Health system goals: A discrete choice experiment to obtain societal valuations☆,☆☆

Margreet Franken [a],*, Xander Koolman [b]

[a] Institute of Health Policy and Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, The Netherlands
[b] Department of Health Sciences, VU University Amsterdam, De Boelelaan 1085-1087, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

Objective: To improve previous approaches to health system goals valuation.
Methods: We reviewed literature on health system performance and previous comparative performance assessments, and combined this with literature on process utility to create a theoretical foundation for health system goals. We used a discrete choice experiment to elicit goal weights. To obtain social justice weights respondents were placed behind a 'veil of ignorance'. To ensure that respondents understood their task, we instructed them in a classroom setting.
Results: We identified five health system goals. All five goals significantly affected choice behavior. An equitable distribution of health obtained the highest weight (0.34), followed by average level of health (0.29) and financial fairness (0.24). Both process outcomes (utility derived from the process and its distribution) received much lower weights (0.07 and 0.06, respectively).
Conclusions: Our framework adds to that of the World Health Organization. We demonstrated the feasibility of measuring societal valuation of health system goals with a multi-attribute technique based on trade-offs. Our weights placed much greater emphasis on health and health inequality than on process outcomes. Our study improves the methodology of international health system performance comparison and thereby enhances global evidence-based health policy information.

1. Introduction

Health systems around the world have contributed to better health and life expectancy with varying degrees of success. Even in countries with seemingly similar resources outcomes vary markedly [1]. To date, policy effects on the performance of health systems remain largely unclear. Monitoring and evaluating performance can generate this vital policy information. Moreover, cross-country comparisons enable countries to learning from others.

The challenge, however, is how to assess health systems that are extremely complex and have multi-dimensional goals. This complex task has been explored by the World Health Organization (WHO) and the Organisation for Economic Co-operation and Development (OECD).

Our aim is to improve the valuation of health system goals. To do so, we first unify literature on health system goals, equity, and process utility to create the underpinnings of a theoretical framework for health system evaluation. Second, we review previous approaches to deriving relative weights for health system goals. Third, we suggest an enhanced methodology based on a multi-attribute choice technique to elicit goal valuations using a

'veil of ignorance' perspective. Last, we present goal valuations for the Netherlands based on our proposed method.

2. Theoretical framework for health system goals

Until the 1990s health economics was dominated by the assumption that 'health' was the dominant outcome of health systems. This links with the consequentialism moral theory, which focuses solely on outcomes irrespective of the process that led to them, and Jeremy Bentham's 'act utilitarianism', which states that the greatest amount of happiness for the greatest number of people determines choice behavior. Although utility derived from health is an obvious outcome of a health system, research showed that people also care about the processes that precede health outcomes, irrespective whether they affect health [2–4]. Therefore, processes are not just means to an end, not just instrumental to an intrinsic goal, but are an intrinsic goal of the health system.

This utility derived from processes, procedural utility, has a base in social sciences. Parsons' social action theory (1937) already described the necessity of the subjective dimension of human action [5]. Psychologists have developed a comprehensive notion of basic psychological needs for the human self, evident in the "self-determination theory of intrinsic motivation" by Deci and Ryan [6,7]. The theory maintains that human motivation originates from three innate needs: autonomy, competence, and relatedness; individual well-being therefore depends on procedures that address them [6,7].

The theory of procedural utility can be directly applied to health care. Consequently, both health outcomes and the process attributes of non-health outcomes are health system goals. Furthermore, it is widely recognized that health systems' costs should be related to capacity to pay rather than the risk of illness [8]. Therefore, health systems have three independent outcome-oriented objectives: health utility, process utility, and financial fairness.

2.1. Health utility

Health systems aim to improve health and strive for the highest possible health status of the entire population, taking both morbidity and mortality into account. Behind a 'veil of ignorance' the distribution of health also matters; empirical evidence indicates that the public is willing to trade efficiency for social objectives such as equity [9–11]. Therefore, health utility consists of two goals: average level of health and the equitable distribution of health.

2.2. Process utility

Procedural utility can arise from two sources [12]. First, interaction between people can generate utility since people evaluate actions by how they are treated by others. Second, people have preferences for good institutions in addition to health outcomes (e.g. preferences on allocative and redistributive decisions) that address the innate needs of human motivation (autonomy, competence, and relatedness). Institutions also establish the fundamental rules for societal decision making. As a result, process attributes of health systems are twofold: utility derived from interaction between people and the health system ("how people are treated by the health system"), and utility obtained from living under institutions ("how allocative and redistributive decisions are taken"). Although distributional fairness of process utility is not well founded in moral theory, we followed the WHO framework and therefore included both process utility and its distribution in our framework.

2.3. Financial fairness

Murray et al. [13] claim that a health system is fairly financed "if the ratio of total health system contribution of each household through all payment mechanisms to that household's capacity to pay is identical for all households, independent of the household's health status or use of health system." This signifies two key challenges. First, households should not pay an excessive share of their income for health care or become impoverished [14]. Second, wealthy households should contribute more than poor households reflecting vertical equity and an element of progressivity.

3. Existing international frameworks

Several countries, such as the USA, United Kingdom, the Netherlands, Australia, and Canada, have designed and implemented national schemes and indicators to measure health system performance [15]. Cross-country comparison, however, requires a comprehensive international framework such as those of the OECD and WHO.

OECD framework. The three main goals of the OECD framework are (i) health improvement and outcomes, (ii) responsiveness and access, and (iii) financial contribution and health expenditure [16]. Without suggesting any relative importance of the system goals it provides a framework to measure performance in several dimensions that seem to be based on the historical development of health systems. A composite score requires, however, each goal to be independent. The OECD framework consists of input and output variables, and intermediate as well as end goals. Consequently, using the OECD framework gives rise to methodological problems when weighing goals.

WHO framework. The WHO framework for performance measurement consists of three intrinsic goals of health systems: health, responsiveness, and fairness in financing [1]. The first two are assessed on both level and fairness of the distribution. The framework satisfies the required conditions (i.e., a complete set of intrinsic goals) to facilitate global performance assessment.

WHO's health system goals closely resemble those identified for our own theoretical framework. Health utility and its distribution are reasonably comparable to WHO's level and the distribution of health. Our two sources of process utility can be described by WHO's assessment of quality and equity of responsiveness. Last, one could suggest that financial fairness reflects WHO's fairness in financing.

4. Valuing health system goals

4.1. Previous approach

WHO's goal weights were acquired by measuring preferences of individuals with health system knowledge via an internet-based questionnaire, which included interactive, weight-assigning pie charts, descriptive multiple choice questions, and ranking tasks. The final weights were rounded to the nearest one-eighth for the World Health Report 2000 (WHR) to make the composite goal easier to understand [17]. In 2000 and 2001, the WHO performed a follow up multi-country study to measure preferences from the public [18]. These questionnaires consisted of ranking tasks and pie charts. The Dutch sample included 1566 respondents: 1068 face-to-face and 498 postal interviews [18].

4.2. Critiques on the WHO approach

Much critical attention and debate followed the publication of the WHO results. One of the major concerns was that WHO's relative weights were highly subjective since they were derived from respondents who were far from representative [19–21]. The WHO justified its method by stating: "the purpose of the first survey was not to describe preferences in a population, but rather empirically derive a set of weights reflecting normative choices" [22].

Richardson et al. [23] maintained that effective weights depend on variation in scores across countries as well as the nominal weights. Consequently, if there is no difference in, for example, health inequality then, regardless of the weight of 0.25, it would contribute nothing to the ranking scores. However, applying a standard set of weights appears to deny differences between countries' ideologies and theories of social justice.

Moreover, the Oswaldo Cruz Foundation [24] concluded that the composite index is very sensitive to modification in the relative weights and claimed that some countries can shift the scale by more than thirty points by small weight adjustments. On the contrary, Lauer et al. [25] concluded that even large changes would have a small impact and that all rankings remained within the uncertainty intervals. Nevertheless, Lauer et al. [26], in yet another publication, argued that the differences could be large for individual countries because of the impact of publication of the rankings.

The most essential critiques concern the valuation methodology. Smith [21] argued that WHO's methodology was highly questionable and that it is unlikely that it would elicit the required relative marginal valuation of an extra unit of performance. Moreover, Williams [20] claimed that the main issue was the use of rankings, scores, and rating scales rather than facing respondents directly with trade-offs.

Making trade-offs is at the heart of economics. In a multi-attribute environment, such as a health system, individuals choose between attributes based on their relative importance. The WHO limited their survey by using pie-charts, rankings, and descriptive multiple choice questions; their instrument did not allow for deliberation.

4.3. Enhanced methodology

The Multi Attribute Utility Theory (MAUT) provides a foundation for valuing complex and multi-dimensional objectives. It has its roots in classical measurement theory and theories of economic choice behavior. In the last decade, multi-attribute valuation methods and especially discrete choice experiments (DCEs) have become increasingly popular. DCE draws upon Lancaster's economic theory of value [27] and is firmly rooted in the random utility theory [28].

Furthermore, Dolan et al. [29] distinguish respondents' perspectives in two dimensions: who the respondent should have in mind and at what point in time. The former dimension concerns oneself, other people, or all people; that is, preferences are personal, social, or socially inclusive, respectively. The time dimension relates to the context in which the valuation is obtained, that is, whether it is *ex ante* or *ex post*. The structure of a health system depends in large part on society's choices concerning resource allocation; relatedly, a person's societal position often reflects a feeling of social (un)fairness. Therefore, it is essential to elicit *ex ante* socially inclusive personal preferences for social justice valuations. Valuations should therefore be obtained in a procedurally fair way reflecting a social justice perspective; i.e., respondents assume a Rawlsian "veil of ignorance". The main characteristic of Rawls's approach is that people make choices without knowing where their own position might be in a society [30]. The "veil of ignorance" ensures that the principles of justice are blind to age, health status or societal position.

An enhanced methodology should therefore be based on a multi-attribute choice technique such as a DCE. This technique applies direct trade-off questions, improves conscientious deliberation, and elicits marginal valuations. To ensure social justice valuations, goal valuations should be obtained using an *ex ante* perspective.

5. Methods

We conducted a DCE to obtain goal valuations in the Netherlands. Our five-step procedure, typical of a DCE, was to: (1) identify and describe the attributes for health systems; (2) assign attribute levels based on goal variation; (3) combine attribute levels and create hypothetical scenarios; (4) establish goal valuations; and (5) analyze and interpret the data.

5.1. Attributes and levels

Our health system goals satisfy the attribute criteria by being complete, operational, decomposable, non-redundant and minimum-sized [31]. For use in a questionnaire, attributes must be meaningful, relevant and easy to understand, and their levels should be plausible, actionable and tradable [32]. Variation in their levels should mimic existing variation.

We pilot-tested attribute and level descriptions in 54 participants [data not shown]. They were interviewed face-to-face or by telephone and were asked to fill out a questionnaire while thinking aloud. Pilot questionnaires

Table 1
Attributes and descriptions of attributes and levels used in the DCE exercise.

Attribute	Attribute description	Level description
Average level of health	Average health adjusted life expectancy	74, 72, 70 and 68 years
Distribution of health	Differences in health adjusted life expectancy across social groups	3, 5, 7 and 9 years difference
Average level of process outcome	Patient experiences	very good experiences, good experiences, rather good experiences, reasonable experiences
Distribution of process outcome	Differences in experiences between patients across social groups	no difference, small difference, some difference and fair difference
Financial fairness	Persons in poverty because of health system payments	0, 1, 2 and 3 person(s) per 200 persons

offered 30 questions comprising paired comparison of two attributes in which vignettes varied by one level. The interviewer could ask about the basis for respondents' trade-offs and the extent to which they understood the descriptions. Answers provided insight into respondents' ways of thinking and revealed any misinterpretation or misunderstanding of the notions of the attributes. Respondents initially found it difficult to make a clear distinction between the intrinsic and instrumental contributions of process factors. The pilots also showed that respondents changed their behavior when choosing from behind a veil of ignorance, some explicitly mentioning that equity was more of a concern. Preferences concerning solidarity versus individualism were partly balanced out. Table 1 shows the attributes and the final descriptions of the attributes and levels which were used in the DCE exercise.

Health outcome. Utility obtained from health outcome reflects the average level of health in a population expressed by life expectancy that takes both morbidity and mortality into account. The levels are actual Health-Adjusted Life Expectancy (HALE) years in the Netherlands: 74, 72, 70 and 68 years.

Distribution of health outcome. We assessed existing inequalities in HALE in the Netherlands. Inequality is present across social classes, educations groups, income levels, ethnic groups, gender and geographic areas. The differences across social groups in HALE were 3, 5, 7 and 9 years.

Process outcome from interactions. Because the pilots revealed that respondents experienced difficulties in understanding the process-outcome concept, the concept and its implications were clarified and illustrated with recognizable examples from the Dutch health system during the DCE exercise. The levels in experience were very good, good, rather good, and reasonable. This variation mimics the actual variation in patient experiences in the Netherlands as measured by a population based study of the WHO (average responsiveness score of 85.7 on a scale of 0 to 100) [33].

Distribution of process outcome. Although distributional fairness of process utility is not well founded in moral theory, we followed the WHO framework and allowed our empirical test to indicate the value respondents attached to an unequal distribution of process utility. This attribute thus indicates the inequality in the distribution of process outcome across groups. In the Netherlands options exist to

bypass waiting lists and "buy" quality but the majority of the population sees this as inequitable. The Health Insurance Act has since 2006 been moving toward managed competition in health care and it is thus likely that more options will become available to buy additional insurance or pay out-of-pocket for extra quality and other process attributes in the future. Therefore, our levels are actual differences in distribution and partly reflect potential future developments. The levels of difference offered were none, small, some, and fair. The differences in the levels were clarified and illustrated with recognizable examples from the Dutch health system.

Financial fairness. The last attribute describes financial protection against costs of illness. We explored national health insurance premiums and potential co-payments, and connected these financial flows with catastrophic payment and impoverishment. We selected a quasi-relative poverty measure found to be recognizable and accurate for the core perception of poverty within the Dutch population [34]. We expressed the levels as consequences of variations in (co-) payments. The levels offered were 0, 1, 2 and 3 extra person(s) per 200 persons unable to satisfy basic needs because of health system payments.

5.2. Experimental design

A full factorial design would generate 1024 (4^5) scenarios. To reduce the DCE exercise to a manageable level, we applied a fractional factorial design. We obtained an orthogonal array from the online Sloan library in order to assign 16 choice sets (http://www.research.att.com/~njas/oadir). We applied an optimal design generator (12332, 21123 and 33211) based on strategies described by Street et al. [35]. Such a fold-over strategy provides an optimal design with a high efficiency for estimating main effects while satisfying the statistical properties of level balance, orthogonality, minimal overlap, and utility balance [32]. Main effects usually account for 70–90% of explained variance [28]. The order of vignettes was randomly varied for all three choice sets.

We did not include an opt-out option because (i) it can generate problems such as applying heuristics to prevent making difficult choices or preferring a status quo and (ii) it is impossible to opt-out by choosing not to have any health system. Our respondents were choosing from behind a veil of ignorance and thus we assumed that there was no default system, i.e., an opt-out option was non-existent.

Table 2

Results conditional logit model.

Attribute	Coefficient	Std. error	z value	P > |z|	95% Conf. interval	
Average level of health	−1.097	0.067	−16.31	0.000	−1.229	−0.966
Distribution of health	−1.294	0.073	−17.84	0.000	−1.436	−1.152
Average level of process outcome	−0.277	0.058	−4.73	0.000	−0.391	−0.162
Distribution of process outcome	−0.236	0.061	−3.84	0.000	−0.356	−0.116
Financial fairness	−0.927	0.063	−14.64	0.000	−1.051	−0.803
Log likelihood	−880.674					
Chi-Square test	948.75					
Pseudo R^2	0.350					

5.3. Data collection

We selected 63 persons familiar with health systems to ensure that respondents understood their task. Moreover, respondents were instructed by the principal investigator in a classroom setting via a PowerPoint presentation and a 20-minute interactive discussion about the meaning of the attributes and their levels. In particular, the distinction between the intrinsic and instrumental contribution of both process attributes was explained and the levels were made operational by illustrative examples from the Dutch health system. Each respondent was provided with the information on paper to refer to if needed. Respondents had to choose in which country they would prefer to be born.

5.4. Data analysis

The software program STATA was used to perform data analysis. Given the exploratory character and aim of our study to investigate the feasibility to estimate health system goal valuations using a multi-attribute technique based on trade-off questions, we used a conditional logit model to estimate the coefficients. The estimated function for the valuations of health system goals was: Yla-tent = $\beta_0 + \beta_1$* average level of health + β_2* distribution of health + β_3* process outcome + β_4* distribution of process outcome + β_5* financial fairness + ε.

6. Results

6.1. Econometric model

All responses were included in our data analysis since all respondents correctly answered the dominant question. Table 2 presents the results of the conditional logit model, which provides good insight into respondents' trade-off behavior because the outcome probabilities are based only on the attributes and their levels. The summary statistics show that the model had a decent fit with a pseudo R^2 of 0.35 and a statistically-significant Chi-square test of 948.75.

The coefficients reflect the relative importance of the system goals. As expected, they revealed that respondents favor the best attainment in all five goals. The associated p-values indicated that all attributes have a statistically significant effect on choice behavior. The table shows that health distribution received the highest importance followed by average level of health, financial fairness, average

level of process outcome, and distribution of process outcome, respectively.

6.2. Marginal rate of substitution

The Marginal Rate of Substitution (MRS) is computed by dividing the coefficients and demonstrates the trade-off between attributes. Table 3 shows that individuals are willing to give up 0.848 level in health distribution to gain one level in average health (β_1/β_2), meaning that individuals will trade 2 years of average health for 1.70 years of health inequality.

6.3. Valuation as percentage

To enable direct comparison between WHO's relative weights and our valuations, our coefficients needed to be converted into percentages by dividing the coefficient of one attribute by the sum of all coefficients. This method is based on the same assumption as the MRS calculation, that is, linearity of the coefficients and a comparable realistic amount of variation between the levels for each of the attributes. Table 4 shows the results.

Although the sum of the quality (0.36 versus 0.37) and equity (0.64 versus 0.63) objectives are similar, their weights are derived differently. Specifically, our valuations place much greater emphasis on health (0.63 versus 0.49) and far less on process outcomes (0.13 versus 0.29). Furthermore, our results show a greater weight for health distribution (0.34) compared to the average level of health (0.29), whereas these weights were almost similar in WHO's survey (0.25 and 0.24, respectively). Fairness in financing is roughly equally weighted in both methodologies.

7. Discussion

The focus of our research was twofold. First, we identified five health system goals and explored their theoretical foundation. We included the process of health care delivery in a utility framework supporting that the process of care giving is an end goal of health systems, irrespectively whether it affects health, and thus can be traded-off against health outcomes. Second, we obtained valuations for the goals using a multi-attribute technique based on trade-offs. We used actual variation in goal attainment, elicited marginal weights from behind a 'veil of ignorance' reflecting the original position, and created a setting to help respondents understand the concepts and their task.

Table 3
Marginal rate of substitution.

Coefficients	Attribute/Attribute	Observed MRS
β_1/β_2	average level of health/distribution of health	0.848
β_1/β_3	average level of health/average level of process outcome	3.968
β_1/β_4	average level of health/distribution of process outcome	4.651
β_1/β_5	average level of health/financial fairness	1.184
β_2/β_3	distribution of health/average level of process outcome	4.677
β_2/β_4	distribution of health/distribution of process outcome	5.482
β_2/β_5	distribution of health/financial fairness	1.396
β_3/β_4	average level of process outcome/distribution of process outcome	1.172
β_3/β_5	average level of process outcome/financial fairness	0.298
β_4/β_5	distribution of process outcome/financial fairness	0.255

Table 4
Goal valuations as percentages from different sources.

	WHO 1st survey[a]	WHR applied weights[a]	WHO 2nd survey (Netherlands) (face-to-face/postal)[b]	Discrete choice experiment valuations
Average level of health (1)	0.24	0.25		0.29
Distribution of health (2)	0.25	0.25		0.34
Average level of process outcome (3)	0.13	0.125		0.07
Distribution of process outcome (4)	0.16	0.125		0.06
Financial fairness (5)	0.22	0.25		0.24
Health outcomes (1 + 2)	0.49	0.5	0.415	0.63
Process outcomes (3 + 4)	0.29	0.25	0.309/0.306	0.13
Financial fairness (5)	0.22	0.25	0.277/0.279	0.24
Quality (1 + 3)	0.37	0.375	0.376/0.432	0.36
Equity (2 + 4 + 5)	0.63	0.625	0.624/0.569	0.64

[a] Gakidou et al., 2000 [22].
[b] Gakidou et al., 2003 [18].

Consequently, we tested all five goals statistically significantly affect choice behavior independently. An equitable health distribution has the highest valuation, followed by average level of health and financial fairness. Both attributes measuring process utility receive much lower weights.

By assuming a linear additive model for a latent preference variable, it was possible to compare our valuations with WHO's weights. Some might argue that our DCE valuations only provide a rough estimate of actual percentages. Nevertheless, our results are quite different from WHO's weights. Specifically, much weight of both process attributes shifts toward both health attributes. The shift may be due to the fact that we specifically made respondents aware that the intrinsic contribution of process attributes do not directly influence health. The extent to which WHO's respondents were conscious of this distinction is unknown.

Several researchers have claimed that people are willing to sacrifice overall health to achieve a more equitable distribution of health [9,11,26,36]. Our results show this greater weight for health inequality whereas WHO's weights are equal for both health attributes. First, an equitable health distribution attains a higher valuation compared to average level of health. Second, the computed MRS suggests that individuals are willing to trade 2 years of average health for 1.7 years in health inequality. We believe that this might be attributable to our enhanced methodology to derive marginal valuations from behind a veil of ignorance.

The complexity of the attributes and levels forced us to make a trade-off between potential interviewer bias caused by an interactive classroom setting and task simplicity. Furthermore, our respondents were well-educated and familiar with the Dutch health system. It is possible that we introduced selection bias and the goal valuations are not representative of the preferences of the Dutch population. Gakidou et al. [18], however, conclude that WHO's valuations varied only slightly between informed respondents and the population at large. We also obtained goal valuations from behind a veil of ignorance, although whether such a scheme guarantees societal preferences that are genuinely impartial is debatable. Two findings in favor of impartiality was that respondents' equity concerns were altered by the 'original position' due to 'the veil of ignorance' and choice behavior in the pilot study changed when respondents assumed an unknown position in society.

At the methodological level, economists are inclined to consider DCEs superior to ranking and rating since they are based on the random utility theory. Therefore, we argue that the research resulting from our enhanced methodology can be seen as a follow-up to the WHO surveys. We do, however, acknowledge methodological issues of DCEs, most of which are related to human cognitive processes. For example, choice experiments assume that people have stable preferences and are willing to trade between all attributes and a 'veil of experience' can influence decision making through status quo bias and the endowment effect [37]. However, these issues have received much attention within health economics and should not be seen as 'threats' to economic methods of valuation [38].

8. Conclusions

Our study demonstrates the feasibility of measuring health system goal valuations using a multi-attribute technique based on direct trade-off questions. We believe

that our valuations improved on WHO's derived weights because we applied a comprehensive enhanced methodology, used actual variation in goal attainment, elicited marginal weights from behind a veil of ignorance, and created a setting to help respondents understand the concepts and their task. And because new weights could affect countries' rankings and comparisons over time, we advocate that appropriate weights be applied in future (international) comparisons of health system performance. Our study provides a promising and challenging basis on which to improve the methodology of global health system performance measurement. Advancing performance measurement is essential to cross-country comparison, which, in turn enhances global evidence-based health policy information.

References

[1] World Health Organization. The World Health Report 2000 – Health Systems: Improving Performance. Geneva: World Health Organization; 2000.

[2] Ratcliffe J, Buxton M. Patients' preferences regarding the process and outcomes of life-saving technology. International Journal of Technology Assessment in Health Care 1999;15:340–51.

[3] Tyler TR. Process utility and help seeking: what do people want from experts? Journal of economic psychology 2006;27:360–76.

[4] van Empel IWH, Dancet EAF, Koolman XHE, Nelen WLDM, Stolk EA, Sermeus W, et al. Physicians underestimate the importance of patient-centredness to patients: a discrete choice experiment in fertility care. Human Reproduction 2011;26:584–93.

[5] Parsons T. The structure of social action. New York: The Free Press (1937); 1968.

[6] Deci EL, Ryan RM. Intrinsic motivation and self-determination in human behavior. New York: Plenum Publishing Co; 1985.

[7] Deci EL, Ryan RM. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry 2000;11:227–68.

[8] Smith PC, Mossialos E, Papanicolas I, Leatherman S. Performance measurement for health system improvement: experiences, challenges and prospects. Cambridge: Cambridge University Press; 2009.

[9] Shibuya K, Boerma J. Measuring progress towards reducing health inequalities. Bulletin of the World Health Organization 2005;83:162.

[10] McKie J, Richardson J. Neglected equity issues in cost effectiveness analysis. West Heidelberg, Victoria: Centre for Health Economics, Monash University; 2005.

[11] Norman R, Hall J, Street D, Viney R. Efficiency and equity: a stated preference approach. Health Economics, First published online 23 April 2012. http://dx.doi.org/10.1002/hec.2827, in press.

[12] Frey BS, Benz MA, Stutzer. Introducing procedural utility: not only what, but also how matters. Journal of Institutional and Theoretical Economics 2004;160:377–401.

[13] Murray CJL, Knaul F, Musgrove P, Xu K, Kawabata K. Defining and measuring fairness in financial contribution to the health system. Geneva: World Health Organization; 2000.

[14] Xu K, Klavus J, Evans DB, Hanvoravongchai P, Zeramdini R, Murray CJL. Understanding household catastrophic health expenditures: a multi-country analysis. In: Murray CJL, Evans DB, editors. Health system performance assessment: debates, methods and empiricism. Geneva: World Health Organization; 2003. p. 566–73.

[15] Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD health care quality indicators project. International Journal for Quality in Health Care 2006;18:5.

[16] OECD. The OECD health project: Towards high-performing health systems. Paris. Organisation for Economic Co-operation and Development; 2004.

[17] Murray CJL, Lauer J, Tandon A, Frenk J. Overall health system achievement for 191 countries. Geneva: World Health Organization; 2000.

[18] Gakidou EE, Murray CJL, Evans DB. Quality and equity: preferences for health system outcomes. In: Murray CJL, Evans DB, editors. Health system performance assessment: debates, methods and empiricism. Geneva: World Health Organization; 2003. p. 667–765.

[19] Almeida C, Braveman P, Gold MR, Szwarcwald CL, Ribeiro JM, Miglionico A, et al. Methodological concerns and recommendations on policy consequences of the World Health Report 2000. The Lancet 2001;357:1692–7.

[20] Williams A. Science or marketing at WHO? A commentary on 'World Health 2000'. Health Economics 2001;10:93–100.

[21] Smith P. Developing composite indicators for assessing overall system efficiency. In: Smith P, editor. Measuring up: improving health system performance in OECD countries. Paris: Organisation for Economic Cooperation and Development; 2002. p. 275–95.

[22] Gakidou E, Murray CJL, Frenk J. Measuring preferences on health system performance assessment. GPE Discussion Paper No. 20. World Health Organization. Geneva; 2000.

[23] Richardson J, Wildman J, Robertson IK. A critique of the World Health Organisation's evaluation of health system performance. Health Economics 2003;12:355–66.

[24] Oswaldo Cruz Foundation. Report of the workshop "Health Systems Performance – the World Health Report 2000", http://www.who.int/health-systems-performance/docs/articles/oswaldocruz_report.pdf; cited November 2011.

[25] Lauer JA, Evans DB, Murray CJL. Measuring health system attainment: the impact of variability in the importance of social goals. In: Murray CJL, Evans DB, editors. Health systems performance assessment: debates, methods and empiricism. Geneva: World Health Organization; 2003. p. 677–81.

[26] Lauer J, Lovell CAK, Murray C, Evans D. World health system performance revisited: the impact of varying the relative importance of health system goals. BMC Health Services Research 2004;4:19.

[27] Lancaster KJ. A new approach to consumer theory. The Journal of Political Economy 1966;74:132–57.

[28] Louviere JJ, Hensher DA, Swait JD. Stated choice methods: analysis and applications. Cambridge: Cambridge University Press; 2000.

[29] Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. Health Economics 2003;12:545–51.

[30] Rawls J. A theory of justice. Cambridge: Belknap Press of Harvard University Press; 1971.

[31] Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs, 2nd ed. Cambridge: Cambridge University Press; 1993.

[32] Kjær T. A review of the discrete choice experiment-with emphasis on its application in health care. University of Denmark. 2005. Health Economic Papers 2005:1.

[33] Ortiz JP, Valentine NB, Gakidou EE, Tandon A, Kawabata K, Evans DB, et al. In: Murray CJL, Evans DB, editors. Inequality in responsiveness: population surveys from 16 OECD countries. Geneva: World Health Organization; 2003. p. 667–75.

[34] Otten F, Bos W, Vrooman C, Hoff S. Poverty notice 2006 [in Dutch: Armoedebericht 2006]. Den Haag: Centraal Bureau voor de Statistiek Sociaal en Cultureel Planbureau; 2006.

[35] Street DJ, Burgess L, Louviere JJ. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing 2005;22:459–70.

[36] Williams A, Cookson R. Equity in health. In: Culyer AJ, Newhouse JP, editors. North Holland handbook of health economics. Amsterdam: Elsevier; 2000. p. 1863–909.

[37] Salkeld G, Ryan M, Short L. The veil of experience: do consumers prefer what they know best? Health Economics 2000;9:267–70.

[38] Ryan M, Amaya-Amaya M. Threats to and hopes for estimating benefits. Health Economics 2005;14:609–19.