gained previously as a result of public health interventions. This classification puts the well-established concerns for severity (stream 2) and age weights (stream 3) into a more complete analytical framework. Stream 4 has not been discussed to date and the paper suggests some moral arguments about the distributive relevance of this stream of health.

\textit{JEL classification:} I1

\textit{Keywords:} Equity-efficiency trade-off; Health gains; Distribution; QALYs

1. Introduction

It is widely accepted that one of the principal objectives of government expenditure on health care is to generate health. Since health is a function of both length of life and quality of life, the quality-adjusted life-year (QALY) has been developed in an attempt to combine the value of these attributes into a single index number. In this paper, it will be assumed that QALYs are an appropriate currency in which to express health benefits, but it is really only necessary to assume that there exists some currency that is considered appropriate (for example, perhaps simply life-years gained).

* Corresponding author. Tel.: +44-114-222-0670; fax: +44-114-272-4095.
E-mail addresses: p.dolan@sheffield.ac.uk (P. Dolan), jano@ism.uit.no (J.A. Olsen).
If health benefits were the only consideration, then the goal of health policy would be to maximise the number of QALYs gained. However, policy-makers, as well as the general public, are also concerned with how QALYs are distributed. Since a fair distribution of health gains is not necessarily the most efficient one, we are faced with a trade-off between efficiency (defined in this paper in terms of the maximisation of QALYs) and equity (which in this paper deals with the distribution of QALYs).

Wagstaff (1991) was one of the first to address an equity-efficiency trade-off in the domain of health. He proposed a social welfare function (SWF) approach to determine the optimal distribution of QALYs across different individuals or groups. In his analysis, the individuals started with identical endowments, i.e. the severity of their illnesses (or no-treatment profiles) were the same and they had experienced the same number of QALYs to that point. Nord (1995) has developed the idea that differences in the no-treatment profiles might be an important equity consideration in its own right. Citing evidence from a number of small-scale empirical studies, he asserts that people favour giving priority to those who have the worst prospects without treatment even if there are others who could gain more from treatment. The focus here, then, is on reducing inequalities in prospective health. Williams (1997) has developed the ‘fair innings’ argument, claiming that the number of QALYs an individual has experienced to date is also relevant and that future gains should be distributed so as to reduce inequalities in lifetime expectations of health.

The aims of this paper are two-fold. First, it seeks to develop a logically consistent framework that will serve both to bring the existing literature on the equity-efficiency trade-off in health together in a coherent way and to highlight the gaps in the literature to date. This is important since, whilst efficiency is often taken to be the sum-ranking of benefits, equity has many different (and sometimes contradictory) meanings. Our framework suggests that preferences for how QALYs are to be distributed depend on differences in four ‘health streams’: (1) the number of QALYs to be gained; (2) the no-treatment QALY profiles; (3) the number of QALYs experienced, thus far and (4) the number of QALYs gained previously as a result of health care (in what follows, we use the term ‘health care’ in a wide sense to refer to any publicly-funded health intervention).

To our knowledge, the last of these considerations has not been discussed before, at least not in the context of outcomes. Rawls (1972) distinguished between those ‘goods’ whose distribution was determined ‘naturally’ and those for which the distribution was ‘socially’ determined. In some ways, this is analogous to our distinction between streams 3 and 4. However, Rawls was concerned only with the allocation of ‘primary social goods’ and, somewhat mistakenly, considered health to be naturally determined. Therefore, his theory applies to individual’s who are “normal, active, and fully co-operating members of society over the course of a complete life” (Rawls, 1972). In addition, Rawls’ main concern was with the social arrangements that ensured each individual’s rights to primary goods, whereas our concern here is exclusively with outcomes (rather than rights or procedures).

The failure to distinguish between outcomes in this way is surprising since concepts of entitlement have widespread appeal, not only in health but also in other areas of public policy. For example, the unemployment benefit that a person is entitled to in the UK is in part a function of the amount of benefit they have received previously and, in education, everybody is entitled to a given number of years of ‘free’ university education. Therefore, the second aim of this paper is consider whether a morally relevant attribute of the equity-efficiency
trade-off is the extent to which the amount of health a person has experienced up to this point has been the result of previous benefits from health care. After setting out the various moral positions on this issue, it will be our contention that the distinction between previous health that was ‘free’ and previous health that was the result of health care is a morally relevant one but only in very limited circumstances. We begin, though, by describing the different ‘streams of health’ in more detail.

2. Four different streams of health

Our categorisation of the different streams of health for any given individual is shown in Fig. 1. Of the four streams, only one is taken into account when evaluating the cost-effectiveness of different health care interventions, namely, the health gains profile, often measured as the expected number of QALYs gained (stream 1). Given that cost-effectiveness analysis is becoming increasingly widespread, most health economists have focused attention on this stream only (Drummond et al., 1997). Consequently, the need for health care has most often been defined solely in terms of capacity to benefit, i.e. ‘need as capacity to benefit’ (Culyer, 1997). If the objective is to maximise health, then this stream is the only relevant one.

The second stream represents the expected QALY profile of an individual or group that is left untreated. This stream corresponds with an alternative definition of need, namely, ‘need as ill health’ (or, more accurately, need as expected ill health over the remaining lifetime). There are two reasons why we would be interested in differences in the no-treatment profiles. The first is that we might simply care for those with poor health prospects. Two recent Norwegian commissions on priority setting in health care have identified that an important rationale for government involvement in health care is to provide benefit to those with the worst health prospects (Olsen, 1997). There is also evidence from a number of empirical studies that people wish to devote considerable resources to improve the health of seriously ill people, particularly those facing an immediate risk of death (Hadorn, 1991).

<table>
<thead>
<tr>
<th>Retrospective health</th>
<th>Prospective health</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Gained’ health</td>
<td></td>
</tr>
<tr>
<td>i.e. from health care</td>
<td>4 Previous QALYs gained from health care</td>
</tr>
<tr>
<td></td>
<td>1 Expected QALY benefit from health care</td>
</tr>
<tr>
<td>‘Free’ health</td>
<td></td>
</tr>
<tr>
<td>i.e. not from health care</td>
<td>3 Previous QALYs yielded without health care</td>
</tr>
<tr>
<td></td>
<td>2 Expected QALY profile without health care</td>
</tr>
</tbody>
</table>

Fig. 1. A taxonomy of health streams.
The second reason why this stream might be important is because of the concern for fairness in the distribution of health. Economists have usually defined the equity-efficiency trade-off in terms of a trade-off between maximising QALYs and equalising health (Culyer and Wagstaff, 1993). If the objective is to reduce inequalities in future health, it follows that QALY gains should be distributed initially to those who can expect the worst prospective health if they are left untreated. As a result, there now exists a number of different ways in which this trade-off can be measured (Dolan, 1998) and a range of empirical studies which shed light on its extent (Nord, 1995). Note that this trade-off is only in the domain of prospective health in the sense that the discussion relates only to the distribution of QALYs after the point at which a resource allocation decision is made. However, it is entirely plausible that the number of QALYs that a person has experienced up to the decision point might also be a relevant equity consideration and, therefore, streams 3 and 4 relate to retrospective health.

There is increasing recognition in the literature that past health is something that can influence the value of future health gains. This has usually taken the form of a discussion about the relevance of age, either at the individual or social level. At the individual level, it has been shown that the importance attached to health varies according to the individual’s life stage (Williams, 1988). This means that the size of the expected QALY gains will be a function of the age of the potential recipient. This is, of course, an efficiency condition, and in principle it can be captured within an appropriate measure of benefit. At the social level, the equity condition relates to the concern that many have for reducing inequalities in life expectancy or in lifetime health more generally (Williams, 1997).

Discussions of these issues, thus far have not separated out streams 3 and 4 from one another. They have implicitly assumed that it is not relevant how much of the health that a person has received up to the decision point is due directly to health care and how much is due to other things, such as lifestyle or luck. Our distinction in Fig. 1 between streams 3 and 4 follows logically from our distinction between streams 1 and 2. Prospective health distinguishes between QALYs gained as a result of future health care and those gained if no health care were to be provided. Likewise, retrospective health distinguishes between QALYs gained as a result of past health care and those gained ‘free’ of health care.

In practice, of course, the four streams of health are difficult to separate out from each other. In particular, it would be almost impossible to know precisely how much of an individual’s previous health was directly due to health care. And while it is also difficult to assess the expected future health gain from current health care, this is nevertheless the objective of economic evaluation, health technology assessment and evidence-based medicine. We must emphasise, however, that our concern is with the conceptual relevance of separating out health into the four streams shown in Fig. 1, and not with the extent to which it is empirically possible to do so. Having said this, in order to facilitate the various distinctions, we will need to make some assumptions about the costs, effectiveness, timing and number of previous gains.

Since we are concerned with the distribution of health gains (and not with the allocation of resources per se), it will be assumed that all previous and future decisions have been — and will be — made on the basis of the expected cost per discounted QALY. In this way, only the most cost-effective interventions have been — and will be — undertaken. We make no claims here about how uncertainty and time should be accounted for in the QALY
algorithm and leave it to others to consider the implications for our framework of different models of dealing with uncertain and intertemporal decisions. Against this background, we can now consider the implications for the equity-efficiency trade-off if different streams of health are distinguished from one another in the way shown in Fig. 1.

3. Defining and measuring the efficiency-equity trade-off

The socially optimal distribution of health gains across two or more individuals or groups depends on the location of the feasibility set within the total health space. The actual size and shape of this set depends on the health care budget, the health production function and the no-treatment (or endowment) point. Given the Pareto condition, there is always a point on the health possibility frontier that is superior to an interior point and, in this sense, any point on the frontier is an efficient one. However, in this paper, efficiency is defined as that point on the frontier at which the sum of all individuals’ health is maximised — this follows from Wagstaff (1991) and is often referred to as the utilitarian solution. The important question in the context of this paper is “which frontier point is preferred?” the answer to which would then reflect society’s trade-off between equity and efficiency.

This question will be addressed using a diagrammatic exposition of a Bergsonian-type SWF which is symmetric around the equity line and which has an iso-elastic curvature (Atkinson and Stiglitz, 1980). A health-related SWF (HRSWF) with these properties was originally used by Wagstaff (1991) and, more recently, by Dolan (1998). Given that there are four different streams of health outlined in the taxonomy in Fig. 1, which frontier point is preferred will depend on the relative weight given to each stream, or the extent to which each equity line ‘pulls’ the frontier point away from the efficient one.

First consider Fig. 2 which shows a health feasibility set with a health frontier $h_A h_B$. The axes, $PGA$ and $PGB$, reflect the amounts of prospective gains for two individuals (or groups of similar individuals), A and B. The shape of the frontier suggests that A has a greater capacity to benefit from treatment than B. The efficiency point is labelled $H_{\text{max}}$. Ceteris paribus, in terms of the individuals being identical according to their (i) no-treatment profiles;

![Fig. 2. Trade-offs in the prospective health space.](https://example.com/fig2.png)
(ii) previous total health and (iii) previous health gains, then the appropriate trade-off would be against the equal prospective gains line, $E_{PG}$ (i.e. equity with regard to stream 1 in Fig. 1). Hence, the arrow labelled 1 is oriented towards $E_{PG}$, suggesting that the iso-welfare curve is symmetric around the $E_{PG}$-line. This is the type of trade-off discussed by Wagstaff (1991).

We can relax each of our ceteris paribus conditions in turn, starting with the no-treatment profiles. The issue here is to locate the feasibility set within the prospective health space of Fig. 2, which is constrained by the axes $PH_A$ and $PH_B$. In this case, A is assumed to be in a more severe condition than B. If equity preferences relate to prospective health, the appropriate trade-off would now be against the equal prospective health line, $E_{PH}$ (i.e. equity vis-à-vis the sum of streams 1 and 2). Hence, the arrow labelled 2 is oriented towards $E_{PH}$, suggesting that the iso-welfare curve is symmetric around the $E_{PH}$-line rather than the equal prospective gains line, $E_{PG}$.

In the example shown in Fig. 2, the person who has the most to gain also happens to be in the most severe state. Thus, for reasons of reducing inequalities in prospective health, A would receive even more health care than what the health maximising solution suggests, i.e. a point that lies somewhere on the frontier between $H_{max}$ and $E_{PH}$. If the feasibility set were located on the other side of the equal prospective gains line, this would have illustrated the type of trade-off between ‘need as capacity to benefit’ and ‘need as ill health’, which

![Fig. 3. Trade-offs in the total health space.](image)
Dolan (1998) expressed in terms of a Cobb–Douglas HRSWF. However, our rationale for Fig. 2 is to illustrate that equity preferences with regard to prospective health might pull in the opposite direction to equity preferences over prospective gains.

The next ceteris paribus condition to be relaxed is that of equal previous health. The issue here is to locate the feasibility set, as well as the prospective health space, within the total health space, as shown in Fig. 3. This is constrained by the two individuals total health; the axes $TH_A$ and $TH_B$. If $B$ is now assumed to have experienced less health than $A$, and if society has some preference for reducing inequalities in lifetime health, the iso-welfare curve would be symmetric around the equal total health line, $E_{TH}$ (i.e. equity vis-à-vis the sum of the four streams). Hence, the arrow labelled 3 is oriented towards $E_{TH}$.

Notice that in Fig. 3, arrow 3 points in the opposite direction to arrow 2. This illustrates a hitherto neglected trade-off in the literature between preference for the most severely ill (as argued for by Nord (1995)) and ‘age-egalitarianism’ (as argued for by Williams (1997)). This particular trade-off is potentially one of the most policy-relevant ones since it forces us to consider how we would weight the health gains to a less severely ill younger person against the health gains to a more severely ill older person.

The first aim of this paper is now complete; we have brought the existing literature on the equity-efficiency trade-off together and highlighted the gaps in it. But, of course, this
still leaves open the questions relating to stream 4. We, therefore, need to introduce the final equity line that takes account of the relative contribution of health care to retrospective health gains. In Fig. 4, \( E_{TG} \) is the equal total gains line (i.e. equity vis-à-vis the sum of streams 1 and 4). If part of B’s health is attributable to previous health care use, and if equity preferences are related to total health gains, then the arrow labelled 4 is oriented towards \( E_{TG} \), suggesting that the iso-welfare curve is symmetrical around the \( E_{TG} \)-line.

The directions of the four arrows indicate partial effects, i.e. a trade-off between the efficient solution, \( H_{max} \), as compared with one equity line only. In practice, however, and as illustrated in Fig. 4, there may be differences in all the four streams of health. Therefore, there are likely to be many situations in which the iso-welfare curves will not be symmetric around any one equity line, due to the ‘pull’ from other equity lines. Moreover, depending on the relative amount of health in each of the four streams, there are circumstances under which two, three or even all four of the equity lines may ‘pull’ us away from \( H_{max} \) in the same direction (e.g. consider a young severely ill person who has no previous gains from health care).

It is not the purpose of this paper to discuss the relative weights that might be given to each of the equity considerations. Rather, since this paper has introduced a distinction between previous health gained as a result of health care and health that was experienced ‘free’ of any health care, we consider whether stream 4 has any moral relevance at all. Whilst stream 4 may also have implications for efficiency in that the threat of receiving lower priority if unhealthy behaviour is repeated may act as an incentive to prevent that unhealthy behaviour, such ‘moral hazard’ issues are not the focus here.

4. The potential relevance of previous health gains

Essentially, the question here is: Is it morally relevant to take account of the QALYs people have gained from health care in the past when making decisions relating to the distribution of QALYs in the future? As with many questions, the answers are basically ‘yes’, ‘no’, and ‘it depends’. An answer of ‘yes’ can be explained by the adoption of a lifetime perspective, whereby the distribution of QALY gains would be looked at from the time when a person is born rather than from any particular decision point. Such preferences over equality with respect to total gains could then be explained by a view that everyone is entitled to a given QALY gain from health care over the course of their lifetime.

An answer of ‘no’ would be to consider that all retrospective gains from health care are ‘sunk’, and should have no bearing on the distribution of prospective gains. This would seem to be the implicit argument in much of the health economics literature that has discussed equity in health (see Williams and Cookson (2000) for a review). An argument here might be that health care rarely generates utility from consumption in itself, and so it would be inappropriate to discriminate against people simply because they have previously needed health care to improve their health.

Those who say ‘it depends’ to the above question are expected to ask for information about the causes of ill health and, in particular, about the extent to which the need for health care has been endogenously determined (e.g. through risky behaviour or an unhealthy lifestyle). This is obviously a controversial issue, both insofar as the degree of control that
individuals have over their own actions is concerned, and the extent to which there exists a relationship between a particular action and subsequent ill health. We make no substantive claims about these issues. For our purposes, we need only to assume that all conditions can be located along an analytical spectrum from being exogenously determined (due to ‘bad luck’) through to being endogenously determined (as a result of well-informed ‘own choices’).

The following three intermediate viewpoints are based on the notion that an individual’s entitlement to collectively funded resources should depend on the extent to which she can be held responsible for why she needs those resources. This discussion is set within the context of a tax-based system in which the motivation for the cross-subsidisation of health care is provided in part by a ‘caring externality’ (Culyer, 1971). It may be the case, however, that careless behaviour makes other people care less.

The first viewpoint, and the one closest to an answer of ‘yes’ to the above question, would take account of past health gains if the previous need for health care was partly the result of ‘own choices’ — even if the current need is due to ‘bad luck’. The argument would be that, while such people were given the same entitlement as others previously, they lose some of their entitlements to health care afterwards. This viewpoint can be challenged on the grounds that the current need for health care is the result of ‘bad luck’ and this is what is relevant to the current decision.

The second viewpoint would take account of past gains if the previous and the current need for health care were both partly endogenously determined — even if those ‘own choices’ may have been different. So, a person who required treatment for lung cancer as a result of smoking would be given less weight if they had previously required treatment for a drinking-related illness, on the grounds that there is a limit to what a person can expect to receive if health care is needed as a result of her own actions. The challenge to this argument is that, since current need is unrelated to previous need, these people have not had any direct previous experience of the relationship between their current actions and their current need for health care.

The third viewpoint considers past health gains to be relevant in very limited circumstances and, as such, comes closest to an answer of ‘no’ to the above question. Here, past gains are relevant only when the previous and the current need for health care are both partly the result of the same endogenously determined actions. The argument is that those who have received health care as a result of ‘own actions’, and who have also been informed that the continuation of those actions might mean that more health care will be needed in the future, should have less entitlement to health care if they do not modify their behaviour. For example, obese people who in the past were treated for heart disease caused by their inactivity and choice of diet, and who now require treatment due to the same causes would have less entitlement to health care than other obese people requiring treatment for the first time. The counter argument is that people have different intellectual and psychosocial capabilities to learn from past actions, and so we should not punish those who cannot modify their behaviour.

The crucial issue in these three intermediate viewpoints is the association between a previous unhealthy action that caused the past need for health care and the cause of the current need for health care. According to the first view, past unhealthy actions make one lose entitlement to health care in the future — ‘one strike and you are out’. In the second
view, it is repetitive unhealthy actions that are to be punished — ‘two strikes and you are out’. And for the third view, it is repetitive unhealthy actions of the same kind that ought to result in less entitlement — ‘two similar strikes and you are out’.

5. Concluding remarks

The aims of this paper have been, first, to develop a framework within which to discuss different types of equity-efficiency trade-off within the domain of health, and second, to consider whether the extent to which a person’s health is the result of previous benefits from health care is a morally relevant attribute of this trade-off. We have shown that the distribution of the benefits from health care can be analysed according to four different streams of health. If the sole objective of health policy was to maximise health gain, or if there was only a concern for the distribution of health gain in isolation of any other considerations, then only prospective gains (stream 1) would be relevant. However, there is evidence that the no-treatment health profile (stream 2) is important to policy-makers and the public alike. And recently, economists and others have focused attention on the level of health that people have experienced prior to the point at which a prioritisation decision is made (streams 3 and 4).

We have discussed these considerations simultaneously rather than in isolation of one another using a diagrammatic exposition of the HRSWF. In doing so, it appears to us that the relevant literature to date has overlooked the importance of identifying the health frontier in the context of the different equity lines spelled out in this paper. Either the literature has not asked the question ‘equality of what?’ or it has assumed that other streams of health are identical. Our framework has identified that an important area for future research would be an investigation into the trade-off between preferences for the most severely ill (stream 2) and for those with a lower stock of health (streams 3 and 4). Since the analytic distinction between streams 3 and 4 (i.e. how much of people’s retrospective health levels came about without health care and how much resulted from health care) is a new one, we have discussed different views about the moral relevance of the distinction.

Various arguments related to duty and desert can be used to support an answer of ‘it depends’ to the question about the moral relevance of past health gains. It might be seen as our duty to lead a healthy lifestyle. This could be because people should try to avoid the unnecessary use of services paid for by their fellow citizens, or simply because a healthy lifestyle is held to be morally superior (see Skrabanek, 1994 on ‘healthism’). And if entitlements to health care depend on the goodness of past social actions, a citizen who has been destructive to her health in the past, might deserve less entitlement to health care in the future (LeGrand, 1991). Duty and desert are closely related to the view that social justice requires ‘equality of opportunity’ (rather than, say, equality of outcomes). Here, different ‘own choices’ are seen as legitimate grounds for different entitlements — provided that proper allowance has been made for the fact that some ‘types’ of people are better able to make informed choices than others (see Roemer, 1998).

We feel that it would be wrong to have a general principle that gave less weight to the future health gains of people whose previous gains were the result of ‘own actions’. We also think that it would be wrong to ‘blame’ someone for their previous actions if their current
actions are different, since the lessons learned from a past ‘mistake’ cannot be expected to have much impact upon a different ‘mistake’ now. But people can learn from making the same ‘mistake’. And so it is our contention that there is a moral argument for ‘looking back in anger’ if the cause of current ill health is the same as the cause of previous ill health and if the individual was informed when they were ill previously that future health care might be required if they did not modify their behaviour. This corresponds to the third intermediate viewpoint above.

Of course, readers are free to disagree with us, and it is an interesting question both for normative debate and empirical research as to which of the viewpoints outlined in this paper society ought to or would want to adopt. There has been some empirical investigation into how the general public would prioritise self-inflicted conditions in general (see, for example, Dolan et al., 1999), but there has been little investigation into how they view repeatedly self-inflicted conditions in particular. And so this represents one obvious avenue for future research into people’s preferences.

Whatever the conclusion others may reach regarding these issues, the framework we have developed provides a background against which these issues can be discussed. It might also serve to enhance our understanding of the factors that policy-makers take into account when making resource allocation decisions. Finally, the conceptual framework highlights possible ways in which future empirical research into equity-efficiency trade-offs could be conducted.

Acknowledgements

This paper has benefited from the comments made by participants at the RES conference in Nottingham, the iHEA conference in Rotterdam and at seminars in Bergen, Oslo, Tromsø, Sheffield and Calgary. We would also like to thank two anonymous referees for their extremely helpful suggestions.

References