2007/90

■

Unfair inequalities in health and health care

Marc Fleurbaey and Erik Schokkaert

CORE DISCUSSION PAPER
2007/90

Unfair inequalities in health and health care

Marc FLEURBAEY [1] and Erik SCHOKKAERT[2]

November 2007

Abstract

Inequalities in health and health care are caused by different factors. Measuring "unfair" inequalities implies that a distinction is introduced between causal variables leading to ethically legitimate inequalities and causal variables leading to ethically illegitimate inequalities. An example of the former could be life-style choices, an example of the latter is social background. We show how to derive measures of unfair inequalities in health and in health care delivery from a structural model of health care and health production: "direct unfairness", linked to the variations in medical expenditures and health in the hypothetical distribution in which all legitimate sources of variation are kept constant; "fairness gap", linked to the differences between the actual distribution and the hypothetical distribution in which all illegitimate sources of variation have been removed. These two approaches are related to the theory of fair allocation. In general they lead to different results. We propose to analyse the resulting distributions with the traditional apparatus of Lorenz curves and inequality measures. We compare our proposal to the more common approach using concentration curves and analyse the relationship with the methods of direct and indirect standardization. We discuss how inequalities in health care can be integrated in an overall evaluation of social inequality.

Keywords: equity in health care delivery, health inequality, social welfare

JEL Classification: D63, I10

[1] University Paris-Dauphine, CNRS, LSE and IDEP, France. E-mail: marc.fleurbaey@univ-paris5.fr

[2] Department of Economics, KULeuven and CORE, Université catholique de Louvain, Belgium. E-mail: erik.schokkaert@econ.kuleuven.be

1 Introduction

There is by now a very large literature on different aspects of inequity in health, both from a theoretical and from an empirical point of view. This literature focuses mainly on socioeconomic inequalities in health and in the delivery of health care (Wagstaff and Van Doorslaer, 2000a). While different methods (including the calculation of odds ratios) have been proposed in the public health literature (Mackenbach and Kunst, 1997), the concentration curve has become the workhorse in most health economic studies. Recently, a number of papers have been published which propose a welfare economic foundation for its use (Wagstaff, 2002; Koolman and van Doorslaer, 2004; Bleichrodt and van Doorslaer, 2006) or advocate alternative approaches (Bommier and Stecklov, 2002; Becker et al., 2005; Abul Naga and Geoffard, 2006; Zheng, 2006; Dias and Jones, 2007; Fleurbaey, 2006b).

When moving from the measurement of inequality (in health or in health care) as such to socioeconomic inequality, one implicitly assumes that policy-makers are more concerned or should be more concerned about some causes of observed overall inequality, such as socioeconomic background, than about other causes. In the literature on health inequality, it is implicitly accepted that health inequalities within a socioeconomic group are less problematic than health inequalities between socioeconomic groups. And in the literature on equity in health care delivery, it is quite reasonably taken for granted that differences in use which reflect differences in needs are not only unproblematic, but even desirable. All in all, this strongly suggests that some inequalities are "legitimate" while others are not. The most obvious justification for making this distinction between "legitimate" and "illegitimate" differences is that the former can be attributed to causes that belong to individual responsibility. Given this background, it is striking that, while there are clear links between the literature on income inequality and the literature on socioeconomic inequalities in health and in health care, there has been until now hardly any link with the growing literature in social choice on equity, responsibility and compensation (Roemer, 1998; Fleurbaey, 2008). This paper tries to bridge part of that gap.

As soon as one formulates the problem of illegitimate or "unfair" inequalities in this

general framework, one is immediately confronted with the observation that, in addition to socioeconomic background, there are many more causes of inequalities that may be a cause of ethical concern. Why should we then focus almost exclusively on socioeconomic inequalities in health care consumption? Is it less problematic if someone is denied health care because she lives in another region of the country? (Or, for that matter, in another country?) Why should we not be interested in equality of health as such? Because part of these health differences is unavoidable, or not created by socioeconomic institutions? And, suppose we restrict ourselves to avoidable health inequalities, why then focus again almost exclusively on socioeconomic health inequalities? And what if socioeconomic differences in health can to some extent be explained by differences in lifestyle? In this paper we will define an "equitable" situation as a situation without unfair inequalities - and inequalities are defined as unfair when they follow from causes which do not belong to the sphere of individual responsibility. The socioeconomic background of individuals is one of these causes - but, although very important, it is not the only one. It is necessary to get a more complete perspective on these different causes.

The method we propose consists of three steps. In the first (explanatory) step one has to construct a structural model to estimate the relative importance of the different causes of inequality and to get a better insight into their possible interactions. In a second (normative) step, one decides which of these causes lead to legitimate and which to illegitimate (or unfair) inequalities. The third step involves the measurement of these unfair inequalities. We will focus on the choices to be made in that third step.

There are of course different opinions in society with respect to what belongs to the sphere of individual responsibility. Some will claim that equality of access is a better criterion than equality of use, because individuals should be held responsible for their own choices. Some will claim that health differences following from differences in life-style are not problematic, because individuals should be held responsible for their smoking and drinking behavior. Some will claim that health differences reflecting differences in age or in genetic endowments are not unfair, because they are unavoidable. But in each of these cases there are also proponents of the opposite view. We will show how different

4

views about equity (or about unfair inequalities) can be interpreted as different views about where to draw the line between legitimate and illegitimate causes of differences, i.e. as different options taken in the second step referred to above. Our method to measure inequality (the third step) works for any of these options and can therefore accommodate many different ethical views. This has the advantage that one can also compare the results for different approaches within one general encompassing framework.

In our view, unfair inequalities in the health domain cannot be separated from unfair inequalities in other domains. The overall social objective is to minimize unfair inequalities in welfare. Health is important because it is one of the most crucial dimensions of welfare. Health care is important because it contributes to better health, and perhaps also directly to a higher welfare level. Although there is this clear hierarchy, we agree that it is useful to consider also inequalities at the lower levels, not in the least because health care and health policy are separate policy domains.

We introduce our basic concepts in section 2 for a simple case with two variables. We propose two possible approaches to measuring unfair inequalities. Direct unfairness refers to inequalities in health or health care after one has removed the effect of all legitimate variables. The fairness gap measures the distance between the actual distribution and a fair distribution in which all the effects of illegitimate variables have been removed. We show why, in general, these two approaches do not yield the same results. Section 3 sketches the broad contours of a structural model of health and health care. Section 4 shows how the concepts from section 2 and the structural model of section 3 can be combined to conceptualize unfair inequalities in health care and in health. We also discuss within our framework the problem of aggregating the different elements in the health care vector for the purpose of measuring inequity. In section 5 we argue that some additional normative choices have to be made when moving from the empirical model to the calculation of inequality. We show that direct unfairness is analogous to direct standardization and that the calculation of the fairness gap is related to indirect standardization, when the latter technique is reinterpreted to include all relevant variables. In section 6 we compare our approach to the traditional work on socioeconomic inequalities using the concentra-

tion curve and we suggest one possible way to integrate health issues in a broader concern for equality of welfare. Section 7 concludes.

We focus on conceptual issues with respect to the definition of unfair inequalities and we do not really go into problems of implementation. Although we are well aware that the level at which, e.g., health is measured, may have an influence on the measurement instruments to be used (see, e.g., van Doorslaer and Jones, 2003; Erreygers, 2006), we completely neglect this issue.[1] Moreover, throughout the paper we work within an absolute approach to measuring inequalities. This means that we will be referring to absolute Lorenz curves and to inequality measures which satisfy translation invariance, i.e. which do not change when a constant is added to all the elements of the vector. This is not in line with the dominant practice in economic inequality measurement (including the measurement of socioeconomic inequalities in health and health care), in which relative Lorenz curves and scale invariant inequality measures have been much more popular. Our choice in favour of the absolute approach brings our paper more closely to the social literature on responsibility and compensation where absolute distances have been used more often than relative proportions. However, this choice is not necessary, and all the axioms and results of this paper can be easily reformulated within a relative approach.

2 Direct unfairness and the fairness gap: a simple example

Let us introduce the basic issues of this paper with a simple example. For illustrative purposes, we will focus on inequalities in health. In later sections, we will apply the same ideas in a more elaborate model and also consider the issue of equity in health care delivery. Let us assume that the health of individual i $(i = 1, ..., n)$ is determined by her

---

[1] Although we will be concerned with measurability and comparability of individual welfare in the very last section.

income y_i and by her life-style $l_i{}^2$, i.e.

$$h_i = h(y_i, l_i) \tag{1}$$

Neglecting all problems of measurability -as we will do throughout the paper- it would be straightforward to construct Lorenz curves for health or to calculate inequality in health. However, from an ethical point of view, we are mainly interested in ethically objectionable or unfair inequalities. Let us for the sake of the argument take it for granted that health inequalities due to differences in life-style are unproblematic, because we want to hold people responsible for these.[3] Therefore, a measure of unfair inequalities should not reflect health differences due to differences in life-style. In our simple example, this means that we only want to measure so-called "socioeconomic inequalities in health".

How to go from "overall inequality" to "unfair inequality"? One way to approach the problem is to see it as an exercise of removing from the overall inequality measure all differences which are due to lifestyle. What then should remain is a measure of health inequalities due to income differences, and to income differences only. In general (but very loose) terms, we can summarize this condition for later reference as

Condition 1 *(NO INFLUENCE OF LEGITIMATE DIFFERENCES). A measure of unfair inequality should not reflect legitimate variation in outcomes, i.e. inequalities which are caused by differences in the responsibility variables.*

Another approach starts from the concept of a fair distribution. In our example, in a fair distribution there should be no health inequalities due to income differences. This implies that if two individuals have the same life style, they should have the same health

---

[2]For the purpose of this simple example, we use income as an indicator of socio-economic status. As we will see in the next section, in a broader setting individuals may be held partly responsible for their income. Moreover, in the real world the health situation of the individuals is determined by many more variables, not in the least their genetic endowment. We come back to this issue in the later sections. For the purpose of the simple example in this section, we assume that all these other variables are identical for all individuals.

[3]This starting point can be -and has been- hotly debated. We will return to that issue later on in the paper. At this stage, we only want to illustrate the basic issues related to measuring unfair inequalities.

level, whatever their income. Again, more generally (but very loosely) formulated, we can say that a measure of unfair inequality should satisfy the following condition:

Condition 2 *(COMPENSATION) If a measure of unfair inequality is zero, there should be no illegitimate differences left, i.e. two individuals with the same value for the responsibility variable should have the same outcome.*

At first sight, both conditions are perfectly clear and it seems obvious that a good measure of unfair inequality should satisfy both. However, there is a basic problem in that the two conditions are incompatible as soon as the effect of income on health is not independent of the life-style. This basic problem is well documented in the social choice literature and discussed in a long series of publications (a synthesis can be found in Fleurbaey, 2008). Its consequences, however, have not yet been fully realized in the literature on the measurement of socioeconomic inequalities in health care or in health.[4] Without going into the formal details, we can use our example to convey the basic intuition in a straightforward way. Let us first introduce two methods to measure unfair inequality. These two methods are closely related to the concepts of conditional equality and egalitarian-equivalence in the literature on fair allocation (see, e.g., Fleurbaey, 2008).

The first method (*conditional equality*) focuses on condition 1. It removes the legitimate differences by fixing the value of l_i in (1), i.e. by defining a "corrected" value of health $\widetilde{h}_i = h(y_i, \widetilde{l})$. This is the health level that individual i with income y_i would reach if he had the reference lifestyle. Inequality in $\widetilde{h}$ can immediately be measured with the traditional apparatus of Lorenz curves and inequality measures. We propose to call this inequality **direct unfairness.** By construction, a measure of direct unfairness can only reflect variation due to income differences, since differences in life style are kept constant. Therefore it satisfies condition 1. However, there is no reason why it would satisfy condition 2: if there is no inequality in $\widetilde{h}$, this does not at all guarantee that two individuals with the same life style will also have the same health level.

---

[4]Gravelle (2003) and van Doorslaer et al. (2004) touch the issue, but do not really go into the normative implications. Schokkaert et al. (1998) and Schokkaert and Van de Voorde (2004, 2006) have shown its relevance for the problem of risk adjustment.

The latter condition is satisfied automatically by a second method (*egalitarian-equivalence*), in which we first explicitly define a fair distribution, i.e. a distribution in which all the illegitimate sources of variation have been removed. A straightforward way to do this is to fix the value of y_i in (1) and to define a reference health level for i as $h_i^* = h(y^*, l_i)$. In a fair distribution, the difference between this ideal reference situation and the actual situation should be zero or at least equal for all i. Unfair inequality can therefore be measured by applying the traditional inequality measurement apparatus to the vector $(h_i - h_i^*)$. We call this the approach of the **fairness gap**. It is immediately clear that it satisfies the compensation condition 2. However, in general it does not satisfy condition 1: the fairness gap may be influenced by life style, because the differences $h(y_i, l_i) - h(y^*, l_i)$ may depend on the value taken by the variable l_i.

In general, the two approaches will not yield the same conclusions. Measures of direct unfairness satisfy condition 1, but not condition 2. Measures of the fairness gap satisfy condition 2, but not condition 1. There is one interesting case in which they do coincide, however. Suppose that eq. (1) is additively separable, i.e. that it can be written as $h(y_i, l_i) = f(y_i) + g(l_i)$. This implies that the effect of income differences on health is independent of the life style (and vice versa). In this case, direct unfairness measures the inequality in the vector $(f(y_i) + g(\widetilde{l}))$, while the fairness gap measures the inequality in the vector $(f(y_i) - f(y^*))$. The two will give the same result in our absolute measurement approach.[5]

A picture may illustrate the issues. Take income to be a continuous variable and suppose there are two different lifestyles in society, denoted l^A and l^B. The figure shows the functions $h_i^A = h(y_i, l^A)$ and $h_i^B = h(y_i, l^B)$. We assume that l^A is the healthier lifestyle

---

[5]Remember that we opt in this paper for an absolute approach to inequality measurement, in which adding a constant to all elements of a vector does not change inequality. As mentioned before, the same basic intuitions hold also for the relative approach. To be more specific, one could instead define the fairness gap in relative terms: h_i / h_i^*. With this formulation, the direct unfairness and the fairness gap approaches are equivalent if the health function is multiplicatively separable: $h(y_i, l_i) = f(y_i) g(l_i)$. Indeed, one then has $\widetilde{h}_i = f(y_i) g(\widetilde{l})$ and $h_i / h_i^* = f(y_i) / f(y^*)$. In this case, relative inequality measures and the Lorenz curve are identical for $\widetilde{h}_i$ and for h_i / h_i^*.
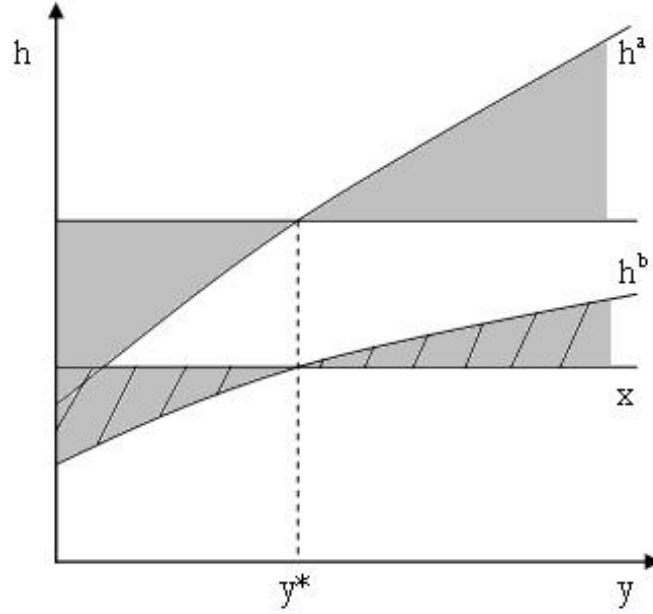
Figure 1: Direct unfairness and the fairness gap

and that for both lifestyles there is a positive relationship between health and income. Differences in health due to differences in lifestyle are considered to be unproblematic, but fairness requires that all individuals with the same lifestyle should have the same health level whatever their income position, i.e. that the curves in the figure should be horizontal lines.

When measuring *direct unfairness*, we fix the lifestyle at a reference value. Let us say that we focus on one specific curve (say, we put $\widetilde{h} \equiv h^B$). We will then measure the inequality in the distances between this curve and the horizontal line x.[6] These distances are contained in the striped area in the Figure. It is obvious that this procedure does not satisfy the compensation condition: indeed we fully neglect the unfairness which is implicit and (in this case larger) for lifestyle h^A. On the other hand, the procedure satisfies condition 1, since the only health differences reflected in the inequality measure by construction are due to differences in income.

---

[6]Given that we focus on absolute inequality measures, the exact position of line x does not matter, as long as it is horizontal.

When we calculate the *fairness gap*, we fix y at a given level (say y^*) and we compute for each individual the difference between his actual health and the health level that he would reach with his actual lifestyle in the hypothetical situation that he had income y^*. These distances are contained in the shaded areas in the Figure. Note that condition 2 is now satisfied: all individuals are taken into account and the fairness gap will only be zero if both health curves are horizontal. However, the fairness gap also takes into account the fact that the slopes of the curves h^A and h^B are different, while we ideally would like to neutralize the effect of life style differences, and, hence, the differences in the slopes. Therefore, the fairness gap does not satisfy condition 1. Both approaches lead to the same result if the only difference between h^A and h^B is a vertical shift. This is the case of additive separability.

Note that we basically propose to use the standard apparatus of inequality measurement to the corrected health outcomes or to the individual fairness gaps: we do *not* work with concentration curves, as is the dominant procedure in the literature on socioeconomic inequalities in health. There are two reasons for this. First, the use of concentration curves is only possible if one considers inequality in one dimension (e.g. income) with a natural ordering which can be used to construct the concentration curves. Our approach allows for many legitimate and illegitimate variables. Second, even if one only considered socioeconomic inequalities we still think that there are severe limitations to the use of concentration curves. In the following sections, we will come back to these issues more explicitly.

3   A sketch of a structural model

Fairness does not only relate to socioeconomic inequalities. Some variables influencing health or health care can be considered to be legitimate, others are ethically objectionable sources of differences. Moreover, as the simple example in the previous section has shown, the empirical interactions between these different variables may have a crucial influence on the inequality measurement. It is therefore necessary to have a structural model in mind when thinking about specific measures of unfair inequality. Starting from a structural

11

model makes it possible to close part of the gap between the large health economics literature on explaining health and health care differences and the normative literature on unfair inequalities. In this section we will organize our thinking by introducing such a structural model in very general terms. Of course, many simplifications are needed if one wants to apply the model to real empirical data. We will return to this issue in section 5.

We state that the health level h_i of individual i is produced by a health technology $H(.)$, which can be written as follows

$$h_i = H(m_i, c_i, e_i, \varepsilon_i, o_i, s_i) \tag{2}$$

where m_i is a vector of medical consumption (e.g. the number of GP visits, the number of specialist visits, pharmaceutical consumption and so on), c_i is a vector of consumption goods, including life style goods (smoking, drinking, physical activities), o_i is a vector of job characteristics (including leisure), and s_i is social background. We therefore leave open the possibility that, in addition to life style and job characteristics, there is also a direct effect of social background on health. Further crucial variables are e_i, the genetically determined health endowment, and ε_i, which is a (stochastic) health shock. The health technology, as described in (2), is determined mainly by biological considerations and is objectively given to the individual. However, individual behavior has an influence on health through the choices of m_i, c_i and o_i.

Labour income y_i is endogenous and is determined through a mixture of endowments and choice variables:

$$y_i = Y(c_i, o_i, h_i, a_i, s_i) \tag{3}$$

where a_i is the innate productive capacity of the individual, for which she cannot be held responsible. Earnings capacities are also influenced by the present health status of the individual. Individual choices of leisure and job characteristics o_i will endogenously determine gross labour income y_i. Note that we again include social background explicitly: this is meant to capture not only the effect of discriminatory practices by employers, but also the differences in the quality of the social networks that are available to various individuals and that will influence their search behavior and their final outcomes.

To model the individual choices of m_i, c_i and o_i, we assume that individuals maximize a utility function $U_i(m_i, c_i, o_i, h_i)$.[7] We thus assume that m_i enters the utility function directly. With this we want to capture the idea that individuals may have specific tastes about medical care consumption, for which they can (perhaps) be held responsible. They also care for their health. However, the health production function (2) is not perfectly known to individuals and there may also be differences between social groups in this respect. Representing the information available to individual i by I_i, we explicitly define her "perceived" health production function as

$$h_i = \widehat{H}(m_i, c_i, e_i, \varepsilon_i, o_i, s_i; I_i) \tag{4}$$

and we assume that choices are based on this perceived health production function.

Individuals maximize their utility under a budget constraint, which we write as

$$pc_i + B(m_i, r_i) = y_i - T(y_i, c_i) - \rho(r_i, e_i) \tag{5}$$

where y_i is income, $T(y_i, c_i)$ gives taxes paid (or transfers received) as a function of labour income and consumption, and p are the consumption prices. To arrive at a general description of the health financing constraints, we introduce two additional functions. The first $(B(m_i, r_i))$ gives the out-of-pocket payments. These are dependent on the level (and structure) of medical care consumption and on the degree of supplementary insurance coverage r_i. The form of the function $B(.)$ is determined by the health care system in which the individual lives. In a National Health Service-system where all expenditures are taken care of by the government and there are no co-payments or deductibles, the value of $B(m_i, r_i)$ can be zero. If individuals take supplementary insurance, they will have to pay a premium $\rho(r_i, e_i)$: the premium amount will depend on the degree of insurance coverage r_i, and on a private insurance market with premium differentiation it will also depend on the genetic health endowment e_i. Buying supplementary insurance is an individual decision, taken at an earlier stage.[8]

---

[7] This (ordinal) utility function is to be interpreted as the representation of a preference ordering $\Re_i$ for individual i. In section 6 we will argue that we do not assume that subjective utility is cardinally measurable or interpersonally comparable.

[8] This assumption of a two stage decision-making process is only made for convenience. However, it

In addition, while individuals have some freedom in their choice of health care consumption, it is generally accepted that they are restricted by the decisions taken by health care professionals. The behavior of the providers will be influenced by the characteristics of the health care system, more specifically by the way in which they are remunerated. Moreover, in many countries there is huge interregional variation in the availability of health care services. This acts as a kind of quantity rationing constraint. We will summarize these supply side influences by saying that an individual i can only choose his medical consumption vector m_i from a restricted choice set M, the shape of which is determined by supply side variables z_i, by his health endowments e_i and the stochastic shock ε_i, by his level of supplementary insurance coverage and by his social background:

$$m_i \in M(z_i, e_i, \varepsilon_i, r_i, s_i) \tag{6}$$

Note that eq. (6) captures restrictions on choice determined by the supply side, not the choice behavior of the individuals themselves. The influence of social background and supplementary insurance coverage in (6) refers to the situation in which providers differentiate their behavior according to the social background of the individual or according to whether they have supplementary insurance or not (e.g. in the situation where they can raise additional supplementary fees from patients with a supplementary insurance). In the extreme case where individual patients had no freedom of choice at all, the set M reduces to a singleton, and medical care consumption is fully determined by the providers.

We can now summarize our stylized model of individual behavior as follows. In a first stage, individuals choose to take supplementary insurance or not. This decision will be influenced by their health and income prospects (affected by e_i and a_i respectively) and by their time and risk preferences R_i. Moreover, we know that the information about insurance opportunities is unequally distributed over the population and that some social groups might find it more difficult to buy insurance than others. Rather than modelling this decision process explicitly, we summarize it by the following reduced form specifica-

seemed important to us to introduce supplementary insurance into the model, because it also plays an important role in recent attempts to explain inequity in delivery (see e.g. Jones et al., 2006).

tion:

$$r_i = \Theta^R(R_i, e_i, a_i, s_i) \tag{7}$$

In a second stage individuals decide about (m_i, o_i, c_i) by maximizing individual utility $U_i(m_i, c_i, o_i, h_i)$ under the budget constraint (5), the information constraint (4), and the supply-side constraint (6). The resulting behavior can be expressed as a function of the exogenous individual characteristics as follows:

$$m_i = m(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{8}$$

$$o_i = o(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{9}$$

$$c_i = c(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{10}$$

The values of health, of income and of achieved welfare are endogenously determined. Introducing the decision variables in the utility function, in eq. (2) and in eq. (3) we get the following reduced form expressions:

$$h_i = H^R(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{11}$$

$$y_i = Y^R(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{12}$$

$$u_i = U^R(s_i, a_i, e_i, \varepsilon_i, z_i, I_i, R_i, U_i) \tag{13}$$

Hidden behind these reduced form expressions (8)-(13) are market characteristics (defining p and the shape of the functions $\rho(.)$ in (5) and $Y(.)$ in (3)) and policy variables (the shape of the functions $B(.)$ and $T(.)$ in the budget constraint (5)). The supply effects on medical care consumption in (6) can be seen as resulting from a mixture of market forces and policy decisions. Evaluating the inequality in medical care consumption (8), in health (11) and in welfare (13) then indeed boils down to an evaluation of the whole social structure. The advantage of the reduced form expressions is that they neatly distinguish different exogenously given characteristics of the individuals. At the same time, considering the whole structural model clearly shows where and how these characteristics enter the analysis. This is of crucial importance to decide whether these characteristics should be treated as legitimate or illegitimate sources of interpersonal differences in health and health care.

In the following section, we will start from this structural model to give a more explicit content to the notions of direct unfairness and of the fairness gap in the context of health and health care. To simplify the notation, we will group the exogenous characteristics in the reduced form equations in five groups: health endowments (also indicating health care needs) $N = \{e, \varepsilon\}$, social background variables $S = (a, s)$, individual preferences $P = (R, U)$, available information I and supply-side variables z.

4 Unfair inequalities in health and in health care

Unfair inequalities in health care (or "horizontal inequity" in health care delivery) relate to the distribution of m in the population. There are many different approaches to the concept (see e.g. Hurley, 2000; Wagstaff and van Doorslaer, 2000a; Williams and Cookson, 2000 for surveys). In this paper we do not want to go deeply in that conceptual and ethical debate. We want to show, however, that the most important arguments can be reinterpreted within a framework with legitimate and illegitimate causes of inequalities. The same is true for inequality in health.

We will first analyze unfair inequalities in health care under the assumption that m can be treated as a scalar variable. We will then discuss the specific questions linked to the fact that health care consumption is better seen as a vector. Finally, we discuss unfair inequalities in health.

4.1 Unfair inequalities in health care

Let us start from eq. (8) and indicate how different approaches to equity in health care delivery can be interpreted in terms of the distinction between legitimate and illegitimate sources of inequalities:

- there seems to be almost general consensus that for given N, P, I and z, differences in S should not lead to differences in health care consumption. This is the basic issue of socioeconomic inequality, which has dominated the health economics literature;

- a different but related stream of literature has focused on supply-induced inequalities. Again, there seems to be near consensus that given N, P, I and S, individuals cannot be held responsible for supply-side influences, and that differences in z should therefore not lead to differences in health care consumption. If we interpret z in terms of the regional distribution of health care resources, we recover here the problem of regional inequities in health care;

- for given P, S, I and z, differences in N lead to legitimate (and desirable) differences. Note that we are considering here only the issue of horizontal equity and ignore vertical equity, which would consist in checking whether the observed differences in health care consumption are an ethically appropriate reaction to differences in health care needs. We take for granted the existing relationship between health care consumption and N[9];

- there is more discussion about the relevance of differences in P and I. This question indeed brings us right into the debate between proponents of equality of use, equality of access, and equality of informed access. In the past, there have been some heated exchanges of opinions in this respect. As an example, after the initial paper by Wagstaff et al. (1991) about equity in delivery, there was a sharp reply (Mooney et al., 1991) pointing to the importance of equality of access rather than use.[10] Proponents of equality of access emphasized that individuals can (and should) be held responsible for their preferences and their information. Proponents of equality of use explicitly rejected responsibility for both preferences and information. An intermediate position (that of equality of informed access) holds individuals responsible for their personal choices, if these are based on good information. We can summarize these different views in our notation. Emphasizing equality of access means that given N, S, and z, differences in P and I lead to unproblematic (legitimate) differ-

---

[9]See Sutton (2002) for an explicit analysis of vertical inequity.

[10]Another example is Goddard and Smith (2001). These authors discuss the issue of equality of access in a broad setting and with a general approach which is close to ours (including the specification of a structural model and the distinction between legitimate and illegitimate causes of differences).

ences. Emphasizing equality of use boils down to including both P and I among the illegitimate sources of variation. Equality of informed access holds people responsible for P but not for I. In fact, the methodology we propose can accommodate all three approaches and allows to compare the results, i.e. to identify the part of inequality which is due to preference variation and to variation in information.

In this paper we only want to describe a specific methodology to implement the distinction between legitimate and illegitimate reasons for differences, without taking a specific position on where exactly to draw the dividing line. For the sake of the argument, however, we will further on treat N and P as legitimate sources of differences, and S, I and z as illegitimate sources of differences.

We can now apply the main concepts from section 2 to eq. (8). To compute the degree of **direct unfairness**, we first remove legitimate differences (due to differences in N and P) by fixing these to a reference value. We then apply the traditional apparatus of (absolute) Lorenz curves and (absolute) inequality measurement to the vector $\widetilde{m_i} = m(S, \widetilde{N}, z, I, \widetilde{P})$, i.e. the variation in medical expenditures which remains after all legitimate sources of variation are kept constant.[11] To compute the **fairness gap**, we first define an "ideal" solution, i.e. a situation in which all the illegitimate sources of variation have been removed, as $m_i^* = m(S^*, N, z^*, I^*, P)$. We then apply the traditional apparatus of (absolute) Lorenz curves and (absolute) inequality measurement to the vector $(m_i - m_i^*)$.

As shown before, these two approaches will in general yield different results, unless the health care consumption function is additively separable in the legitimate and illegitimate variables. Which of the two is then preferable? Remember first that direct unfairness does not satisfy the compensation condition 2 - it may get a zero value even if there are illegitimate differences left. This seems to be a decisive weakness. The fairness gap approach does satisfy the compensation condition. However, it does not satisfy condition

---

[11]The results in general will depend on the choice of the reference values ($\widetilde{N}$ and $\widetilde{P}$ for direct unfairness, S^*, I^* and z^* for the fairness gap). It is natural to choose average values in the population as reference, but other possibilities may also be considered, e.g. the choice of an "efficient" level of z for the reference value z^*.
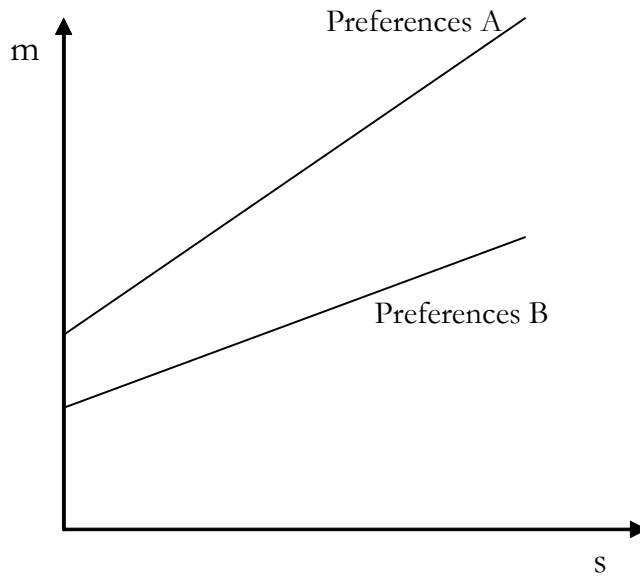
18

Figure 2: Socio-economic background and preference differences

1 and may also take up inequalities due to needs and preferences, if these interact with socioeconomic background and information and supply side variables. The problem was sketched in Figure 1. In order to avoid confusion, that Figure is repeated here for the context of health care consumption as Figures 2 and 3.

Look first at Figure 2. In that figure we distinguish two preference groups. The fairness gap will take up into the inequality measure that the relationship between social background and health care consumption is stronger for preference group A than for preference group B. This goes against condition 1 (no influence of legitimate differences), if individuals are held responsible for their own preferences (after controlling for all the other variables that appear in the structural model). Or, looked at from another angle: with a uniform distribution of individuals, the fairness gap suggests that the contribution of preference group A to overall unfair inequality is larger than the contribution of preference group B, although these differences are only due to preference variables for which individuals are held responsible. But look now at Figure 3. This is exactly the same, except that we now interpret the two curves as corresponding to different needs groups. Again, the fairness gap will take into account that the relationship between social
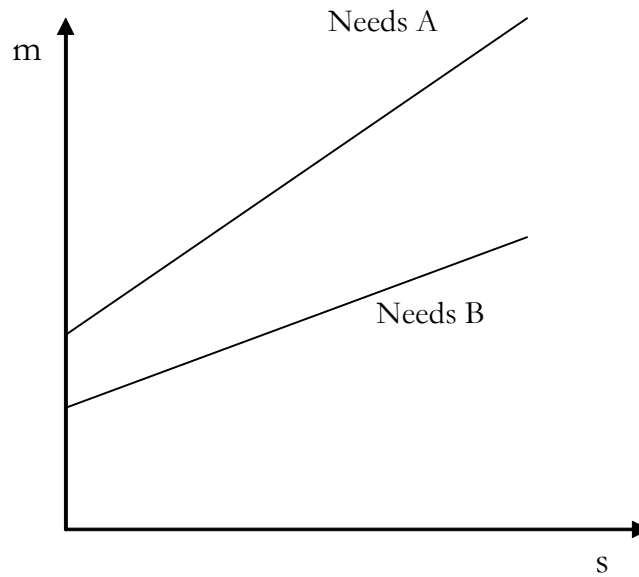
Figure 3: Socio-economic background and differences in needs

background and health care consumption is stronger for one needs group than for the other and will therefore take up differences in consumption reflecting differences in needs. However, in this case the ethical implications are quite different, because individuals are not responsible for the needs curve which is relevant to them - needs reflect differences in health endowments or health shocks. It therefore makes perfect sense to claim that the interaction between the effect of needs and of socioeconomic background should be taken up in the measure of unfair inequality. There seem to be good normative reasons to prefer the fairness gap to the direct unfairness approach.

4.2 How to aggregate use?

In the previous subsection we provisionally adopted the assumption that health care consumption can be seen as a scalar variable. We therefore sidestepped the aggregation problem: how to aggregate such diverse items as visits to a general practitioner and regular hospital dialysis? One approach could be to calculate unfair inequality for each of the items separately. The Ecuity-project (see, e.g. van Doorslaer et al., 2000) has shown

20

that this may lead to intriguing results: one of these is the rather common finding in many countries of "pro-poor" inequity for general practitioners and "pro-rich" inequity for specialists. However, this disaggregated approach implies a very strict interpretation of equality of use, which seems unreasonable given that there are obvious substitution possibilities between the various health care items. Moreover, it does not allow to arrive at an overall evaluation of equity. For this purpose, we have to aggregate the different items in one way or another.[12]

Given that we are focusing on the fair use of *resources*, the most natural aggregator is the global opportunity cost of the health care resources used by the individual.[13] Although this approach is natural, it still is useful to interpret it in terms of the distinction between legitimate and illegitimate inequality. Indeed, the use of this aggregator implies that there are no unfair inequalities in health care left if there are no unfair inequalities in overall health care resources used. Individuals are then held responsible for the allocation of the total resources over the different items. A relatively larger use of emergency services in the hospital can (from the point of view of equity) compensate for a relatively smaller number of preventive visits to a doctor.

To see the consequences of this, let us go back to the structural model and consider the different variables influencing health care consumption m. If differences in the composition of m are motivated by differences in the "objective" health production function (2), they may be desirable. If they reflect differences in preferences, we are back in the discussion about equality of use and equality of access that was introduced before. If one holds individuals responsible for differences in P, the resulting differences in the composition of m are unproblematic. However, different choices by the patients may also reflect differences in I and therefore in the perceived health technology (4), or differences in the supply side mechanism (6). As argued already before, it would be difficult to state that individual patients are responsible for supply-side features (e.g. the density of specialists

---

[12]A possible alternative, which to the best of our knowledge has not yet really been explored in the literature, would be the construction of multidimensional inequality measures.

[13]In this spirit, the Ecuity-project (van Doorslaer et al., 1992, 2000) has focused on imputed expenditures, computed as volumes times average imputed cost.

in low-income versus high-income areas). There can be more debate about responsibility for information, but it seems reasonable to accept that informational differences are due, at least partly, to factors that are beyond individual responsibility.

All this suggests that we should care about inequalities in the composition of m, even with no inequality in the total amount of resources used, if these differences in composition lead to differences in the quality of treatment. If people get treatment of different quality because of factors for which they cannot be held responsible, it is difficult to maintain that there would be no "unfair inequality in health care". We therefore propose an alternative aggregator which takes into account the effect of treatment on health. The basic idea is introduced in Figure 4. Suppose we consider two health care items: GP visits and visits to a specialist, and we start in a hypothetical situation where there would be equality of use between group A (say, the rich) and group B (say, the poor), if equality of use is interpreted in terms of the global opportunity cost of the resources of health care for A and for B. This global opportunity cost is given by the "budget" line CC'. Suppose also, for the sake of the argument, that the *objective* health production technology is the same for A and for B and is represented by the iso-health curves in Figure 4. Group A consumes more specialist services and therefore reaches a higher iso-health curve than group B.[14] Or, formulated in another way, the health care resources are used in a less efficient way by B if efficiency is defined in terms of health "production". Let us assume that the "choice" of B reflects poorer information (or supply-side discrimination) and that B is not held responsible for her lower efficiency. An attractive possibility is then to define for each individual the minimum health care budget, needed to bring her at the health level she has reached with her actual health care consumption. In Figure 4 this is the budget represented by DD' for the rich and the budget represented by EE' for the poor. We can then apply the whole measurement apparatus described in the previous section to these "minimal efficient budgets".

The concept of "minimal efficient budgets" may be hard to implement because its measurement for any particular individual should theoretically depend on this individ-

---

[14]Note that the figure does not represent a traditional budget line and traditional indifference curves.
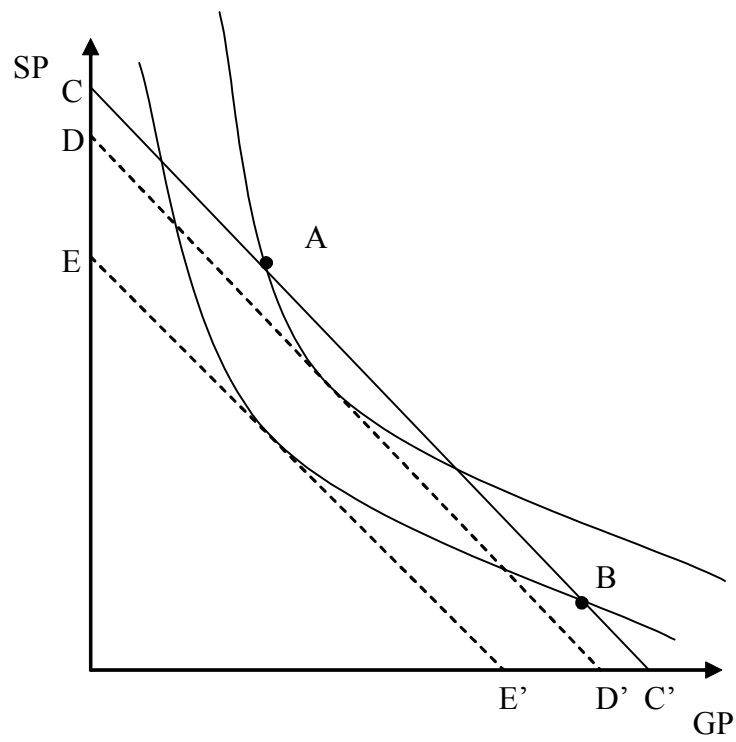
Figure 4: The minimal efficient budget

ual's characteristics, as can be seen from eq. (2). With this aggregator, equation (8) may also be hard to estimate because individual preferences may affect the value of the minimal efficient budget not only through quantitative demand behavior, but also through qualitative demand influencing the composition of m.

The basic idea underlying this aggregator is that what ultimately matters is the impact of medical care on health, not the expenditure as such. To take an extreme example, a patient receiving useless treatments is not viewed as getting a favor in the metric of minimal efficient budgets. In this light, our proposal is perfectly in line with the idea of Culyer and Wagstaff (1993) that what basically matters is health. This immediately brings us to a potential point of criticism. The shift from "budgets" to "minimal efficient budgets" moves us away from the basic intuitions about "equity in delivery" and in the direction of equal health. To some extent, it erodes the relevance of looking at equity in delivery. Unfair inequalities in health are the topic of the next subsection.

4.3   Unfair inequalities in health

The most convincing justification to strive for equity in health care delivery seems to be a concern for equity in health, combined with the belief that health care contributes to a better health. A separate treatment of equity in delivery may be useful, because it allows for an evaluation of a specific domain of government policy. However, removing unfair inequalities in health care is not sufficient to remove unfair inequalities in health. It therefore remains necessary to calculate unfair inequalities in health, too.

Unfair inequalities in health can be analyzed starting from the reduced form equation (11). Again different ethical conceptions will lead to different views about where to draw the boundary line between legitimate and illegitimate variables in (11):

- there is hardly any debate about the idea that given N, I and P, differences in z and S should not lead to differences in health status. It is generally accepted that these differences are induced by social arrangements or by policy, and that they are therefore avoidable and hence unfair. The literature about socioeconomic inequalities in health (i.e. inequalities related to differences in S) is especially impressive.

24

- as argued before, there are also good reasons to include information I among the illegitimate causes of health differences.

- there is much more debate about the classification of the preference variables P. The debate is somewhat confused because it often starts from the empirical finding that a large part of socioeconomic inequalities in health is due to differences in lifestyle (smoking, alcohol consumption, nutritional equilibrium) and therefore focuses upon the statistical correlation between P and S. We will come back to this issue of statistical correlation in the next section. The question we raise here is different: do differences in P lead to legitimate health differences under the condition that S, z, I and N are fixed? To make the question more specific: if within one social group there are differences in health which can be traced back completely to differences in smoking behavior, are then these differences in health unfair?

- in our view, there is surprisingly little debate about the question whether differences in N lead to "unfair" inequalities in health, conditional on given values for P, z, I and S. In fact, it seems to be almost generally accepted that the variables in N are biological characteristics (genetic endowments or stochastic shocks) which are not caused by the social structure, and cannot even be modified by social policies - and are therefore not unfair. It is argued, e.g., that differences in health related to age cannot be unfair because they are a fact of life, or that innate handicaps do not lead to unfair health differences because they are unavoidable.[15] We agree that it makes sense to focus on avoidable differences for policy analysis. We would argue, however, that "unfair" does not necessarily coincide with "avoidable" and that it is also useful to calculate a broad notion of unfairness, which includes differences in unavoidable factors and even differences in unavoidable consequences. Suppose an unavoidable natural disaster hits a given region. Are the induced welfare losses to be called fair because the natural disaster was an unavoidable cause? Are the welfare losses which cannot be remedied to be called fair because they are unavoidable

---

[15] In the same spirit, Gravelle (2003) makes the distinction between what he calls "policy irrelevant" or standardizing variables and "policy-relevant" variables.

consequences of the disaster? It is sometimes objected against a broad notion of unfairness that reducing the unfairness due to inequality in health would require lowering the health of the healthy. This is of course not desirable. Yet, reducing the unfairness due to health inequality does not necessarily mean that one has to lower the health level of the healthy: one can also go beyond the health domain, and compensate the lower health level of sick persons by a better result on other dimensions (e.g. income). In order to evaluate or to stimulate these policies a broad measure of unfair health inequality is useful. Moreover, what are avoidable or policy-relevant differences in health is not always clear and will change over time. Some diseases which were incurable a few decades ago can now well be treated. Would one not be willing to draw the conclusion that this implies that unfair inequality in health has decreased over time? Similarly, should we not consider improvements in the treatment of ageing as reducing undesirable inequalities between individuals of different ages? We therefore need a measure which is able to capture these changes and these differences.[16]

While we therefore think that it is preferable to include N among the illegitimate rather than among the legitimate sources of inequality, the methods we propose allow for both options. And the same is true for different opinions about P. Let us again for the sake of the argument accept that differences in S, N, I and z lead to unfair inequalities in health, while health differences following from differences in P are no cause of concern.

We then can apply our basic concepts in the health sphere. To compute the degree of **direct unfairness of the health distribution**, we remove legitimate differences (due to differences in P) by fixing these to a reference value, and we apply the traditional apparatus of (absolute) Lorenz curves and (absolute) inequality measurement to the vector $\widetilde{h}_i = H^R(S, N, z, I, \widetilde{P})$, i.e. the variation in health which remains after preferences are kept constant. To compute the **health fairness gap**, we first define an "ideal" situation, i.e. a situation in which all the illegitimate sources of variation have been removed, as

---

[16] A particularly tricky example of differences in genetic endowments is that of gender differences in health and mortality. See, e.g., Tsuchiya and Williams (2005).

$h_i^* = H^R(S^*, N^*, z^*, I^*, P)$, and we apply the traditional apparatus of (absolute) Lorenz curves and (absolute) inequality measurement to the vector $(h_i - h_i^*)$. The two approaches will yield the same results if the function $H^R(.)$ is additively separable in the legitimate and the illegitimate variables. In general, the health fairness gap will satisfy the compensation condition, but may include some effects of P. The direct unfairness approach does not satisfy the compensation condition, i.e. it can yield a value of zero even if there are still health differences between individuals with identical preferences.

Our approach bears some similarity with the proposal made by Bommier and Stecklov (2002). These authors write that in an equitable situation, "we should expect two individuals with equivalent health endowments to reach the same health level, regardless of their socioeconomic status" (p. 503). This is basically a reformulation of our compensation condition in the case where inequalities due to health endowments are unavoidable and therefore do not lead to unfairness (a position Bommier and Stecklov explicitly support). They implement this ethical starting point through the requirement that "the expectation of actual health given social background, $E(h_i|S_i)$, should be equal to the average health level in the society, $E(h_i)$, and therefore independent of income" (p. 505),[17] and more specifically propose to draw the Lorenz curve for this variable $E(h_i|S_i)$. Their focus on social background (and on social background only) implies that differences in inequality following from differences in P, N, I and z are all considered to be legitimate. If these variables are not explicitly included in the estimations, the measure of socioeconomic inequality will be biased, as we will argue in the next sections. Moreover, even allowing for this, drawing the Lorenz curve for $E(h_i|S_i)$ is not identical to drawing the Lorenz curve for $\widetilde{h}_i = H^R(S, \widetilde{N}, \widetilde{z}, \widetilde{I}, \widetilde{P})$. Yet, the inspiration of Bommier and Stecklov is very similar to our idea of direct unfairness: it is easily seen that the Bommier-Stecklov-approach does not satisfy the compensation condition, despite their claim to the contrary.

---

[17]Bommier and Stecklov (2002) propose to work with $E(h_i|y_i)$, but we substituted social background for (endogenous) income in line with the approach followed in this paper.

5 Towards empirical application

Empirical application of the concepts we propose is very easy in principle. The first step is the estimation of a structural model, as described in section 3. The best available econometric techniques can be used, because increasing the complexity of the empirical model does not increase to the same extent the complexity of the normative exercise. What finally matters for the latter are the reduced form expressions (8) and (11) - and the hypothetical values used in the calculation of direct unfairness or the fairness gap can be calculated with simulation techniques if necessary. In fact, we believe that the clear distinction between on the one hand the descriptive (or explanatory) step of estimating the best econometric model which is possible with the available data, and on the other hand the normative step of calculating unfair inequalities is an important advantage of our approach. It has become usual practice in the literature to compute first an index of socioeconomic inequality and then in a second step to decompose this index in order to get a better insight into the causes of the inequality (see, e.g., Wagstaff et al., 2003; van Doorslaer et al., 2004; van Doorslaer and Koolman, 2004). Linear approximations are needed to keep this decomposition feasible. Moreover, the decomposition approach focuses on a reduced form without the possibility of deriving additional insights from a theoretically acceptable structural model. In a certain sense, such an approach clips the wings of the econometricians to fit them into the normative straitjacket. Nothing of this kind is needed in our approach. The effect on unfair inequality of policy changes or hypothetical changes in the environment can be calculated through simulation with the full-fledged structural model.

While the empirical economists are completely free to estimate the best structural model possible, there are important normative questions which should be considered when going from the estimated model to the calculation of inequality. We will first discuss some of these questions. We will then show the link between our approach and the traditional methods of direct and indirect standardization.[18]

---

[18] The discussion parallels the application to risk adjustment in Schokkaert and Van de Voorde (2004, 2007).

5.1 Normative choices

Although the normative analysis can in principle start from a sophisticated empirical analysis, for our purposes it is sufficient to illustrate the issues with a simple linear example. The points we raise are equally relevant for the topics of equity in health care and equity in health, but we will illustrate them for health care. Neglecting measurement issues and taking all the variables, including m, as scalars for the sake of simplicity, we consider a situation in which eq. (8) takes the following simple form:

$$m_i = \alpha + \beta S_i + \gamma z_i + \delta N_i + \eta S_i N_i + u_i \tag{14}$$

in which u_i is a disturbance term. If $\eta = 0$, eq. (14) is additively separable.

To apply the concepts from the previous section, one first has to decide for each variable in the model whether it is a legitimate or an illegitimate source of differences. As described before, different decisions will reflect different views about equity in health care. Let us in line with what was done before assume that S and z are illegitimate sources of inequality and that N is a legitimate source of inequality. There then remains the crucial problem of interpreting the disturbance term, however. A natural interpretation would be to say that u_i reflects individual idiosyncrasies and therefore mainly individual preferences P and information I. In this interpretation, theories of equality of access would include the disturbance in the list of legitimate sources of inequality, while theories of equality of use would still see it as an illegitimate cause of differences. An approach in terms of inequality of informed access would take an intermediate position. This choice makes a crucial difference. In the equality-of-access interpretation direct unfairness would measure the inequality in the vector $(\beta S_i + \gamma z_i + \eta S_i \widetilde{N})$, while the fairness gap would consider inequality in the vector $(\beta S_i + \gamma z_i + \eta N_i (S_i - S^*))$.[19] In the equality-of-use interpretation u_i should be added to both these expressions, which can then be rewritten as $(m_i - \delta N_i + \eta S_i (\widetilde{N} - N_i))$ and $(m_i - N_i (\delta + \eta S^*))$ respectively.

---

[19]Note that we omitted all constant terms from these expressions. Since we follow in this paper an absolute approach to inequality measurement, these constants are irrelevant because of the translation invariance.

The issue gets even more complicated, because in all real-world applications the disturbance term will not only reflect individual preference and informational idiosyncrasies but also the effects of measurement errors and omitted variables - and the errors and omissions will relate to both legitimate and illegitimate sources of inequalities.[20] In fact, the safest procedure seems to be to calculate unfair inequalities in both cases, i.e. with and without u_i included. The results could then be interpreted as giving an upper and a lower boundary to unfair inequality respectively. In any case the important message is that it is necessary to think explicitly about the interpretation of the disturbance term.[21]

A second general question refers to the choice of the reference values in the calculation of direct unfairness and of the fairness gap. A natural choice is to choose the mean values in the sample, but this choice is not innocuous and will influence the results. It seems advisable to perform sensitivity analysis for different reference values.[22]

A third general point is obvious. In the approach we propose it is important to estimate the best structural model possible and it would make no sense to deliberately omit some variables from the estimated model, because this could induce omitted variables bias. Suppose for instance that one is mainly interested in socioeconomic inequality, and that one would omit from eq. (14) the variable z related to supply-side differences. If there is a correlation between z and S (as may be expected) the estimate of β - and therefore the estimates of unfair inequality will be biased. The problem of omitted variables bias is even more obvious if one estimated eq. (14) with the variable S omitted, as used to

---

[20]An obvious example of a legitimate source of differences entering into the disturbance terms would be the stochastic health shocks ε_i.

[21]In this respect model (14) may be misleadingly simple, in that the disturbance term is additively separable from all the explanatory variables in the model. As soon as one turns to more elaborate statistical specifications, this is no longer necessarily true - and the conflict between the two conditions from section 2 will involve the disturbance term too.

[22]Gravelle (2003) discusses the case of what he calls "essential non-linearity". This boils down to the absence of additive separability. He also advocates the use of mean values, and indicates that the results depend on this choice. In the context of income distribution, Luttens and Van de gaer (2007) show some interesting theoretical results about the consequences of choosing different reference values. Some of their results are dependent on the specific model they use, however. More theoretical work along these lines would certainly be useful.

be the case in the early work on equity in delivery.[23] We will come back to this when we discuss the relationship with direct and indirect standardization.

Correlation between legitimate and illegitimate variables raises not only statistical issues, however (see, e.g. the analysis of Roemer, 1998, 2002). Let us consider the example of a life style variable (e.g. smoking). If people are not to be held responsible for their socioeconomic background, can they then be held responsible for their smoking behavior, if this behavior turns out to be very highly correlated with (and therefore perhaps caused by) their social background? In our opinion, the only adequate way to tackle this difficult issue is the construction of good structural models of behavior, which in principle should create the possibility to disentangle the social class and the pure lifestyle effect. This issue is related to the deeper question that one might be willing to argue that in some cases individuals are "partly" responsible for their health or health care outcomes. This would make it difficult to interpret the resulting differences unequivocally as either legitimate or illegitimate. We think that such intuitions of partial responsibility reflect the fact that some variables work through different channels or result from the interplay of different underlying mechanisms (some of which are unobservable). In our view, disentangling the different structural mechanisms is a necessary prerequisite to evaluate in a meaningful way the degree to which individuals can be held responsible. This is precisely one of the main reasons to build first a structural model rather than immediately focusing on the reduced form.

5.2   Direct and indirect standardization

Both in the literature on equity in delivery (Wagstaff and van Doorslaer, 2000a, 2000b) and on equity in health (Gravelle, 2003) an explicit link has been made between different calculation methods and the methods of direct and indirect standardization. While Wagstaff and van Doorslaer (2000b) argue in favour of the indirect standardization method, Gravelle (2003) takes the opposite position and claims that this method leads to inconsistent

---

[23]The recent work takes into account the problem of omitted variables bias - see, e.g., the discussion in van Doorslaer et al. (2004) and Van Ourti (2004).

estimates. Our approach offers some additional insights into this issue. Note that all these authors work with concentration instead of inequality indices: we come back to that issue in the next section, but it is not essential for our discussion at this stage.

First, the direct standardization method is exactly analogous to what we have called direct unfairness, if one takes the mean value of the legitimate variables as the reference value to calculate $\widetilde{m}$ or $\widetilde{h}$.[24]

Second, there is also a clear link between the method of indirect standardization and the fairness gap, but here we have to be careful. The traditional approach to indirect standardization (see, e.g., van Doorslaer et al., 2000) is first to run a regression with the legitimate (or "standardizing") variables only, i.e. estimate a model

$$m_i = \rho + \zeta N_i + e_i \tag{15}$$

and then to compute the inequality in $(m_i - \zeta N_i)$.[25] If the true model is (14), estimating eq. (15) will lead to a biased estimate of the effect of the needs variable on medical expenditure, if there is non-zero correlation between on the one hand N and on the other hand z and S. This is the basic insight underlying the statements of Gravelle (2003) about the inconsistency of indirect standardization. In some cases the consequences of this are quite dramatic. Suppose that health endowments are lower in regions with a lower supply of doctors and that a lower supply of doctors leads to lower expenditures: then the estimate of ζ in (15) will be an underestimate of the true value δ in eq. (14). Suppose now that there is also a positive correlation between health endowments and socioeconomic background. Then the simple method of indirect standardization will in

---

[24] Wagstaff and van Doorslaer (2000b) claim that the method of direct standardization requires the use of grouped data. As Gravelle (2003) has already argued, this claim is mistaken and there is no problem to do direct standardization with individual data. In fact, a model with separate utilisation equations per income group, can easily be rewritten as one general model with a full set of cross-effects between income and the needs variables included.

[25] We again leave out all the constant terms, which are irrelevant in our absolute approach to measurement. Note also that in any application all the theoretical coefficients in these expressions have to be replaced with estimates. Moreover, remember that in some non-linear specifications it is not always the case that $E(e_i) = 0$.

general lead to an underestimation of the "justified" health care expenditures of those from lower socioeconomic background - and therefore also to an underestimation of unfair socioeconomic inequality.

However, it is obvious that one can easily solve this problem by estimating the full model (14) (see, e.g. van Doorslaer et al., 2004). The method of the fairness gap, i.e. putting all the illegitimate variables equal to a reference value and then calculating the inequality in $(m_i - m_i^*)$, is completely analogous to the basic idea of indirect standardization but takes care of the omitted variables bias that would be induced by estimating (15).

With this extension of the idea of indirect standardization, the choice between direct and indirect standardization boils down to the choice between calculating direct unfairness and the fairness gap (or between conditional equality and the egalitarian-equivalent solution). This choice can now be based on normative considerations. First, if there is additive separability between the legitimate and illegitimate causes of differences in health care consumption or in health, direct and indirect standardization will lead to the same results.[26] Second, and more importantly, in the general case the choice between direct and indirect standardization will depend on the relative importance given to the "no influence of legitimate differences condition" 1 and the "compensation condition" 2. Researchers who (like us) give a greater weight to the compensation condition should prefer the fairness gap and, hence, the (extended) method of indirect standardization.

Let us repeat that we consider the explicit distinction between empirical and normative questions one of the main advantages of our approach. Statistical and econometric considerations should play the crucial role in the choice of the explanatory model. Once this model has been estimated, however, the further choice of inequality concepts should be based on normative considerations. In this respect, our approach differs from previous papers such as Wagstaff and van Doorslaer (2000b), Gravelle (2003), van Doorslaer et al. (2004) and Van Ourti (2004), which all focus on statistical issues.

---

[26]Remember that the estimation of separate equations per income group, as is common in the direct standardization approach, implies that there is no additive separability.

6 Socioeconomic inequalities and welfare inequality

Until now, we worked in a setting where unfair inequality can reflect a whole range of illegitimate causes of differences, and not only socioeconomic background. Let us now consider in more detail the latter. We first show how socioeconomic inequality is a special case of the approach in the previous sections and we discuss the use of concentration curves instead of Lorenz curves. We then suggest one possible way to integrate health inequalities in a broader view on welfare inequalities.

6.1 Socioeconomic inequalities, Lorenz curves and concentration curves

Suppose one wants to focus exclusively on socioeconomic inequalities. In one extreme interpretation, this would mean that one takes all the other causes of differences (needs, information, preferences and supply side restrictions) as *legitimate* sources of inequality. Socioeconomic inequalities are then simply a special case of our general framework, and one could easily compute direct unfairness and the fairness gap. Note that this implies a specific ethical position, some aspects of which are highly debatable: that inequalities due to supply side restrictions are not illegitimate, that one prefers equality of access over equality of use (because individuals are responsible for their preferences), that health inequalities following from differences in lifestyle are unproblematic, that differences in health endowments are unavoidable and therefore not unfair.

There is a less extreme and more pragmatic interpretation possible, however. In that interpretation it is accepted that socioeconomic inequalities are only part of total unfair inequality. Calculating socioeconomic inequalities is then the calculation of only that part - without necessarily implying that the other parts are irrelevant (Wagstaff and van Doorslaer, 2004). This is a reasonable position, but one should not forget the points raised in the previous section. If one does not work within a fully specified structural model, the results will reflect a mixture of different considerations, including the statistical correlations between different variables that may depend on the sample used. Moreover,

depending on the method, the disturbance terms will sometimes be interpreted as part of socioeconomic inequality, sometimes not. To really understand what one has measured, it is necessary to reason within the full framework that has been sketched in this paper.

There is another important difference between the traditional approach to socioeconomic inequalities and our proposal. We propose to use (absolute) Lorenz curves, while most health economic studies work with the concentration curve. Adapted to our notation, this latter approach means that one cumulates the values of $\widetilde{h}_i$ ($\widetilde{m}_i$) or $(h_i - h_i^*)$ $((m_i - m_i^*))$ not from small to large, but according to S_i. Of course, if the rankings in terms of $\widetilde{m}_i$ (or the other variables) and in terms of S_i are the same, then the Lorenz curve and the concentration curve coincide.[27] The move to the concentration curve does only make a difference if the rankings are not the same. Should we then use it? Some recent papers on health inequality have argued convincingly that we should not. Fleurbaey (2006b) shows that interpreting a higher generalized concentration curve to yield a higher level of social welfare means that it is always good to reduce health in higher social classes in order to increase it in lower classes. Bleichrodt and van Doorslaer (2006) have characterized the concentration index with (among other axioms) what they call the "principle of income-related health transfers", conveying the same idea: it requires that transferring health from someone who is better off in terms of socioeconomic status to someone who is worse-off does not lead to a reduction in social welfare, provided the transfer does not change the ranking of the individuals in terms of socioeconomic status (pp. 954-955). This principle is hard to accept. In fact, they note that it is "more acceptable the stronger the correlation between health and other attributes such as income" (p. 955). This is a paradoxical conclusion, because it basically states that looking at concentration curves instead of Lorenz curves is more acceptable if the concentration curve looks more like the Lorenz curve. Then why not work always with the Lorenz curve?[28]

---

[27]Bommier and Stecklov (2002) also mention that the Lorenz curve and the concentration curve in their approach coincide, if the ranking in terms of $E(h_i|S_i)$ and the ranking in terms of socio-economic background are identical.

[28]An alternative proposed by Erreygers (2006) is the income concentration curve in which individuals are ranked by health status. As argued by Erreygers, this curve has similar properties as the standard

The basic reason for the popularity of the concentration curve and the corresponding indices seems to be the intuition of "cumulative deprivation": a social situation is more worrying if the economically weak are in addition less healthy, or are in addition treated in an inequitable way by the health care system. We fully accept the ethical relevance of this idea of cumulative deprivation. However, the use of concentration curves is not an adequate way to do justice to it. What is really necessary is to broaden the analysis beyond the health care sector and, in addition to health, also consider other dimensions of welfare. We sketch one possible procedure to do so in the next subsection.

6.2 The healthy-equivalent consumption

As soon as one wants to combine different dimensions of welfare in an overall evaluation, one needs a kind of aggregation procedure. The traditional economic approach to this aggregation problem is to start from the utility function (13). In such an approach, one would apply the basic ideas of direct unfairness and the fairness gap to the individual utility levels. However, this implies that one accepts subjective welfare (or utility) as the ultimate criterion to evaluate social states. It is well known that this raises difficult issues of interpersonal comparability. Moreover, the ethical criticism on welfarism is now widely spread, also in health economics. On this basis, many health economists have drawn the conclusion that one should discard all preference information and leave the aggregation problem to be solved by the decision-maker. This extreme approach is partly based on the belief that rejecting welfarism implies that one can no longer use individual preferences. This belief is mistaken, however. Moreover, in a democratic society it seems hard to defend an approach which does not at all satisfy the Pareto principle in terms of (informed) individual preferences. Is it then possible to define utility functions which do not measure subjective utility although they represent preferences? One possible way to proceed has been proposed in Fleurbaey (2005, 2006b). We will simply describe it here

---

health concentration curve but is preferable when health is measured only ordinally. Nonetheless, it has a similar (symmetric) puzzling feature as the original concentration curve, i.e., it is always good to transfer income from a healthy person to a sick person.

without going into any details, but with some refinements since our structural model is richer than the basic framework adopted by Fleurbaey (2005, 2006b).[29]

Recall that individual utility is defined here as $U_i(m_i, c_i, o_i, h_i)$. In order to construct our criterion, we first compute the indirect utility function which depends on expenditure x_i instead of the consumption bundle c_i :

$$V_i(m_i, x_i, o_i, h_i) = \max \{U_i(m_i, c_i, o_i, h_i) \mid pc_i = x_i\}.$$

Consider first a subpopulation of individuals who are perfectly healthy, i.e., $h_i = h^P$ for all i.[30] One may assume that in this case their utility no longer depends on m_i, since they do not need to be treated. Equivalently, we can simply posit that $m_i = m^P$ for all i, for some suitable value of m^P. We can therefore simplify the problem and focus on their preferences over (x_i, o_i). There is a literature on the definition of social criteria for the comparison of consumption-labor allocations on the basis of fairness criteria (for a survey, see Fleurbaey 2006a). This literature implicitly assumes that the population under consideration is healthy, which nicely coincides with our problem here. It proposes a variety of criteria, which correspond to various fairness principles involving different views about responsibility for labor choices. For the sake of illustration, we will retain one of them here. It consists in computing the equivalent income that would give individuals the same utility with no labor duty (i.e., activities are free but do not pay any earnings). Let this situation of no labor duty be denoted o^P. The equivalent income x_i^* is computed as the solution to

$$V_i(m^P, x_i^*, o^P, h^P) = V_i(m^P, x_i, o_i, h^P).$$

i.e., individual i is indifferent between the bundles (m^P, x_i^*, o^P, h^P) and (m^P, x_i, o_i, h^P).

Now let us introduce differences in health and assume that we want to satisfy the Pareto principle, i.e. we want to respect individual preferences over all the dimensions

---

[29]Fleurbaey and Schokkaert (2007) discuss how the concept can be introduced in health economic evaluation studies.

[30]In defining the indirect utility function, we only introduced a budget constraint. This is a simplification of the structural model in section 3 where we also had information and supply restrictions. However, for the computation of the healthy-equivalent consumption levels, we focus only on perfectly healthy people. This makes the information and supply restrictions much less relevant.

(m_i, x_i, o_i, h_i). Assume that for every i there is a level of expenditure x_i^* such that

$$V_i(m^P, x_i^*, o^P, h^P) = U_i(m_i, c_i, o_i, h_i). \tag{16}$$

We propose to call x_i^* the "healthy-equivalent consumption" of i.[31] The Pareto principle then implies that society should be indifferent between the actual situation $[(m_1, c_1, o_1, h_1), ..., (m_n, c_n, o_n, h_n)]$ and the hypothetical situation $[(m^P, x_1^*, o^P, h^P), ..., (m^P, x_n^*, o^P, h^P)]$. It is then natural to state that our normative evaluation of inequalities in well-being should be the same for both situations. Given that (m^P, o^P, h^P) is the same for all individuals in the latter situation, this means that we can apply our apparatus of inequality measurement to the vector of healthy-equivalent consumption levels (x_i^*).

To interpret this approach, note that the healthy-equivalent consumption x_i^*, when it is well defined, actually yields a utility function that represents the individual preference ordering $\Re_i$. In order to see this, consider (m_i, c_i, o_i, h_i), (m_i', c_i', o_i', h_i') such that

$$(m_i, c_i, o_i, h_i) \Re_i (m_i', c_i', o_i', h_i'),$$

and let $x_i^*, x_i'^*$ denote the corresponding healthy-equivalent consumption levels. Using (16) and transitivity of preferences, one necessarily has

$$(m^P, x_i^*, o^P, h^P) \Re_i (m^P, x_i'^*, o^P, h^P),$$

which, by monotonicity of preferences, is equivalent to $x_i^* \geq x_i'^*$. We can therefore denote this particular utility function simply as $x_i^*(m_i, c_i, o_i, h_i)$.

Although this approach satisfies the Pareto principle, it is not welfarist. The welfarist approach also satisfies the Pareto principle, and does so by defining social welfare as a function of individual subjective utility $U_i(m_i, c_i, o_i, h_i)$. The equivalent consumption $x_i^*(m_i, c_i, o_i, h_i)$ is not a welfarist notion because it is not a measure of subjective utility. Indeed, the two functions $U_i(.)$ and $x_i^*(.)$, although ordinally equivalent and both representing $\Re_i$, are different. One way of illustrating the difference is to look at what

---

[31]Rigorously speaking, x_i measures expenditures, but the word "consumption" is more pleasant and captures the essential meaning of the concept.

happens when two individuals i and j have the same preferences $\Re_i = \Re_j$ but different utility functions $U_i \neq U_j$. Within a welfarist approach, the difference in utilities will lead to (possibly unfair) inequality. For the healthy-equivalent consumption approach, in contrast, this difference in subjective utility does not lead to unfair inequality, because i and j have the same healthy-equivalent consumption functions $x_i^*(.) = x_j^*(.)$ just as they have the same preferences $\Re_i = \Re_j$.[32]

Measures of unfair inequality will in this approach depend on the choice of perfect health h^P and no labor duty o^P as the reference values. We took o^P as an example borrowed from the literature in order to fix ideas. However, we do think that perfect health is a natural choice of a reference for h. It seems beyond discussion that one considers an individual to be better off than another whenever he has a better consumption-labor bundle, provided both are healthy. This makes sense, whereas the same kind of judgment would appear questionable if both individuals were not healthy. Imagine that they have the same mediocre health and, say, slightly unequal consumption-labor bundles. In this case it is not obvious that the individual with a better consumption-labor bundle is better off. Maybe he cares more about health, and therefore suffers more from his health condition than the other one. This problem cannot occur with healthy individuals. We do not claim that some healthy do not enjoy their good health more than others. We simply say that it would be a strange conception of fairness to advocate lowering the consumption of the healthy individuals who care about health in order to increase the consumption of other equally healthy individuals who care less about health. If one accepts the idea that some kind of equality with respect to consumption-labor bundles would be a sound ideal for a uniformly healthy population, then the reference to h^P should appear acceptable.

Looking at eq. (16) shows that computing unfair inequality in healthy-equivalent consumption is indeed equivalent to calculating unfair inequality in the bundles (m_i, x_i, o_i, h_i) while respecting individual preferences. Using eq. (5), these bundles can in fact also be

---

[32] Such a neglect of utility functions (as opposed to preferences) can be viewed as an application of the principle of no influence of legitimate differences. If we hold individuals responsible for how they transform ordinal satisfaction into numerical utility, it makes sense to ignore it in the evaluation of their situations.

expressed as

$$(m_i, y_i - B(m_i, r_i) - T(y_i, c_i) - \rho(r_i, e_i), o_i, h_i).$$

This immediately suggests that calculating inequality in healthy-equivalent consumption offers also a coherent approach for thinking about equity in health care finance (Wagstaff et al., 1992, 1999).

7 Conclusion

Equality of access to health care is one of the proclaimed objectives of social policy in many countries. There is also growing recognition that inequality in health (e.g. in life expectancy) is one of the most shocking instances of injustice in our societies. In both domains, the focus of the social debate is often on socio-economic inequalities. This focus has been taken over by the huge amount of academic research that has been produced on these topics in recent years.

While we think that the concern about socio-economic inequalities in health care delivery and in health is certainly justified, we have argued that it should be integrated in a broader perspective on distributive justice. The perspective should be broadened in at least two respects. First, unfairness in health and health care do not only follow from socio-economic factors. A balanced view on the strengths and weaknesses of social arrangements should integrate these other factors. Second, there is a hierarchy of principles. Equality of access to health care matters because it is important for health. And inequality in health is worrying because health is one of the main components of personal well-being. This is not to say that it is not interesting (or even necessary) to consider these subdomains of social policy. However, at the end, the evaluation at a lower level of the hierarchy should be consistent with the evaluation at the higher level. As an example of this idea, we proposed to evaluate equity in delivery in terms of a "minimal efficient budget". Along the same lines, we argued strongly against the use of the concentration curve. The use of the latter implies some unattractive ethical assumptions, more specifically that it is always good to reduce health in higher social classes in order to increase it in lower

classes. The main intuition for the concern about socio-economic inequalities is one of "cumulative deprivation": we argue that this concern can be captured in a more attractive way by evaluating socio-economic inequalities in health in terms of inequalities in overall personal well-being. We have shown that a coherent measure of individual well-being can be constructed which respects individual preferences but is non-welfarist.

Our approach emphasizes the importance of drawing a clear distinction between the empirical (or positive) question of explaining behavior (and health) on the one hand and the normative question of evaluating unfair inequalities on the other hand. To get a better understanding of the different explanatory factors, it is necessary to build a coherent structural model. Once this information is available, two important *normative* decisions have to be taken.

First, one has to decide for each of the causal variables in the model whether they lead to ethically legitimate or ethically illegitimate differences in health or health care. Different ethical views will lead to different choices. Those advocating equality of access as an ideal will hold people responsible for their preferences, and will therefore consider preference differences as a legitimate source of inequalities. Proponents of equality of use will take the opposite position. Or, to give another example: while it seems to have been rather generally accepted that inequalities in health due to biological differences are not unfair, we have argued that this is not at all obvious. The framework we propose is sufficiently flexible to accomodate these different views and to compare the empirical results.

Second, in general there will exist no inequality measure which satisfies two interesting conditions. The first ("no influence of legitimate differences") states that a measure of unfair inequality should *not* reflect variations in outcomes which are caused by differences in the responsibility variables. The second ("compensation") states that, if a measure of unfair inequality is zero, there should be no illegitimate differences left. We proposed two measures, inspired respectively by the conditional-egalitarian and the egalitarian-equivalent solutions from the theory of fair allocation. The first ("direct unfairness") satisfies the first condition, but not the second. The alternative measure ("fairness gap")

41

satisfies the compensation condition, but will in general take up some differences due to responsibility variables. We have argued that the fairness gap is probably the better of the two measures, but this again reflects a normative choice. A closer look shows that direct unfairness and the fairness gap correspond to direct and indirect standardization respectively. In this way we have also shown what are the normative (rather than statistical) implications of both standardization procedures.

This paper has explored some links between health economics on the one hand, and the social choice theory of fair allocation on the other hand. Given the relative popularity of non-welfarist approaches in health economics, it is surprising that the connection between these two streams of literature has until now been rather tenuous. In our view, the theory of fair allocation offers a coherent framework to analyse important normative issues in the domain of health and health care. Further exploration of these links may be a fruitful path for further research.

8 References

Abul Naga, R. and Geoffard, P.-Y. (2006). Decomposition of bivariate inequality indices by attributes. Economics Letters 90(3):362-367.

Becker, G., Philipson, T., and Soares, R. (2005). The quantity and quality of life and the evolution of world inequality. American Economic Review 95(1):277-291.

Bleichrodt, H. and Van Doorslaer, E. (2006). A welfare economics foundation for health inequality measurement. Journal of Health Economics 25:945-957.

Bommier, A. and Stecklov, G. (2002). Defining health inequality: why Rawls succeeds where social welfare theory fails. Journal of Health Economics 21:497-513.

Culyer, A. and Wagstaff, A. (1993). Equity and equality in health and health care. Journal of Health Economics 12:431-457.

Dias, P. and Jones, A. (2007). Giving equality of opportunity a fair innings. Health Economics 16:109-112.

Erreygers G. (2006) Beyond the health Concentration Index: An Atkinson alternative for the measurement of the socioeconomic inequality of health. Paper presented at the conference Advancing Health Equity, Helsinki, WIDER-UNU, September.

Fleurbaey, M. (2005). Health, wealth and fairness. Journal of Public Economic Theory 7(2):253-284.

Fleurbaey, M. (2006a). Social welfare, priority to the worst-off and the dimensions of individual well-being. In F. Farina, E. Savaglio eds, Inequality and Economic Integration, London: Routledge, 2006.

Fleurbaey, M. (2006b). Health, equity and social welfare. Annales d'Economie et Statistique 83-84: 21-59.

Fleurbaey, M. (2008). Fairness, Responsibility, and Welfare. Oxford: Oxford University Press.

Fleurbaey, M. and Schokkaert, E. (2007). Economic evaluation of health care: distribution matters. Mimeo.

Goddard, M. and Smith, P. (2001). Equity of access to health care services: theory and evidence from the UK. Social Science and Medicine 53:1149-1162.

Gravelle, H. (2003). Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics 12:803-819.

Hurley, J. (2000). An overview of the normative economics of the health sector. In A. Culyer and J. Newhouse (eds.), Handbook of Health Economics, Vol. 1A. Amsterdam: Elsevier (North-Holland), 55-118.

Jones, A., Koolman, X., and Van Doorslaer, E. (2006). The impact of supplementary private health insurance on the use of specialists in selected European countries. Annales d'Economie et Statistique 83-84: 251-275.

Koolman, X. and van Doorslaer, E. (2004). On the interpretation of a concentration index of inequality. Health Economics 13:649-656.

Luttens, R. and Van de gaer, D. (2007). Lorenz dominance and non-welfaristic redistribution. Social Choice and Welfare 28(2):281-302.

Mackenbach, J. and Kunst, A. (1997). Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Social Science and Medicine 44(6):757-771.

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

Roemer, J. (1998). Equality of opportunity. Cambridge, Mass.: Harvard University Press.

Roemer, J. (2002). Equality of opportunity: a progress report. Social Choice and Welfare 19(2):455-471.

Schokkaert, E., Dhaene, G., and Van de Voorde, C. (1998). Risk adjustment and the trade-off between efficiency and risk selection: an application of the theory of fair compensation. Health Economics 7:465-480.

Schokkaert, E. and Van de Voorde, C. (2004). Risk selection and the specification of the conventional risk adjustment formula. Journal of Health Economics 23:1237-1259.

Schokkaert, E. and Van de Voorde, C. (2006). Incentives for risk selection and omitted variables in the risk adjustment formula. Annales d'Economie et Statistique 83-84: 327-352.

Schokkaert, E. and Van de Voorde, C. (2007). Direct versus indirect standardization in risk adjustment. Mimeo.

Sutton, M. (2002). Vertical and horizontal aspects of socio-economic inequity in general practitioner contacts in Scotland. Health Economics 11:537-549.

Tsuchiya, A. and Williams, A. (2005). A "fair innings" between the sexes: are men being treated inequitably? Social Science and Medicine 60:277-286.

van Doorslaer, E. and Jones, A. (2003). Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics 22:61-87.

van Doorslaer, E. and Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries. Health Economics 13:609-628.

van Doorslaer, E., Koolman, X., and Jones, A. (2004). Explaining income-related inequalities in doctor utilisation in Europe. Health Economics 13:629-647.

van Doorslaer, E. et al. (1992). Equity in the delivery of health care: some international comparisons. Journal of Health Economics 11:389-411.

van Doorslaer, E., et al. (2000). Equity in the delivery of health care in Europe and the US. Journal of Health Economics 19:553-583.

Van Ourti, T. (2004). Measuring horizontal inequity in Belgian health care using a Gaussian random effects two part count data model. Health Economics 13:705-724.

Wagstaff, A. (2002). Inequality aversion, health inequalities and health achievement. Journal of Health Economics 21:627-641.

Wagstaff, A. and van Doorslaer, E. (2000a). Equity in health care finance and delivery. In A. Culyer and J. Newhouse (eds.), Handbook of Health Economics, Vol. 1B. Amsterdam: Elsevier (North-Holland), 1803-1862.

Wagstaff, A. and van Doorslaer, E. (2000b). Measuring and testing for inequity in the delivery of health care. Journal of Human Resources 35(4):716-733.

Wagstaff, A. and van Doorslaer, E. (2004). Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations. Health Economics 13:297-301.

Wagstaff, A., van Doorslaer, E., and Paci, P. (1991). On the measurement of horizontal inequity in the delivery of health care. Journal of Health Economics 10:169-205.

Wagstaff, A., van Doorslaer, E., and Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics 112:207-223.

Wagstaff, A. et al. (1992). Equity in the finance of health care: some international comparisons. Journal of Health Economics 11:361-387.

Wagstaff, A. et al. (1999). Equity in the finance of health care: some further international comparisons. Journal of Health Economics 18:263-290.

Williams, A. and Cookson, R. (2000). Equity in health. In A. Culyer and J. Newhouse (eds.), Handbook of Health Economics, Vol. 1B. Amsterdam: Elsevier (North Holland), 1863-1910.

Zheng, B. (2006). Measuring health inequality and health opportunity. Mimeo: Department of Economics, University of Colorado.