

Societal Preferences for Distributive Justice in the Allocation of Health Care Resources: A Latent Class Discrete Choice Experiment

Chris Skedgel, PhD, Allan Wailoo, PhD, Ron Akehurst, BSc, Hon, MFPHM

Economic theory suggests that resources should be allocated in a way that produces the greatest outputs, on the grounds that maximizing output allows for a redistribution that could benefit everyone. In health care, this is known as QALY (quality-adjusted life-year) maximization. This justification for QALY maximization may not hold, though, as it is difficult to reallocate health. Therefore, the allocation of health care should be seen as a matter of distributive justice as well as efficiency. A discrete choice experiment was undertaken to test consistency with the principles of QALY maximization and to quantify the willingness to trade life-year gains for distributive justice. An empirical ethics process was used to identify attributes that appeared relevant and ethically justified: patient age, severity (decomposed into initial quality and life expectancy), final health state, duration of benefit, and distributional concerns. Only 3% of respondents maximized QALYs with every choice, but scenarios with

*larger aggregate QALY gains were chosen more often and a majority of respondents maximized QALYs in a majority of their choices. However, respondents also appeared willing to prioritize smaller gains to preferred groups over larger gains to less preferred groups. Marginal analyses found a statistically significant preference for younger patients and a wider distribution of gains, as well as an aversion to patients with the shortest life expectancy or a poor final health state. These results support the existence of an equity-efficiency tradeoff and suggest that well-being could be enhanced by giving priority to programs that best satisfy societal preferences. Societal preferences could be incorporated through the use of explicit equity weights, although more research is required before such weights can be used in priority setting. **Key words:** equity in distribution; resource allocation; survey methods; psychometric methods; cluster analysis. (*Med Decis Making* 2015;35:94-105)*

Conventional health economic theory suggests that health care resources should be allocated in the way that produces the greatest aggregate health gains. These gains have most often been measured in terms of the quality-adjusted life-year

(QALY), which summarizes quality of life and years of life.^{1,2} Under this model, any increase in quality, duration, or the number of patients benefiting is associated with a proportional increase in societal well-being.³ This leads naturally to a policy of QALY maximization. Critically, this model assumes “distributive neutrality,” or that society is indifferent to how gains are distributed between different individuals.^{3,4}

Welfare economics has typically justified its focus on maximization, and its disregard of distribution, by arguing that maximizing benefits allows for a redistribution that could make at least some people better off

Received 24 February 2014 from the School of Health & Related Research, University of Sheffield, Sheffield, UK (CS, AW, RK); Atlantic Clinical Cancer Research Unit, Capital Health, Halifax, Canada (CS); and Maritime Strategy for Patient-Oriented Research Support Unit, Capital Health, Halifax, Canada (CS). Financial support for this study was provided by grants from the Capital Health Research Fund and the Canadian Centre for Applied Research in Cancer Control (ARCC). ARCC is funded by the Canadian Cancer Society Research Institute. The funding agreements ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The authors have no conflicts to declare. Revision accepted for publication 19 July 2014.

© The Author(s) 2014

Reprints and permission:

<http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/0272989X14547915

Supplementary material for this article is available on the *Medical Decision Making* Web site at <http://mdm.sagepub.com/supplemental>.

Address correspondence to Chris Skedgel, Capital Health, Centre for Clinical Research, Room 242, 5790 University Ave., Halifax, NS B3H 1V7, Canada; telephone: +1 902 473 3303; fax: +1 902 425 1611; e-mail: chris.skedgel@cdha.nshealth.ca.

without making anyone worse off.⁵ The question of how (or whether) gains should be redistributed is seen as a political concern, outside the scope of economics. This logic may not hold in health care, though, as it is not usually possible to consider the production of health separately from its distribution; production, in the form of health care, and the allocation of health to a specific patient happen simultaneously.⁶ Furthermore, there is evidence of a societal willingness to sacrifice some potential QALY gains in order to promote equity in the distribution of those gains.^{3,7,8} For these reasons, the allocation of health care resources can be seen as a matter of distributive justice as well as technical efficiency. Technical efficiency requires that the output from a given set of resources be maximized, while distributive justice requires that the costs and benefits of that output be fairly distributed. However, these two principles can often conflict: The most efficient allocation may be less fair, while the fairest allocation may be less efficient.

The degree to which society is willing to sacrifice efficiency for fairness or equity is known as the “equity-efficiency tradeoff.”^{9,10} This tradeoff implies that a more equitable intervention can be less efficient and still be preferred over a more efficient but less equitable intervention. There is a limit to this tradeoff, though, and at some point an equitable but inefficient intervention will be less preferred than a more efficient intervention with a less equitable distribution of benefits. Prioritizing programs that better satisfy societal preferences could arguably improve societal well-being,¹¹ and one way to do this is with “equity-weighting,” which would weight life-year or QALY gains to reflect their value to society, not to the individual patient. The necessary first steps are to identify what characteristics are relevant to society’s conception of equity and to quantify the acceptable equity-efficiency tradeoff over each characteristic. These tradeoffs could be defined by decision makers, but it is argued that the moral legitimacy of the priority-setting process,¹² as well as public trust in the resulting decisions,¹³ can be enhanced by societal participation.

A number of recent stated preference studies have explored the equity-efficiency tradeoff,^{14–19} but they disagree over the magnitude and whether societal preferences are in fact consistent with a QALY-maximizing decision rule. This study contributes to the debate by administering a discrete choice experiment (DCE) in a representative sample of the Canadian population as well as a convenience sample of

decision-making agents. There were two main objectives: to test the consistency of respondents’ preferences with QALY maximization, and if respondent preferences were not consistent with strict QALY maximization, to quantify the equity-efficiency tradeoff for equity-relevant characteristics. To provide some insight into whether decision-making agents reflect the way they believe the public *would* act, or the way they *should* act, a secondary objective was to compare the preferences of agents and the general public. It was expected that agents would be more consistent with QALY maximization and less likely to demonstrate a significant equity-efficiency tradeoff.

The next section describes a process of empirical ethics used to identify a set of attributes that could be considered both relevant and fair, as well as the development of the experimental design and the statistical methods used to model choice observations. The results of the analysis of the DCE responses are then described, including estimates of the equity-efficiency tradeoff for the different attributes included in the elicitation and a ranking of the different scenarios in terms of choice probabilities and equity weights. We then discuss the implications and provide a conclusion.

METHODS

The conduct of the DCE was based on a process outlined by Ryan²⁰: identify attributes and levels, develop the experimental design, administer the DCE, and analyze the choice responses.

Identification of Attributes and Levels

Attributes were included in the elicitation through a process of empirical ethics, by which each attribute had to have empirical evidence of public support and be consistent with some coherent and defensible theory of justice.²¹ An empirical filter ensures that the attributes are relevant to society, while an ethical filter contributes to the legitimacy of those attributes and avoids the “moral relativism” that may result from strictly empirical methods, whereby what is morally right or wrong is reduced to social consensus or a simple majority.²²

Empirical studies were identified using a pearl growing strategy, beginning with 3 comprehensive reviews.^{3,7,8} Their bibliographies were searched, and the “related articles” feature of PubMed and Web of Science was used to identify other potentially relevant studies. Keywords from the reviews were

Table 1 Attributes and Levels Presented to Respondents

Level	Age	Initial Utility	Initial Life Expectancy	Final Utility	Gain in Life Expectancy	Patients Treated
1	10	1/10	1 month	1/10	1 year	100
2	40	5/10	5 years	5/10	5 years	2500
3	70	9/10	10 years	9/10	10 years	5000

Note: The values of aggregate QALYs gained ranged from 10 to 55,000. Utilities were presented on a 10-point scale but analyzed as decimal values.

also searched in PubMed, EconLit and Google Scholar. This review identified 14 unique characteristics or concepts, which were then reviewed with respect to their consistency with a defensible theory of justice. The theories of need, maximization, and egalitarianism, including Rawls' Difference principle²³ and Prioritarianism,^{24,25} were the principal theories used to justify the inclusion of an attribute, and these are briefly outlined in the online supplement.

The consistency of each attribute with different theories of justice was subjectively evaluated by the authors. The relevance of patient age, for example, was judged to be consistent with "fair-innings" egalitarianism as well as with gain and productivity-maximizing principles. The inherent, and arguably essential, subjectivity of this process is discussed later. Through this process, 4 attributes were judged to have evidence of public support and a defensible ethical justification: patient age, severity before/without treatment, final health state with/after treatment, and the distribution of health gains. To distinguish between severity as proximity to death and severity as a poor health state, this factor was decomposed into life expectancy without treatment and initial health state. Duration of benefit also appeared relevant despite some ambiguity over its interpretation, although it would have been included in the elicitation regardless of the empirical or ethical evidence in order to facilitate the calculation of QALYs to test the principle of QALY maximization. It is also worth highlighting that a number of attributes included in similar elicitations were excluded by the process used here. Most notably, prioritization on the basis of a patient's responsibility for his or her illness had empirical support but was judged to have little ethical justification,²⁶ while priority on the basis of social role had ethical justification but little empirical support, and both attributes were excluded. See the online supplement for a complete list of the attributes considered and a summary of the empirical and ethical justifications for each.

To balance information on the shape of the utility* function with statistical efficiency,²⁷ each attribute was assigned 3 levels to allow for the identification of nonlinear preferences over a minimal number of scenarios. The assigned levels were evenly spaced across plausible ranges. The minimum initial life expectancy of 1 month represented imminent death while avoiding implausible combinations of zero life expectancy but positive utility, and the maximum life-year gain was chosen to be plausible in combination with maximum age and initial life expectancy. Distributional concerns were incorporated in the number of patients treated, but assigning appropriate levels was a challenge. An upper limit of 5000 patients was felt to be small enough to be comprehensible but large enough to allow for a meaningful distinction between the high and low levels. Finally, aggregate QALY gains were calculated as a function of the other attributes and included in each alternative to test respondent consistency with the principles of QALYs maximization. The set of attributes and their associated levels are shown in Table 1.

Experimental Design

The number of possible combinations of attributes and levels is given by L^A , where L is the number of levels and A is the number of attributes.²⁷ With 6 attributes of 3 levels each (3^6), there were 729 possible combinations. Illogical combinations where the net QALY gain with treatment was negative were excluded, as were scenarios where health state and life expectancy were unchanged before and after treatment. After these exclusions, there were 594 combinations in the candidate set. To reduce this to a manageable number, a D-efficient fractional

*The conception of "utility" used throughout is consistent with a conventional economic interpretation, where it represents the degree of preference for a particular good or alternative, rather than the narrower interpretation of utility as the degree of preference for a specific health state more common in health economic evaluations.

factorial design was developed using published SAS macros.²⁸ Briefly, a D-efficient design seeks to maximize statistical efficiency—the precision of the parameter estimates for a given sample size—by identifying scenarios that minimize the variance-covariance matrix of a prespecified model.^{29–31} In addition to requiring factors such as orthogonality, level balance, and minimal overlap, an efficient design requires utility balance,³² or a roughly equal choice probability for each alternative in each choice set. Paradoxically, this means that an efficient experimental design requires knowledge about the very preferences that are being elicited. The experimental design was able to use preference information derived from a pilot study³³ to improve the utility balance and thus the efficiency of the design. The prespecified model included attribute main effects and an interaction between life-years gained and final health state.

The SAS design macros generated 18 choice sets with 2 alternatives each, which were divided into 2 blocks of 9. The experimental design also included tests of preference stability and nonsatiation, or dominance. In the test of preference stability, the position of the 2 alternatives from a choice set presented in the first half of the elicitation was reversed and they were re-presented in the second half. In the dominance task, a choice set was constructed where one alternative was unambiguously better than the other. The results of these tests were reported for information purposes but were not used to identify or exclude “irrational” respondents from the samples. Recent stated preference research has highlighted that seemingly irrational preferences can be based on rational reasons.^{34,35} Including these tasks resulted in a total of 11 choice tasks in each block.

Data Collection

Two groups were included in the survey. First, an age-sex representative sample of the Canadian population was drawn from an online survey panel maintained by Research Now, a market research firm. Second, a convenience sample of decision-making agents was invited to participate via e-mails and flyers distributed by the pan-Canadian Oncology Drug Review, the Canadian Association of Medical Oncologists, the Canadian Centre for Applied Research in Cancer Control, and provincial cancer authorities in Nova Scotia, Ontario, and British Columbia.

The elicitations were administered via the Internet. To provide a uniform context, respondents were told that the patient groups had different forms

of cancer, but specific diagnoses were not mentioned and the alternatives were presented simply as Program A and Program B. Although labelled alternatives have the advantage of making hypothetical choice tasks more realistic and concrete, respondents may also use such labels to infer information that was not presented or intended as part of the task.³⁶ At the extreme, respondents might ignore tradeoffs between labelled alternatives and make their choices based on their perceptions of the labels alone.³⁷

Respondents were asked to imagine themselves as a societal decision maker responsible for allocating a fixed budget between 2 health care programs. They were told that both programs had the same cost but that the budget was only large enough fund one or the other. The actual cost and the overall budget were not specified. Brief descriptions of each attribute were provided to respondents, along with graphical representations of the individual QALYs gains associated with each alternative. Samples of the DCE task, the attribute descriptions, and the QALY graphs are available in the online supplement.

Choice Analysis

The choice analysis first tested the consistency of respondents' preferences with QALY maximization by counting how many of each respondent's choices maximized aggregate QALY gains. If a substantial proportion of respondents always chose the QALY-maximizing alternative (a dominant preference for aggregate QALY gains), it would imply that the other attribute levels had little or no impact on choices.³⁸ If most responses were not consistent with strict QALY maximization, the second objective was to quantify the equity-efficiency tradeoff for each attribute in terms of the willingness to trade life-year gains for a more equitable distribution of gains.

Choice responses were initially modelled using a pooled multinomial logit (MNL), but a latent class MNL was also tested to allow for unobserved heterogeneity in preferences. Latent class choice models assume that respondents can be categorized into 1 of 2 or more classes that share unobserved characteristics that affect choice. Preferences are assumed to differ between latent classes but to be homogeneous within classes.^{39,40} Different value functions were also tested. A strictly additive value function would imply that the utility derived from a particular attribute was independent of the levels of other attributes, while interacting attributes with life-years gained would imply that as gains derived from a program tend to zero so does utility, regardless of the level of

the other attributes.^{14,41} Within each value function, continuous and dummy-coded parameters were tested. Alternative specifications were compared on the basis of Akaike's information criterion (AIC), which weighs model fit against parsimony by penalizing log-likelihood by the number of parameters in the model.⁴² The specification that minimized AIC was preferred.

The equity-efficiency tradeoff for each attribute was measured in terms of compensating variation (CV), using individual life-years gained (LYg) as the numeraire. For each attribute, CV was estimated for upward and downward changes from the baseline (middle) level and can be interpreted as the willingness to sacrifice individual life-year gains in order to secure greater (lesser) priority for a more (less) desirable level. These foregone life-years would have otherwise accrued to members of the patient group under consideration, not the respondent. Since some scenarios could be much less preferred than the baseline state, it did not appear appropriate to use the usual Small and Rosen model to estimate CV, as this would depend on choice probabilities for scenarios that respondents may have preferred to avoid.⁴³ Instead, CV was calculated in the context of a "state of the world" model, assuming that each change would happen with certainty.⁴⁴

$$CV_{x:LYg} = \frac{1}{\beta_{LYg}} [v^0 - v^1],$$

where β_{LYg} was the coefficient on the life-years gained attribute, or the constant marginal utility of an additional life-year gained, and v^0 and v^1 were the utilities before and after a change in the level of attribute x , respectively.⁴⁴ Confidence intervals around the estimates of CV were calculated using the delta method.⁴⁵

Choice scenarios were also ranked by their predicted utility, allowing for attributes to vary simultaneously. Scenario utilities were calculated by weighting the attribute levels in each scenario by the coefficients derived from the choice model. The probability of choice was calculated relative to a reference scenario with all attributes at their middle level, although this scenario was not actually shown to respondents in the choice tasks:

$$Pr(i|i, ref) = \frac{e^{\beta x_i}}{e^{\beta x_i} + e^{\beta x_{ref}}},$$

where i was a scenario from the experimental design, ref was the reference scenario, and β and x were

vectors of attribute coefficients and levels, respectively.⁴⁶ Equity weights were calculated by dividing the predicted utility of each scenario by the predicted utility of the hypothetical reference scenario.¹⁴

The influence of respondent characteristics on membership in a particular latent class was assessed by transforming the predicted probability of membership in a particular class to the logit scale.⁴⁷ This was regressed against dummy-coded flags indicating agent status, postsecondary graduation, above median income, female, and above age 40. The relative effect of each factor (f) on the overall probability membership in class c ($Pr(c)$) was calculated as $\beta_f[1 - Pr(c)] \cdot Pr(c)$.⁴⁸

The statistical models were analyzed using NLOGIT 4.0, and other analyses and graphical plots were performed using R statistical software (version 2.15.3)⁴⁹ and the ggplot2 package.⁵⁰ A threshold of 0.05 was adopted for most tests of significance, although a threshold of 0.10 was adopted to allow for the most inclusive set of predictive parameters in the choice models.⁵¹ The elicitation and analyses were approved by the University of Sheffield Research Ethics Committee and the Capital Health Research Ethics Board, Halifax, Canada.

RESULTS

Data collection ran from November 2011 to March 2012. A total of 656 questionnaires were completed out of 738 begun, for an 85% overall completion rate: 595 of 640 (93%) from the general population sample and 61 of 98 (62%) from the agent invitations. Among respondents to these invitations, 44 identified as a decision-making agent and the remaining 17 were included with the general population sample. Twenty-one of the 44 (48%) identified as health system decision makers, 15 (34%) as oncology professionals, and 8 (18%) as both. The sample was broadly representative of the Canadian population in terms of age, gender, median income, and postsecondary graduation, although females in the 35- to 44-year age group were overrepresented.

The median completion time in the public sample was 9.5 minutes. Seventy-three percent of all respondents chose the same alternative in the repeated task. Agents were slightly but not significantly less likely than the general public to choose the same alternative in the repeated task (66% v. 73%, $P = 0.41$). Ninety-five percent of all respondents chose the dominant alternative in the test of nonsatiation, but agents were significantly less likely to

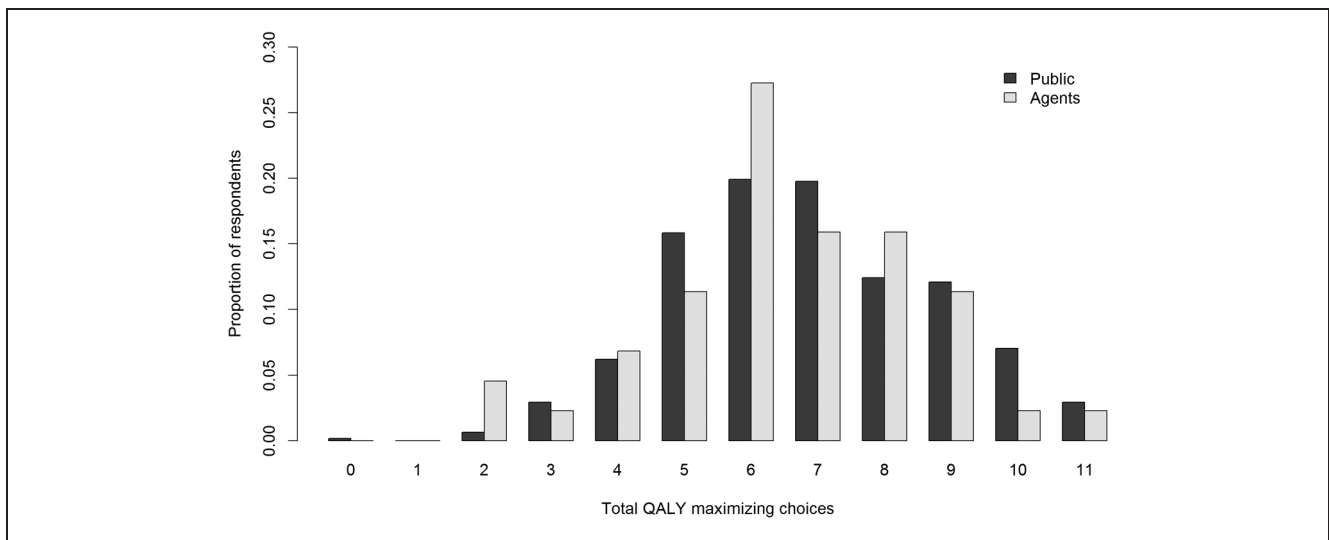


Figure 1 QALY-maximizing choices by public and agent respondents. The number of QALY-maximizing choices made by agents was not significantly different from the number made by the public (6.5 v. 6.8; $P = 0.32$).

choose this alternative than the general public (84% v. 96%, $P = 0.003$).

QALY Maximization

Respondents chose the QALY-maximizing alternative an average of 6.8 times over the 11 choice tasks, and the number of QALY-maximizing choices made by agents was not significantly different from the number made by the public (6.5 v. 6.8; $P = 0.32$). As illustrated in Figure 1, only 2% of agents and 3% of the public chose the QALY-maximizing alternative in every choice task. However, 75% of agents and 74% of the general public made a QALY-maximizing choice in more than half of their tasks, and the average number of QALY-maximizing choices was significantly greater than the 5.5 out of 11 that would be expected by chance alone ($P < 0.001$).

Choice Analysis

A 2-class latent MNL with continuous life-year gains, dummy-coded main effects, and continuous interactions had the best fit by AIC. This was preferred to a pooled MNL despite the penalty associated with the increased number of parameters, suggesting heterogeneity in preferences. The results of this model, overall and by latent class, are shown in Table 2. The positive coefficient on individual life-years gained suggested a preference for larger life-year

gains and supported its use as the numeraire in estimating compensating variation.

The CVs by latent class are detailed in Table 3 and illustrated in Figure 2. They show consistent preferences over age, life expectancy, and worst final health state across the 2 latent classes but significant and offsetting differences over initial health states, best final health state, and the number of patients treated. Respondents had a 48% probability of membership in class 1 and 52% in class 2, but the probability regression found that agents had a statistically significant 33% relative reduction in the overall probability of membership in class 1 (adjusted $P = 0.04$). From an overall class 1 probability of 48%, the probability of an agent belonging to class 1 was therefore 32%, with a corresponding 68% probability of membership in class 2. Agents were distributed between both classes, though, and interacting agent status with specific attributes did not reveal significant differences relative to the public sample. Other respondent characteristics were not significantly associated with class membership.

In interpreting the 2 latent classes, it is of interest to note that confidence intervals in latent class 2 were substantially tighter than the corresponding intervals for class 1 and that agents were statistically more likely to belong to latent class 2. As it seems reasonable to expect that agents might be more familiar with their preferences over the attributes tested here than the public, this suggests that the defining latent

Table 2 Overall and Latent Class Coefficients

Parameter	Overall Coefficients	Latent Class 1 Coefficients	Latent Class 2 Coefficients
Alternative-specific constant	-0.13 (0.18)	-0.67 (0.39) ^a	0.37 (0.06) ^b
Life-years gained (LYg)	0.28 (0.06) ^b	0.28 (0.13) ^c	0.29 (0.04) ^b
Age 10	1.98 (0.86) ^c	4.81 (1.64) ^d	-0.63 (0.14) ^b
Age 70	-1.57 (0.59) ^d	-4.43 (1.08) ^b	1.07 (0.16) ^b
Initial utility 0.1 (U0)	-2.33 (0.76) ^d	-4.13 (1.47) ^d	-0.67 (0.12) ^b
Initial utility 0.9 (U0)	2.09 (0.47) ^b	1.48 (0.95)	2.66 (0.17) ^b
Initial life expectancy 1 month (LE)	-1.32 (0.36) ^b	-2.01 (0.68) ^d	-0.68 (0.08) ^b
Initial life expectancy 10 years (LE)	0.53 (0.22) ^c	1.10 (0.46) ^c	0.01 (0.07)
Final utility 0.1 (U1)	0.85 (0.26) ^d	0.60 (0.52)	1.07 (0.11) ^b
Final utility 0.9 (U1)	-1.86 (0.64) ^d	-3.01 (1.26) ^c	-0.81 (0.17) ^b
100 patients (nPats)	-0.03 (0.15)	0.33 (0.28)	-0.36 (0.11) ^d
5000 patients (nPats)	1.40 (0.49) ^d	2.89 (0.97) ^d	0.01 (0.11)
(1-U0):U1	8.75 (2.08) ^b	8.10 (4.10) ^c	9.35 (0.35) ^b
LYg:Age	0.05 (0.03) ^a	0.18 (0.05) ^d	-0.07 (0.01) ^b
LYg:U0	-0.37 (0.14) ^d	-0.90 (0.26) ^d	0.12 (0.05) ^c
LYg:LE	-0.01 (0.01) ^c	-0.03 (0.01) ^c	0.00 (0.00)
LYg:U1	-0.04 (0.06)	-0.01 (0.12)	-0.08 (0.03) ^c
LYg:nPats	-0.02 (0.01)	-0.04 (0.03) ^a	0.01 (0.01)
Overall probability of class membership		48%	52%

Note: Latent class coefficients are based on the latent class coefficients weighted by the individual probabilities of class membership. Values in parentheses are standard errors.

a. $P < 0.10$. b. $P < 0.001$. c. $P < 0.05$. d. $P < 0.01$.

Table 3 Compensating Variations (CVs) by Attribute Change and Latent Class

Attribute Change	Class 1 CV	Class 2 CV	Difference, Class 1 – Class 2
Patient age, 40 → 10	-7.41 ^a (-13.80, -1.02)	-1.67 ^a (-2.38, -0.97)	-5.74 (-12.18, 0.71)
Patient age, 40 → 70	6.06 ^a (0.15, 11.97)	0.14 (-0.39, 0.68)	5.91 (-0.09, 11.92)
Initial health state, 0.5 → 0.1	2.58 ^a (0.24, 4.91)	-3.35 ^a (-4.17, -2.52)	5.92 ^a (3.54, 8.30)
Initial health state, 0.5 → 0.9	7.02 ^a (0.68, 13.37)	-3.54 ^a (-4.21, -2.86)	10.56 ^a (4.24, 16.88)
Life expectancy, 5 years → 1 month	4.72 ^a (1.54, 7.91)	2.55 ^a (1.93, 3.17)	2.18 (-1.12, 5.47)
Life expectancy, 5 years → 10 years	-1.38 (-2.34, -0.43)	-0.23 (-0.50, 0.03)	-1.15 (-2.18, -0.12)
Final health state, 0.5 → 0.1	3.66 ^a (0.64, 6.69)	2.20 ^a (1.53, 2.86)	1.47 (-1.70, 4.63)
Final health state, 0.5 → 0.9	5.06 ^a (1.72, 8.40)	-3.11 ^a (-4.42, -1.80)	8.17 ^a (4.67, 11.68)
Total patients treated, 2500 → 100	-3.09 ^a (-6.37, 0.19)	1.59 ^a (1.26, 1.93)	-4.69 ^a (-7.99, -1.39)
Total patients treated, 2500 → 5000	-8.51 ^a (-13.97, -3.04)	-0.41 ^a (-0.73, -0.09)	-8.10 ^a (-13.57, -2.63)

Note: Values in parentheses are 95% confidence intervals.

a. Statistically significant CVs and differences between latent classes.

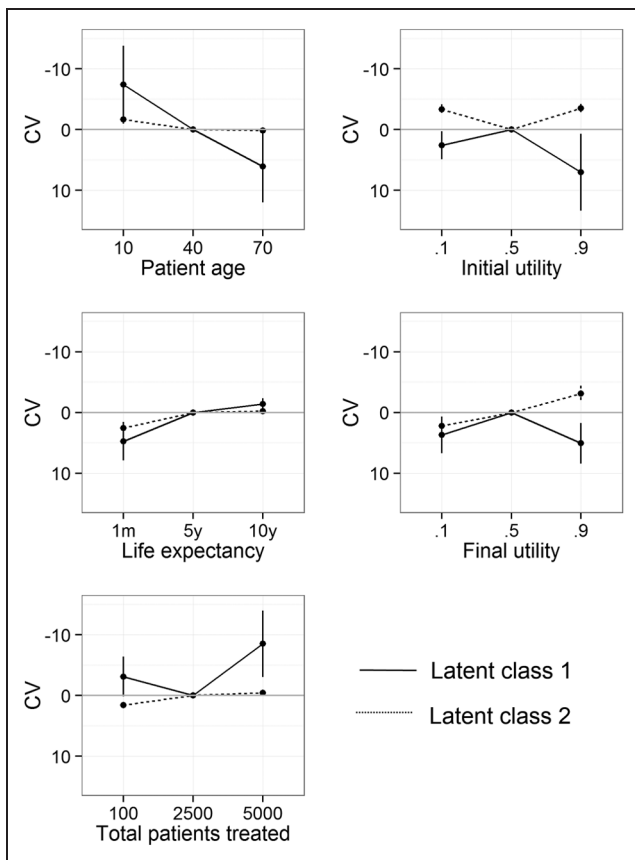


Figure 2 Compensating variations by attribute and latent class. Figures illustrate the compensating variation associated with change from the baseline (middle) level of each attribute. The y-axes have been reversed to show more preferred levels above the baseline and less preferred levels below the baseline.

characteristic of the classes may have been the relative consistency and axiomatic quality of preferences (i.e., completeness, stability, and transitivity), rather than the relative direction or intensity of preferences.

Estimates of overall CV, weighting class-specific estimates of CV by the probability of being in each

latent class, are shown in Table 4 and illustrated in Figure 3. These estimates suggested significant and positive welfare effects associated with prioritizing younger and larger patient groups and patients with the greatest initial life expectancy. Conversely, there were significant and negative welfare effects associated with prioritizing older patients, those with the shortest initial life expectancy and the worst final health state. There were no significant welfare effects over the levels of initial health state or for the best final health state or the smallest patient groups relative to baseline levels.

In the ranking of scenarios by utility and choice probability (available in the online supplement), scenarios with larger individual QALY gains and individual life-year gains tended to be more favorably ranked. The 5 highest ranked scenarios had individual QALY gains in the top 20% of all scenarios, and 3 of the top 5 scenarios had aggregate QALY gains in the top 10%. Likewise, 7 of the bottom 10 scenarios had aggregate QALY gains in the bottom 20% across all scenarios. However, 4 scenarios among the top 10, all presenting 10-year-old patients, had aggregate QALY gains well below the median, and the 2 scenarios with the largest and the second largest aggregate QALY gains, in both cases accruing to 70-year-old patients, were ranked 8th and 29th, respectively, out of the 38 scenarios. Relative to the baseline weight of 1.0, scenario equity weights ranged from 1.77, indicating that that the most preferred scenario was 77% more valuable to respondents than the reference scenario, to negative for the least preferred group.

DISCUSSION

Recent elicitations have consistently found that greater life-year or QALY gains are preferred to less, but they have differed in their conclusions regarding societal support for QALY maximization. Some^{14,17,18,52} have concluded that there is

Table 4 Overall Compensating Variations (CVs) by Attribute Change

Attributes	Attribute Levels	CV, Baseline → Low	CV, Baseline → High
Patient age	10 years old—40 years old—70 years old	−4.36 ^a (−7.45, −1.26)	2.91 ^a (0.91, 4.91)
Initial health state	0.1—0.5—0.9	−0.57 (−1.63, 0.48)	1.41 (−0.55, 3.36)
Life expectancy	1 month—5 years—10 years	3.57 ^a (1.82, 5.32)	−0.77 ^a (−1.30, −0.25)
Final health state	0.1—0.5—0.9	2.88 ^a (1.34, 4.43)	0.71 (−1.27, 2.69)
Total patients treated	100—2500—5000	−0.60 (−2.03, 0.83)	−4.20 ^a (−6.55, −1.86)

Note: Values in parentheses are 95% confidence intervals. CVs are shown for a change away from the baseline (middle) level, holding all other attributes at their baseline level.

a. Statistically significant CVs.

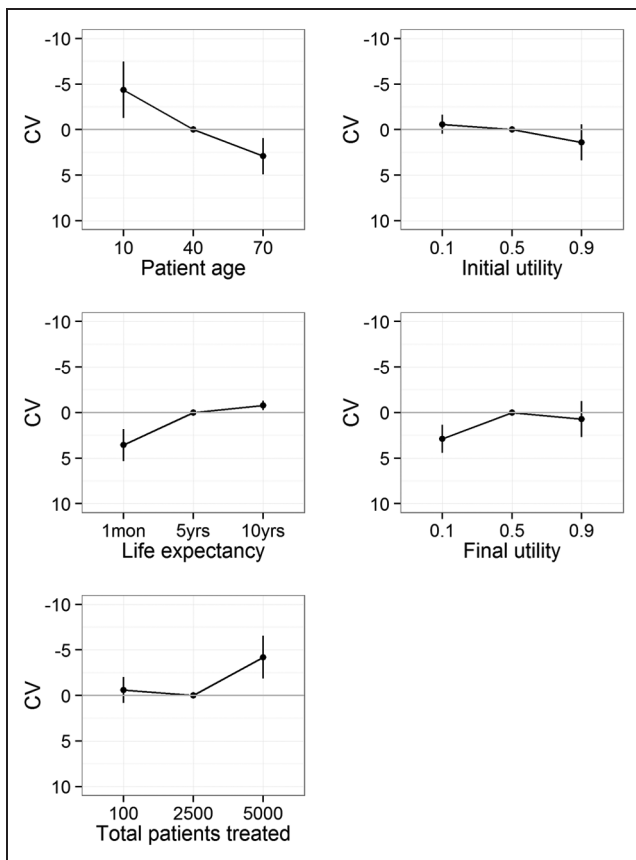


Figure 3 Compensating variations by attribute, combined latent classes. Graphs illustrate the compensating variation associated with change from the baseline (middle) level of each attribute. The y-axes have been reversed to show more preferred levels above the baseline and less preferred levels below the baseline.

a willingness to forego some potential health gains to prioritize particular patient characteristics, rejecting strict QALY maximization. Others^{15,16,19} have concluded that observed choices are broadly consistent with a QALY-maximizing decision rule. Respondents to the DCE reported here tended to favor scenarios with larger QALY gains, but they were also willing to give higher or lower priority to patient groups with particular characteristics, suggesting some preference for distributive justice over strict QALY maximization.

Bryan and others¹⁹ suggested that a majority of respondents maximizing QALYs gained over a majority of their choices may be sufficient to justify a QALY-maximizing decision rule, and by this standard there was some support, as a majority of respondents maximized QALYs in a majority of their choices. The scenario rankings, which allowed attributes to vary simultaneously, also suggested that

respondents were consistent with QALY maximization to the extent that scenarios with larger aggregate QALY gains tended to be ranked more favorably. However, respondents were also willing to prioritize smaller QALY gains to preferred patient groups, particularly younger patients, while older patients and those in the best initial health state were often ranked less favorably, even when their gains were relatively large. This was reflected in the equity weights, where the relative value of a scenario was not proportional to its aggregate QALY gains. Together these suggest a violation of the presumption of distributive neutrality underlying QALY maximization. The relatively strong correlation between rank and aggregate QALY gains despite this apparent tradeoff appeared to reflect a “nested” ordering of preferences. Overall, respondents appeared willing to sacrifice QALY gains to prioritize younger patients, but *within* a particular age level, scenarios with larger QALY gains were preferred to scenarios with smaller QALY gains. Therefore, the preference for larger QALY gains appears to be nested within higher-order preferences for particular patient groups.

The notion of an equity-efficiency tradeoff was supported by the marginal analyses, which found statistically significant preferences for younger and larger patient groups and for greater individual life-year gains. Despite preferences for larger patient groups and greater individual life-year gains, however, the interaction between these two terms was not significant, suggesting that preferences for these factors were not related to a preference for aggregate life-year gains. Indeed, this interaction was negative and statistically significant in latent class 1, suggesting diminishing returns to aggregate life-years gained. Instead, it appeared to reflect a desire to distribute health gains to as many beneficiaries as possible, even if each beneficiary received a smaller gain than if benefits were concentrated among a smaller group. The unexpected overall aversion or indifference to initial severity may have reflected respondent expectations about prognosis or ability to benefit that did not take the presented life-year gains into account. There was also evidence of a reluctance to allocate resources to patients who would remain in a poor health state following treatment, although this did not translate into a preference for patients in the best final health state. This is consistent with an emphasis on achieving some minimum level of quality in the posttreatment health state rather than maximizing the quality of that health state,^{7,53} casting doubt on the presumption of a strictly linear relationship between value and changes in health utility.

Interestingly, the statistical model showed that attribute main effects were significant even after including interactions with life-years gained. This suggests that there may be value associated with treating particular patient groups even in the absence of health gains, consistent with arguments that society may desire a health care system that provides for aspects such as compassion, respect for dignity, and maintenance of hope, in addition to health gains.^{54–56} This possibility was not specifically tested here, but future research would be useful in identifying nonhealth factors that might be associated with societal welfare gains.

The use of the latent class approach in the marginal analysis appeared unique among preference elicitation in health care and allowed for an interpretation of the responses in terms of individual choices as well as personal characteristics. The interpretation of different groups of respondents in terms of their preference consistency and axiomatic quality, for example, would not be possible with a conventional random parameters approach. In a similar vein, latent class methods have been used to identify inattentive respondents⁵⁷ and could help control for differences in the scale, or variance, among different sets of respondents.⁵⁸ The latent class approach also allowed respondents with the same characteristics to be probabilistically categorized into different classes on the basis of their observed choices, whereas a classic interaction approach would presume that everyone with the same observed characteristics must share the same preference for the interacted attribute.^{40,59}

The analysis had a number of limitations. First, the empirical ethics review was unavoidably subjective, both in the interpretation of the different theories of justice and in judging the consistency of each attribute with these theories. Any application of empirical ethics, though, can be seen as a balance between a more objective interpretation of the empirical evidence—which leaves the process open to a charge of moral relativism—and more a subjective interpretation of competing theories of justice. The process described here favored ethical subjectivity over empirical relativity. Second, the final model specification was more complex than the model specified at the experimental design stage. This likely introduced some correlation between the parameters, but as the more complex model had a better fit than the prespecified model, it can be seen as sacrificing some precision in individual parameter estimates for greater overall explanatory power. Third, the elicitation were conducted in a cancer context,

potentially limiting their generalizability. There are suggestions of a “cancer premium” relative to other diseases,^{52,60} although this was arguably negated here by the fact that both groups in each task had cancer, and no specific diagnoses were presented that might lead respondents to favor one group over the other. As a result, we would expect the tradeoffs observed here to hold in other disease contexts. However, it is still possible that there may be some interaction effects, where respondents valued particular attributes or levels differently because of the cancer diagnoses. Finally, although the calculated equity weights were consistent with the notion of an equity-efficient tradeoff, the appropriate calculation for such weights is unsettled.^{14,15} Different methods of calculation, as well as of eliciting respondent preferences,⁶¹ tend to produce different weights, and as such, it would be premature to use these weights in priority-setting. As noted by Norman et al,¹⁴ though, some of the differences observed between studies may stem from differences in the attributes included in the elicitation. In this respect, greater use of empirical ethics, as used here, may help to standardize the attributes over which preferences are elicited and improve the consistency of equity weights.

In conclusion, these results appear to reject strict QALY maximization as a preferred societal decision-making rule, as well as the presumption of distributive neutrality. Instead, societal well-being might be improved by giving relatively greater priority to programs that best satisfy societal preferences, although more research is required before explicit equity weights can be used in priority-setting.⁶²

ACKNOWLEDGMENT

The authors acknowledge the valuable comments of 3 anonymous reviewers and the assistance of Victoria Joy-Warren in preparing the tables and figures.

REFERENCES

1. Culyer AJ. The normative economics of health care finance and provision. *Oxford Rev Econ Policy*. 1989;5(1):34.
2. Coast J, Smith R, Lorgelly P. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. *Soc Sci Med*. 2008;67(7):1190–8.
3. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ*. 2005;14(2):197–208.

4. Nord E, Richardson J, Street A, Kuhse H, Singer P. Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Soc Sci Med*. 1995;41(10):1429–37.
5. Kaldor N. Welfare propositions of economics and interpersonal comparisons of utility. *Econ J*. 1939;49(195):549–52.
6. Coast J. Maximisation in extra-welfarism: a critique of the current position in health economics. *Soc Sci Med*. 2009;69(5):786–92.
7. Schwappach DL. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expect*. 2002;5(3):210–22.
8. Stafinski T, Menon D, Marshall D, Caulfield T. Societal values in the allocation of healthcare resources: is it all about the health gain? *The Patient*. 2011;4(4):207–25.
9. Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ*. 1991;10(1):21–41.
10. Sassi F, Archard L, LeGrand J. Equity and the economic evaluation of healthcare. *Health Technol Assess*. 2001;5(3):1–138.
11. Mooney G. “Communitarian claims” as an ethical basis for allocating health care resources. *Soc Sci Med*. 1998;47(9):1171–80.
12. Fleck LM. Just caring: health reform and health care rationing. *J Med Philos*. 1994;19(5):435–43.
13. Broqvist M, Garpenby P. To accept, or not to accept, that is the question: citizen reactions to rationing. *Health Expect*. 2014;17(1):82–92.
14. Norman R, Hall J, Street D, Viney R. Efficiency and equity: a stated preference approach. *Health Econ*. 2013;22(5):568–81.
15. Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. *J Health Econ*. 2011;30(2):466–78.
16. Green C, Gerard K. Exploring the social value of health-care interventions: a stated preference discrete choice experiment. *Health Econ*. 2009;18(8):951–76.
17. Dolan P, Edlin R, Tsuchiya A, Team NSQ. The relative societal value of health gains to different beneficiaries. *Health Economics and Decision Science Discussion Paper Series*; 2008. Available from: URL: http://eprints.whiterose.ac.uk/11213/1/HEDS_DP_08-12.pdf
18. Schwappach DL. Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. *Health Econ*. 2003;12(4):255–67.
19. Bryan S, Roberts T, Heginbotham C, McCallum A. QALY-maximisation and public preferences: results from a general population survey. *Health Econ*. 2002;11(8):679–93.
20. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Soc Sci Med*. 1999;48(4):535–46.
21. Richardson J, McKie J. Empiricism, ethics and orthodox economic theory: what is the appropriate basis for decision-making in the health sector? *Soc Sci Med*. 2005;60(2):265–75.
22. Hausman DM. The limits to empirical ethics. In: Murray CJL, Salomon JA, Mathers CD, eds. *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva: WHO; 2002. p 663–8.
23. Rawls J. *A Theory of Justice*. Cambridge (MA): Belknap Press of Harvard University Press; 1999.
24. Parfit D. Equality and priority. *Ratio*. 1997;10(3):202–21.
25. Arneson RJ. Luck egalitarianism and prioritarianism. *Ethics*. 2000;110(2):339–49.
26. Olsen JA, Richardson J, Dolan P, Menzel P. The moral relevance of personal characteristics in setting health care priorities. *Soc Sci Med*. 2003;57(7):1163–72.
27. Hensher DA, Greene WH, Rose JM. *Applied Choice Analysis: A Primer*. Cambridge (UK): Cambridge University Press; 2005.
28. Kuhfeld WF. Marketing research methods in SAS. Available from: URL: http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html
29. Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. *Health Econ*. 2003;12(4):281–94.
30. Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental design with marketing research applications. *J Mark Res*. 1994;31(4):545–57.
31. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3–13.
32. Huber J, Zwerina K. The importance of utility balance in efficient choice designs. *J Mark Res*. 1996;33(3):307.
33. Skedgel CD, Wailoo AJ, Akehurst RL. Choosing vs. allocating: discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences. *Health Expect*. Published online June 12, 2013.
34. Miguel FS, Ryan M, Amaya-Amaya M. “Irrational” stated preferences: a quantitative and qualitative investigation. *Health Econ*. 2005;14(3):307–22.
35. Ryan M. Rationalising the “irrational”: a think aloud study of discrete choice experiment responses. *Health Econ*. 2009;18(3):321.
36. De Bekker-Grob EW, Hol L, Donkers B, et al. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value Health*. 2010;13(2):315–23.
37. Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, eds. *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht (The Netherlands): Springer; 2008. p 13–46.
38. Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. *J Econ Psychol*. 2002;23(3):383–98.
39. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res Part B Methodol*. 2003;37(8):681–98.
40. Boxall PC, Adamowicz WL. Understanding heterogeneous preferences in random utility models: a latent class approach. *Environ Resour Econ*. 2002;23(4):421–46.
41. Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJM. The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. *Manag Sci*. 1998;44(6):839–49.
42. Swait JD. Advanced choice models. In: Kanninen BJ, ed. *Valuing Environmental Amenities Using Stated Choice Studies*. Dordrecht (The Netherlands) Springer; 2007. p 229–93.
43. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. *Health Econ*. 2004;13(9):901–7.

44. Ryan M. Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1). *Health Econ.* 2004; 13(9):909–12; discussion 919–24.
45. Oehlert GW. A note on the delta method. *Am Stat.* 1992;46(1): 27–9.
46. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York: Academic Press; 1974. p 252.
47. Clark S, Muthen B. Relating latent class analysis results to variables not included in the analysis. Available from: URL: <http://statmodel2.com/download/relatinglca.pdf>
48. Gujarati DN. *Basic Econometrics*. New York: McGraw-Hill; 1988.
49. R Core Team. R: a language and environment for statistical computing. Available from: URL: <http://www.R-project.org/>
50. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009.
51. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: Wiley; 2000.
52. Linley WG, Hughes DA. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Econ.* 2013;22(8):948–64.
53. Cookson R, Dolan P. Principles of justice in health care rationing. *J Med Ethics.* 2000;26(5):323–9.
54. Donaldson C, Shackley P. Does “process utility” exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med.* 1997;44(5):699–707.
55. Wiseman V. Caring: the neglected health outcome? Or input? *Health Policy.* 1997;39(1):43–53.
56. Mooney G. Beyond health outcomes: the benefits of health care. *Health Care Anal.* 1998;6(2):99–105.
57. Lagarde M. Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health Econ.* 2013;22(5):554–67.
58. Swait JD, Louviere JJ. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res.* 1993;30(3):305–14.
59. Morey E, Greer Rossmann K. Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: classic heterogeneity, random parameters, and mixture models. *J Cult Econ.* 2003;27(3):215–29.
60. Viscusi WK, Huber J, Bell J. Assessing whether there is a cancer premium for the value of a statistical life. *Health Econ.* 2014;23(4): 384–96.
61. Oliver A. Testing procedural invariance in the context of health. *Health Econ.* 2013;22(3):272–88.
62. Wailoo DA, Tsuchiya A, McCabe C. Weighting must wait. *Pharmacoeconomics.* 2009;27(12):983–9.