Equity in health care prioritisation: An empirical inquiry into social value

Elly A. Stolk a, *, Stefan J. Pickee a, André H.J.A. Ament b, Jan J.V. Busschbach c

a Institute for Medical Technology Assessment (iMTA), Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
b Department of Health Organisation, Policy and Economics, Maastricht University, Maastricht, The Netherlands
c Department of Medical Psychology and Psychotherapy, Erasmus MC, Rotterdam; VISPD, Halsteren, The Netherlands

Abstract

The value of QALY gains for different patients may be recalculated using equity weights, but it is unclear which interpretation of equity should be used: severity of illness, fair innings or proportional shortfall. We set up an experiment to analyze which of these equity concepts best reflects people’s distributional preferences. Sixty respondents assigned a priority rank to the treatment of 10 conditions using the paired comparison technique. We described these real-life conditions by their actual QALY profiles, i.e. in terms of age, disease free period, duration of disease, quality of life, and life years lost. Next we determined the priority rank order of the 10 conditions by the three equity concepts, using the weights that each equity concept attributes to the different units of the QALY profile describing the 10 conditions. To explore the social interpretation of equity, we compared the observed and theoretical rank orderings using Spearman correlations. All correlations were significant at a 0.05 level. Fair innings best predicted the observed rank order of the 10 conditions (r = 0.95). Weaker correlations were found for proportional shortfall (r = 0.82) and severity of illness (r = −0.65). This result calls attention to health policy, because actual health care decisions often reflect concerns of severity of illness. This raises the question if health care decision makers evaluate the claims of different patients for health care by appropriate criteria.

Keywords: Health care; Equity; Social value

1. Introduction

The concept of efficiency in the distribution of health is usually taken to imply that people who can gain QALYs in a relatively cheap way are more entitled to treatment than other people are. Societal preferences, however, have demonstrated that other individual characteristics may also affect priority setting, because the principle of health maximization may lead to uneven distributions of health which conflicts with people’s equity concerns. To incorporate social concerns about equality in health in economic evaluations several authors have advocated operationalization of an equity-efficiency trade-off [1–3]. This trade-off reflects the willingness to decrease the total amount of benefits...
from our health care system, if this results in a more equitable distribution of health effects. Unfortunately, equity is an ethical concept that has no precise definition. Ever since the concept of the equity-efficiency trade-off has been introduced, it has been debated which equity concern(s) should be included in the trade-off. To contribute to this debate, we set up an empirical study to explore the social interpretation of equity.

There seems to be agreement that the definition of equity should be found in the health domain, but it is unclear what kind of measure of inequalities in health must be defined. Different authors have tried to persuade others to the use of different equity concepts, like severity of illness and fair innings. The severity of illness approach embodies the feeling that people facing severe illness must be rescued, whilst this urge to help declines when the health conditions are less severe. Therefore, patients in the most critical condition receive the highest priority, e.g. patients facing the threat of immediate death or a severe handicap. In effect, this approach gives highest priority to patients with the poorest health prospects without treatment. There is however no consensus that indeed patients with the poorest health prospects are always the most deserving.

Proportional shortfall has in common with fair innings that the size of the health gap is relevant, but it agrees with severity of illness that also the remaining no-treatment QALY expectation should be taken into account. From this viewpoint equity weights are not simply proportional to the absolute size of the health gap caused by a condition. Rather, equity weights should be determined on the basis of the amount of QALYs that a patient loses proportional to this person’s remaining QALY expectancy in normal health (e.g. calculated as the average expected number of QALYs for the population of that age and sex). Higher equity weights then apply if a patient loses a greater fraction of his or her remaining QALY expectation. Proportional shortfall thus values relative changes in expected QALYs, irrespective of the number of expected QALYs concerned. This reflects the idea that everyone is equally entitled to live out his or her remaining life span, no matter whether the remaining life span is long or short.

In the literature on inequalities such proportionate equity concepts have been discussed frequently, but little is known about the social support for this type of combination principles. Usually the two equity concerns that are combined are evaluated separately. In a recent paper Cuadras-Morató et al. [12] compared support for the absolute and proportionate equity concepts using axiomatic bargaining theory. Cuadras-Morató et al. recruited respondents to solve resource allocations, whereby the possible solutions specified shares of the available budget that would be allocated to different patients. The solutions represented six different distributive and equity concepts among which the utilitarian position, the fair innings argument and a proportionate equity concept similar to the proportional shortfall approach. Respondents had to indicate which solution they found most attractive. In that way this study explored what equity concept prevailed in circumstances where different views would result in different priorities [12]. This experiment found no dominant principle, but strongest support was for the proportional solution and fair innings. Which of these two solutions was preferred in the different situations depended on the differences in the capacity to benefit, the health gap, and the context. The authors conclude that more research into social support for equity concepts is warranted, and they advise to explore benefits of realistic examples in future surveys to study on a less abstract level support for different equity concepts.
Building on the studies discussed above we further explored social preferences for equity. In our study we used realistic cases to test support for different equity concepts, as suggested by Cuadras-Morató et al. [12]. We asked our respondents to priority rank treatments of 10 conditions using the paired comparison technique. To explore the social interpretation of equity, we compared this observed rank order to the rank orderings expected by the three equity concepts. This study contributes to existing literature by concentrating on the way in which people balance different equity concepts in a series of forced choice questions. The purpose is to see whether combined information on the different choices reveals the underlying decision process and weighting of equity concerns.

2. Methods

2.1. Respondents

We recruited a heterogeneous sample of students, researchers, and health policy makers (N = 65). Students and researchers were recruited at the departments of Health Sciences of the Erasmus University in Rotterdam and the University of Maastricht. Health policy makers were employed at the Dutch Healthcare Insurance Board.

2.2. Paired comparison scaling

Respondents had to priority rank 10 conditions using the paired comparison technique. The choices were presented on cards in random order (see for an example Fig. 1). Respondents were asked to indicate which of two patients in a different condition they would treat, if resources were lacking to provide both patients the treatment that they needed. Each respondent compared each condition to each of the remaining conditions, which means that they had 45 choices to make \( \binom{n}{2} \). This design assumes symmetry, i.e. it assumed that asking a person whether A would be preferred to B gives the same result as asking whether B would be preferred to A. Although this assumption not necessarily needs to hold, assuming symmetry was desirable as not to double the number of comparisons that needed to be made. We considered 45 comparisons a feasible number for each respondent to make. Many studies have employed at least 32 choices successfully and recent research suggested that an experiment might include over 40 comparisons ([16], p. 134). Moreover, the task was cognitively not complex, because respondents were familiar with issues of health care priority setting, and because many of the choices were straightforward in the sense that they contained a dominant alternative (better in all respects).

![Fig. 1. Example of a card presenting one of the choices in the paired comparison task.](image)
To introduce the paired comparison task, we described a hypothetical context in which the respondent would possess a wonder pill, which cures any patient who receives it. Unfortunately, in each pair of options two patients are in need of treatment, but only one pill is available. The respondent had to indicate which patient s/he would give priority. The wonder pill would relieve that patient of all described health problems and bring this person back to normal health; the other patient would not be cured and would have to endure the illness.

We determined a scale value for each of the 10 conditions that reflected their position compared to other conditions on the priority scale. To obtain these scale values, the paired comparison scaling technique assumes that the proportion of times that any alternative is chosen over any other alternative reflects discriminant dispersions of the alternatives on the continuum of an underlying unidimensional construct (h.l. priority rank). To determine scale values using paired comparison data, first a matrix must be constructed with the probability that each condition is prioritized over any other. Using the properties of the normal curve, these probabilities can be converted into z-values, corresponding to their location on a normal distribution. These z-scores can be interpreted as measures of priority at interval level. If no extreme values of p (equal to 1 or 0) would be present, the mean z-value for each condition gives its scale value directly. If extreme values do occur, these must be ignored, since the corresponding z-scores approach infinity. The data then should be analyzed using the Thurstone Case V Scaling Model for Incomplete data. See Edwards [17] for a detailed description of paired comparison scaling (pp. 40–47).

2.3. Health state descriptions

In the paired comparison task we presented realistic decomposed QALY profiles of 10 conditions. A QALY profile describes how quality of life develops over time. The specification of these QALY profiles was essential in the context of our experiment, because people adhering to different equity concepts will attribute different values to different characteristics of a patient’s QALY profile. The presented QALY profiles included all attributes that were relevant according to any of the equity concepts. Hence we provided information about the no-treatment QALY profile (to allow for rank ordering according to severity of illness), and the QALYs foregone measured from a predetermined age-related target level for health (allowing for rank ordering according to fair innings). The proportional shortfall combines the two and required no additional factors to be included in the health state description. Accordingly, the health states were described using the five units listed below:

(a) mean age of the patients (because that determines the target level for health)
(b) time without disability
(c) time with complaints
(d) average quality of life loss associated with the complaints
(e) life years lost

We included 10 real-life conditions for which the QALY profile had been established in the paired comparison task, because we assumed that people have clearer preferences for real-life cases. The use of real-life examples therefore might increase the predictive value of the study outcomes for actual decisions in real life. Data regarding the no-treatment QALY profiles of the 10 conditions were available from clinical and economic investigations that were performed in the Netherlands. The original researchers were involved in the process of deriving the correct estimates from the models. These data are presented in Table 1. In addition, the health states were labelled and offered a general description of the health problem, again because we assumed that people’s preferences are clearer when explicit labels are used rather than neutral and abstract descriptions of a health state. We learnt from a pilot that the use of labels and descriptions did not introduce biases or framing effects: this pilot in 20 students confirmed that the observed rank order was unaffected by the presentation of labels and health state descriptions.

2.4. Establishing the theoretical rank orderings

Table 1 also presents the three theoretical rank orderings and the scores of the 10 conditions on the equity scales, which were determined using the weights that each equity concept attributes to the different units of the QALY profile describing the 10 conditions. Below we discuss the specifics of these calculations. Note...
Table 1  Scaling and rank ordering of 10 conditions by different equity concepts

<table>
<thead>
<tr>
<th>Condition</th>
<th>Treatment</th>
<th>Parameters</th>
<th>Sealing (rank)</th>
<th>Proportional shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>Health years</td>
<td>Disabled years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLY</td>
<td>DLY</td>
</tr>
<tr>
<td>Onychomycosis</td>
<td>Terbinafine</td>
<td>60</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Benign Prostate Hyperplasia</td>
<td>Finasteride</td>
<td>70</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Oestrogen</td>
<td>70</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>High Blood Pressure</td>
<td>Antihypertensives</td>
<td>55</td>
<td>15</td>
<td>6.5</td>
</tr>
<tr>
<td>Pneumococcal pneumonia</td>
<td>Vaccination</td>
<td>70</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>High cholesterol</td>
<td></td>
<td>60</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Non-Hodgkin Lymphoma (intermediate/high grade)</td>
<td>CHOP-chemo.</td>
<td>50</td>
<td>13</td>
<td>7.5</td>
</tr>
<tr>
<td>COPD</td>
<td>Ciprofloxacin</td>
<td>50</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Arthritis rheumatoid</td>
<td>Methotrexate</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* Abbreviations: HLY, healthy life years; DLY, disabled life years; QoL, quality of life; LLY, life years lost.

** The estimates of age, HLY, DLY, QoL, and LLY describe the QALY profiles of patients at the moment that the specified treatment is considered. Other estimates apply when other treatments are considered in other stages of the disease. This and other measurement issues are discussed elsewhere [28].

---

that we simplified the calculations by assuming that no quality of life losses would occur due to other illnesses: the age of a patient equalled the number of QALYs consumed so far, and each life year lost equalled a QALY.

2.4.1. Severity of illness

The severity of illness approach suggests that the patient with the poorest health prospects should get the highest priority. Usually this argument is illustrated with an example on quality of life: studies have shown that a movement from 0.6 to 0.9 receives a lower value than a movement from 0.3 to 0.6 when the size of the health gain is the same ([18], p. 75). This approach is most concerned about those patients whose inability to function in ordinary life is most pronounced. From this underlying idea one could infer that in terms of QALYs the severity of illness approach would be more concerned with people that have poor QALY prospects than with those who have better prospects if no treatment is provided (assuming equal potential gains). This can be illustrated using Fig. 2, which represents the development of a patient’s quality of life during the remaining life years. In the taxonomy of this figure, the size of area ‘x’ describes severity of illness in terms of the remaining QALY prospects: the smaller area ‘x’, the greater the severity of illness.

Accordingly, the rank ordering by severity of illness that is presented in Table 1 has been determined on the basis of the no-treatment QALY profile: the higher the absolute number of QALYs that a person will get in spite of his or her condition, the lower his or her QALY.
priority ranking. In these calculations, all disease free years counted for 1 QALY, QALYs from disabled life years where calculated by multiplying the number of life years by the quality of life weights during those years.

2.4.2. Fair innings

Whilst the severity of illness argument considers only inequalities in future health, the fair innings argument is concerned with lifetime health. The fair innings argument requires that everyone be given an equal chance to have a predefined fair innings, and a patient’s entitlement for health care drops when s/he gets closer to the fair innings. To put this equity concept into practice, two issues need to be discussed. The first issue concerns past health losses. The fair innings argument bases priorities on the number of QALYs foregone, no matter whether the health losses occurred in the past or will occur in the future. In practice this is not always easy to include in calculations of the fair innings, because the literature usually concentrates on changes in future health that maybe brought about by treatment. We therefore simplified the calculations by assuming that all patients have had healthy lives until the moment that they applied for the intervention under consideration. This means that the size of area ‘y’ in Fig. 2 determined the rank ordering by fair innings in this investigation. In this figure ‘y’ represents health loss: the greater the health loss, the greater priority is placed on treatment of that patient.

Another question in operationalizing the fair innings argument is what the target level of health should be from which the health gap is measured. When Williams illustrated the framework provided by the fair innings, he assumed a ‘fair innings’ of 70 QALYs [19]. Accordingly, the rank ordering by fair innings could be established by dividing the expected lifetime QALYs without treatment by 70. In this approach the fair innings is defined at birth and is equal for everybody. Alternatively, it could be argued that the target level for health innings should be recalculated for survivors according to where they are at present. The reason is that people’s health prospects improve when they age (expected age of death is delayed, hence the expected lifetime QALY totals increases with age) and that the years that are ‘added’ to their life expectancy at birth will also be considered a health loss. In this case, fair innings weights could be determined by dividing the expected lifetime QALYs without treatment by an age specific lifetime QALY expectation.

Williams discussed arguments in favour of and against this ‘dynamic’ version of the fair innings, emphasizing that a dynamic version has more appreciation for health needs of the old. This implies that this method is less powerful to reduce inequalities in total health outcomes over age groups [2]. Different demands of equality underlie the two versions of fair innings, reflecting different views as to what aspects of people’s health are to be valued in a priority setting context [2] and how people are to be assessed vis-à-vis each other. As there are different ways of seeing this, we decided to put both versions into the experiment: in one version fair innings is fixed at 70 QALYs, in the other version a dynamic fair innings is used, that is calculated using age and gender specific mortality data. Gender was brought in the equation to further individualize the dynamic fair innings, so as not to over- or underestimate the remaining life expectancy of patients in conditions that are gender specific. This means that our dynamic version of fair innings was indifferent to inequalities in life expectancy at birth between males and females. Table 1 shows several reversals in the tabulated rankings of the 10 conditions by the two versions of fair innings.

2.4.3. Proportional shortfall

Whilst both severity of illness and fair innings suggest that equality should be measured in terms of absolute attainments, a third equity concept suggests that comparison may be done in terms of ‘proportionate’ or ‘relative’ attainments [10]. Proportional equality in health corresponds with equalizing persons within their own scope of potential for health, by distributing shares proportionate to people’s remaining life expectancy instead of equal shares. Accordingly, different patients are asked to make the same proportional concession to their target level for health, or to accept the same ‘proportional shortfall’. Hence equity weights are determined not on the basis of absolute health loss as the fair innings argument assumes, but on the basis of proportional health loss. Since it is the remaining life expectancy that counts and that people want to maximize, this approach calculates remaining life expectancy on the basis of age and gender specific mortality data. This means that proportional shortfall does not compensate for inequalities at birth in the life expectancy between
males and females, like the dynamic version of fair innings.

To determine the rank ordering by proportional shortfall first the absolute health loss is measured relative to an age and gender dependent target level for health, similar to the approach in the second version of the fair innings argument. This time however, it is not the absolute loss that counts. Instead, proportional shortfall values the QALY loss relative to the number of QALYs that a patient would receive in the future if he would be in normal health. In other words, proportional shortfall measures the ratio between a patient’s QALY expectation if no treatment is to be received (‘x’ in the taxonomy of Fig. 2), and his or her QALY expectation in absence of the considered condition (‘x + y’).

2.5. Analysis

To explore the social interpretation of equity, we analyzed the congruence between the observed rank order obtained in the paired comparison experiment and the theoretical rank orderings expected by the three equity concepts using Spearman rank order correlations ($p < 0.05$). A Fisher $z$-transformation of the correlations was used to test correlation differences ($p < 0.05$). First, however, we explored the robustness of the preferences that were elicited in the paired comparison task. For this purpose we applied an internal consistency check. Hereto the difference in the $z$-value in a pair of conditions was back-transformed to the expected probability that one of the conditions was preferred over the other ([17], pp. 37–40). When the difference between the expected and the observed data is of the order of 0.05 for most comparisons, the model adequately fits the data ([20], pp. 39–44, [21]). Additionally we ran significance tests for the coefficient of consistency (to evaluated consistency in the individual preference orderings) and the coefficient of agreement (to measure agreement among different respondents). Formulas to compute these coefficients can be found in [17] (chapter 3).

The rationale underlying the paired comparison technique predicts that the occurrence of inconsistent individual judgments increases as the difference between the compared objects on the underlying continuum decreases. The technique of paired comparison ranking therefore allows individual rank-order inconsistencies. However also other factors might contribute to inconsistencies in individual rank-orderings (poor task comprehension, inability to compare the objects, disinterest in the task), so it is desirable to obtain a measure of the degree of consistency in the responses of each subject. Given that in the paired comparison task each scenario was compared to each other, we were able to test for consistency of a subject’s responses. When a respondent is inconsistent, intransitivities occur in the preference ordering. For example, when A is preferred to B, and B is preferred to C, logic predicts that A will be preferred to C. If C on the other hand is considered more favourable these three comparative judgments constitute a circular triad. The number of circular triads is used to find Kendall’s coefficient of consistency, which offers a measure of inconsistency in the responses of a particular judge. To test significance of the observed degree of consistency we explored if consistency was greater than can be expected by chance using the $\chi^2$ distribution.

Next we computed the coefficient of agreement, which reflects diversity of preferences among the respondents. When the coefficient approaches 1.0, the subjects have nearly equal orderings. Complete agreement is reached when all respondents make identical choices during the experiment, in which case half of the entries in the preference matrix presented in Table 2a would be equal to 1.0, while the other half would be zero. Alternatively, if agreement is completely absent among the subjects, all entries will be equal to 0.5. Each time that two test subjects make the same decision in a paired comparison question, we say that we have one agreement regarding this pair. Agreement is measured by counting the number of pairs of test subjects that make the same decision over each pairs of health states that are compared. Again, a $\chi^2$ statistic was used to determine to test agreement among respondents, null hypothesis being that all test subjects cast their preference completely at random.

3. Results

3.1. Sample

Twenty-four students, 24 researchers and 17 health policy makers filled in the questionnaire, which took them on average about 20 min. We found no
differences in the rank ordering of the three groups. The only exception was that students ranked high cholesterol on place 4 and non-Hodgkin lymphoma on place 5, whilst the other two groups reversed these two positions. Given these marginal differences between the groups we only present the aggregated data.

3.2. Paired comparison data

We recorded the frequencies that each alternative was preferred over another alternative. Table 2a presents the \( p \)-matrix that summarizes the frequency data. The corresponding \( z \)-matrix and distances matrix are presented in the Tables 2b and 2c, along with the scale values. These scores indicate the relative distances between the 10 conditions on the priority scale. The internal consistency check demonstrated that the observed proportions agreed well with those to be expected in terms of the derived scale values: the absolute average discrepancy was 0.048. The coefficient of consistence showed a mean value of 0.947 \(( p < 0.001)\), indicating a high level consistency. Participants in the study generally had few circular triads: 30 participants had no triad. The average number of circular triads was 2.1 (S.D. 3.33) with a maximum of 16 (the maximum number of triads possible was 40). The coefficient of agreement was high: 0.721 \(( p < 0.001)\). This means that subjects were consistent among each other.

3.3. Comparison of observed and theoretical rank orderings

Fig. 3 presents the relations between the observed scale values (\( z \)-scores) and the scale values expected by the three equity concepts. Spearman correlations show that all theoretical rank orderings were statistically significant correlated to the observed rank ordering. The highest correlations were found between the rank orderings by the two versions of fair innings and the observed rank ordering: 0.985 \(( p < 0.001; 95\% \text{ CI } = 0.94 \text{ to } 1.00)\) for the dynamic version of fair innings and 0.948 \(( p < 0.001; 95\% \text{ CI } = 0.79 \text{ to } 0.99)\) for the fair innings with a fixed target of 70 QALYs. The Spearman correlations with proportional shortfall and severity of illness were 0.818 \(( p = 0.004; 95\% \text{ CI } = 0.39 \text{ to } 0.96)\) and \(-0.648 \(( p = 0.043; 95\% \text{ CI } = -0.91 \text{ to } -0.03)\), respectively. Note that the negative sign of
Table 2b
*z*-matrix, eliminating *p*-values greater than 0.98 and less than 0.02

<table>
<thead>
<tr>
<th>Condition</th>
<th>Onychomycosis</th>
<th>Benign prostatic obstruction</th>
<th>Osteoporosis</th>
<th>High blood pressure</th>
<th>Pneumococcal pneumonia</th>
<th>High cholesterol</th>
<th>Non-Hodgkin lymphoid</th>
<th>COPD</th>
<th>Arteriosclerosis</th>
<th>Pulmonary hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onychomycosis</td>
<td>0.000</td>
<td>1.542</td>
<td>1.870</td>
<td>1.870</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign prostatic obstruction</td>
<td>-1.542</td>
<td>0.000</td>
<td>1.542</td>
<td>1.683</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>-1.542</td>
<td>0.000</td>
<td>1.683</td>
<td>1.542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High blood pressure</td>
<td>-1.542</td>
<td>-1.683</td>
<td>0.214</td>
<td>0.417</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumococcal pneumonia</td>
<td>-1.683</td>
<td>-1.319</td>
<td>0.000</td>
<td>0.174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High cholesterol</td>
<td>-1.683</td>
<td>-1.319</td>
<td>0.058</td>
<td>1.239</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hodgkin lymphoid</td>
<td>-1.319</td>
<td>-1.239</td>
<td>0.639</td>
<td>0.957</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>-1.239</td>
<td>-1.683</td>
<td>0.334</td>
<td>0.736</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arteriosclerosis</td>
<td>-1.683</td>
<td>-1.683</td>
<td>0.000</td>
<td>0.736</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>-1.683</td>
<td>-1.683</td>
<td>1.870</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2c
Matrix of the differences between the *z*-values*, corresponding to the scale distances

<table>
<thead>
<tr>
<th>Calculation</th>
<th>Onychomycosis (j)</th>
<th>Benign prostatic obstruction (i)</th>
<th>Osteoporosis (h)</th>
<th>Pneumococcal pneumonia (f)</th>
<th>High blood pressure (g)</th>
<th>High cholesterol (e)</th>
<th>Non-Hodgkin lymphoid (d)</th>
<th>COPD (c)</th>
<th>Arteriosclerosis (b)</th>
<th>Pulmonary hypertension (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i - j</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>h - i</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>e - h</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>f - e</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>g - f</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>d - g</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>c - d</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>b - c</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
<tr>
<td>a - b</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
<td>1.542</td>
</tr>
</tbody>
</table>

*Distances were determined by subtracting entries in each column from the corresponding entries in the next column.*
the latter correlations is expected, because priority increases when health prospects decrease.

A Fisher $z$-transformation was used to explore if these correlations differed significantly from each other. This analysis revealed that the correlations of the two versions of fair innings did not differ ($p=0.12$). Also the correlations of proportional shortfall and severity of illness did not differ ($p=0.24$). However, the theoretical rank generated on the basis of fair innings was significantly more correlated with the observed rank ordering than the others. The correlation differences between the two versions of fair innings and both proportional shortfall and severity of illness were all highly significant ($p<0.001$). Visual examination of Fig. 3 also demonstrates the most consistent relation between fair innings and the observed value: there were no major exceptions to the rule that the observed value increased with the fair innings foregone. In general, the observed value also increased when the proportional shortfall increased. There were however two exceptions: non-Hodgkin lymphoma and pneumococcal pneumonia. People were less concerned about the treatment of these two life-threatening conditions for the elderly than expected on the basis of proportional shortfall.

Since the $z$-score is a relative measure, the absolute values of $'z'$ have no meaning unless they can be related to an absolute scale. This was only the case for proportional shortfall. In terms of proportional shortfall, the 10 conditions were nicely spread across the whole scale.
(proportional shortfall ranged from 0.02 to 0.98, on a scale ranging from 0.00 to 1.00). This means that other indications cannot lose smaller or greater fractions of health than the conditions included in this study already illustrate. To emphasize this, the z-scores were linearly transformed onto a scale from 0 to 1. Since the 10 conditions did not cover the whole spectrum of possibilities on the scales of fair innings and severity of illness, the absolute values of ‘z’ have no meaning for these two equity concepts: higher or lower values may have been found when other conditions would have been included in the experiment. Relevant is then only the fact that the z-scores have interval properties, and that we get an idea of the relative differences between different conditions on an equity rank ordering. For that reason values of z were not mentioned in Fig. 3 for severity of illness and fair innings.

4. Discussion

To determine what interpretation of equity should be used in recalculating the value of QALY gains for different patients, we compared the observed rank order of the 10 conditions with the rank orders that were expected by the three equity concepts: severity of illness, fair innings, and proportional shortfall. The results showed that the observed rank order of the 10 conditions was best predicted by the fair innings concept. Proportional shortfall was also highly correlated with the observed rank order. The severity of illness approach showed a moderate correlation with the observed ranking, suggesting that this concept is less consistent with social preferences for equality in health.

Fair innings had the highest correlation with observed preferences, suggesting that this concept received strongest social support. Moreover, proportional shortfall overestimated the value of treatment of two conditions in the elderly considerably (see Fig. 3). However, it may be too soon to claim superiority of fair innings yet, because effect size might have confounded the results. Respondents assumed that the wonder pill would relieve the recipient of all described health problems. Because the effect size equalled the health gap in the no-treatment QALY profile, a preference for QALY maximization could have boosted the correlation between the observed rank ordering and fair innings. One could wonder why we did not use a fixed effect to prevent this collinearity. However, the use of a fixed effect could not solve this problem of collinearity; it merely moves it. Because the proportional shortfall approach is consistent with the maximization of relative QALYs [10], the use of a fixed effect produces collinearity between the observed rank ordering and the rank ordering for proportional shortfall. In future studies the best strategy may therefore be to test preferences for varying effect sizes in different situations depending on the absolute and relative health gap.

The theoretical rank generated on the basis of fair innings was significantly more correlated with the observed rank ordering than the others. The correlation differences between the observed rank ordering and the rank orderings generated by proportional shortfall and severity of illness did not differ significantly. An explanation is that the 95% confidence interval of the Spearman correlations for severity of illness and proportional shortfall were relatively wide. Moreover, results may be produced when a higher number of cases is included. Moreover, our choice to include realistic cases may have limited robustness of the results to some degree, because cases were not selected randomly or most efficiently in terms of the statistic properties of the design. Case selection was conditional on the availability of the QALY profiles, so that discriminative ability of the design was not maximized. The design was slightly unbalanced in the sense that the cases were not equally well divided over the three theoretical scales. In terms of proportional shortfall, the 10 conditions were nicely spread across the whole scale, but for fair innings and severity of illness only half the scales were used (see Fig. 3). Additionally, our case selection did not try to minimize the a priori correlations between the three theoretical rank orderings. The rank orderings by severity of illness and fair innings were not statistically significant correlated \((p < 0.05)\), but the rank orderings by proportional shortfall was correlated to fair innings \((0.717, p = 0.020)\) for fixed target fair innings, \(0.800, p = 0.005\) for dynamic fair innings) and severity of illness \((0.855, p = 0.002)\). Since proportional shortfall combines elements of fair innings and severity of illness, a correlation with the other equity concepts is unavoidable. However, a greater level of orthogonality may be obtained when (hypothetical) cases are carefully selected.

Finally some conceptual issues need to be discussed. First, some researchers may disagree that distributive
justice belongs to the scope of economics. This may be traced back to different opinions about the social objective of health care, the type of utilities that constitute social welfare, or the legitimacy of addressing equity issues in economic models [22]. Based on the idea that both measures of efficiency and fairness depend on value judgments ([23], p. 47), however, we believe that it is appropriate to pay attention to equity issues to improve the descriptive validity of economic models. A second conceptual issue that needs to be discussed is whether all three positions are equally appropriate to inform resource allocation, notwithstanding social support. From a theoretical perspective a strong case can be made for equality in lifetime health, because health is an important condition of human life and a critical constituent of opportunities [24]. People who demand equality in lifetime QALYs may therefore find both the severity of illness and the proportional shortfall approach inadequate [19]. Nevertheless the public also seems to support principles that focus on health achievements rather than on the opportunities of people to accomplish what they value. As objectivity in ethics is warranted, such ‘gut feelings’ may need to be criticized and rationalized. Still, ‘gut feelings’ may also function as eye-opener and bridge towards new ways of moral consideration [25]. We therefore did not a priori exclude particular equity concepts from this investigation.

The information on equity preferences obtained in this study might have implications for policy. In The Netherlands, interventions with vital consequences for the elderly have always been funded, even in spite of unfavourable cost-effectiveness [26]. These policy choices can best be explained by the equity concept of severity of illness, an equity concept that did not reflect the distributive preferences of our sample well. An explanation for the discrepancy between our outcomes and recent policy decisions may be that for the sake of this experiment, a high level of scarcity was assumed. When health policymakers are confronted with terminally ill patients the question is if they perceive the same level of scarcity, or that they will be inclined to increase budget for the health care sector [27].

To conclude, this investigation aimed to contribute to the debate about the social interpretation of equity. An important message is that measurable interpretations of equity make it possible to validate people’s claims on health care. We conclude that the fair innings argument and proportional shortfall may provide a better basis for determining equity weights for recalculating the value of QALY gains for different patients than the severity of illness approach. Given the apparent conflict between recent health care decisions and the results of our investigation, the question can be raised whether current allocation of the health care budget is in line with societal preferences.

Acknowledgements

We acknowledge financial support from NWO, the Netherlands Organisation for Scientific Research under project DO 945-10-034 and ZON-MW, the Netherlands Organisation for Health Research and Development under project VVS-2-7. An earlier version of this paper was presented at the ISPOR meeting in Rotterdam, November 2002.

References


