# Title: The biasing effect of order in multiple time trade-off judgments.

Running head: The biasing effect of order in multiple TTO judgments.

Authors: Michaël Schwarzinger, MD, MPH, Lucile Foulley, MPH, Thea Herz, PhD, Isabelle Durand-Zaleski, MD, PhD.

Departments and institutions involved in the work: Institut National de la Santé et de la Recherche Médicale, INSERM unité 444 (MS), Faculté de Médecine Saint-Antoine, Paris; service de pharmacologie clinique, (MS, LF) Hôpital Henri Mondor, AP-HP, Créteil; Laboratoire de Psychologie Sociale (TH), Université René Descartes, Paris; Service de Santé Publique (IDZ), Hôpital Henri Mondor, AP-HP, Créteil, France.

Part of the results was presented at the 8th biennial meeting of the European Society for Medical Decision Making, Taormina, Sicily, Italy, June 2-5, 2002.

Address correspondence and reprint requests to Dr. Schwarzinger: Inserm 444, Faculté de médecine Saint-Antoine, 27 rue Chaligny, 75571 Paris Cedex 12, France; e-mail: schwarzi@u444.jussieu.fr

# Abstract

**Background**: Most valuation experiments include only a few health states. We tested an order effect in multiple Time Trade-off (TTO) elicitations, i.e. the utility given to a particular health state would be different when the previous health state elicited was a mild health state as compared to a bad health state. **Methods**: A convenience sample of 137 nurses and student nurses was randomly assigned to the elicitation of: (1) either a mild or a severe disease stage, and (2) "moderate persistent asthma". The initial mild and severe disease stage had either similarities (physical limitations only) or dissimilarities (physical and mental limitations) with subsequent "moderate persistent asthma". **Results**: Median utility of "moderate persistent asthma" in respondents having elicited a mild disease stage first was significantly higher than in those having elicited a severe disease stage first (0.88 vs. 0.75, $p<0.02$). This order effect did not change whether the first disease stage shared similarities or not with "moderate persistent asthma", and it was mainly related to an assimilation effect between the first and subsequent disease stages in those having elicited a mild disease stage first (39% vs. 21%, $p<0.02$). **Conclusion**: TTO utility assessment using multiple ratings may be subject to substantial bias.

**Keywords**: order effect; assimilation effect; contrast effect; social judgment; time trade-off; QALY; DALY; cost-effectiveness analysis.

# Introduction

Quality-Adjusted Life Years (QALYs) take into account both the number of years lived in a particular condition and the relative severity of a year lived in that condition. QALYs have been recommended to measure effectiveness of health interventions in cost-effectiveness analysis [1], and the number of published cost-effectiveness analyses relying on QALYs increased exponentially with time [2]. However there is no gold standard valuation method to elicit the relative severity of year lived in a particular condition. Various valuation methods have been developed following rational theories of riskless or risky choice (time trade-off (TTO) and standard gamble (SG), respectively) [3]. In both valuation methods, respondents should assume that they are in a given health state during their lifetime, and they should trade-off life expectancy or survival probability in full health (respectively) until they are indifferent between the two options. The repetition of trade-off exercises with different health states allows to reveal the individual preferences for health states by the computed utilities that range from 0 (dead) to 1 (full health).

A basic assumption of rational theories of choice is the principle of invariance. It states that the relation of preference between health states should not depend on the description of the options (i.e., framing invariance) nor on the method of elicitation (i.e., procedure invariance). Without stability across equivalent descriptions and equivalent elicitation procedures, respondents' preferences cannot be represented as maximizing some subjective measure. Several experiments have addressed this issue in health economics. Within a given valuation method, framing invariance was not supported by different types of health state description [4-7], nor when the reference anchor varied from full health to death [4,7-9]. Procedure invariance was not supported by different types of survey (i.e. auto-questionnaire, interview with props [10], or computer-assisted survey [11]), the search procedure for the indifference trade-off number (i.e., when the usual “ping-pong” search procedure that narrows the trade-off number from both extremes to the indifference point was challenged by a top-down titration that decreases the trade-off number from full health to death [12]), nor when the usual TTO method that compares a period of ill-health with a shorter period in a higher quality-of-life was challenged by an unconventional TTO method that compares a period of ill-health with a longer period in a lower quality-of-life [13].

Although framing and procedure effects show that preferences for health states are constructed, not merely revealed, in the process of valuation, they are inherent in the valuation method. The psychological mechanisms underlying the all valuation process, i.e. from the health state to a utility elicited within a valuation method, received little attention. Two common findings indicate however that psychological mechanisms should play a crucial role in the valuation process. First, the personal knowledge of the condition impacts strongly utilities, i.e. valuation surveys showed consistently that the same disease stage received a decreasing utility from patients to the general public and healthcare professionals (see Dolan [14]). Second, disease labeling may decrease significantly the utility of an unlabeled health state as shown by the labeling of mental conditions [15], or conditions of bad prognosis [6,16] in the general population.

Assuming that human judgment is comparative in nature, previous findings may be explained by the selective knowledge respondents consult to make their comparative evaluation. In the first case, the intimate knowledge of the condition may be used as a standard that varies sharply from patients to healthcare professionals, i.e. patients may actually cope with their condition whether healthcare professionals may recall predominantly bad clinical experiences. When valuing the proposed disease stage, the former may assimilate a rather bad disease stage to their actual well-perceived condition, whether the latter may assimilate a rather good disease stage to previous bad clinical experiences. In the second case, disease labeling may trigger off social fears and sad personal knowledge of the condition to which respondents may assimilate the proposed disease stage. When valuing unlabeled health state, as described by a generic health state profile [17], pieces of information on conditions are less accessible, and the respondents are less likely to use them in the valuation process.

Psychological mechanisms at stake in comparative evaluation were recently integrated in the selective knowledge accessibility model [18]. Three major stages are involved in processes of comparative evaluation: standard selection, target-standard comparison, and valuation. Once a standard has been selected for comparison, respondents further have to determine on which particular features of the standard and the target the comparison should be based on. In an initial step, respondents engage in an overall assessment of the target and the standard, in which they

briefly consider a small number of particularly salient features to determine whether both are generally similar or dissimilar. If this assessment indicates that the target is generally similar to the standard, respondents will engage in a process of similarity testing by selectively searching for evidence indicating that the target's standing on the judgmental dimension is indeed similar to that of the standard, and will then move its valuation closer to the standard (assimilation effect). The reverse occurs if overall assessment indicates that the target is dissimilar to the standard (contrast effect).

Most valuation surveys conducted in health economics included a few health states related to a single[19] or several conditions [20]. The aim of this study is to evidence that psychological mechanisms other than framing and procedure effects interact in the valuation of a disease stage. We hypothesized that the utility stated for a given disease stage will be influenced by the disease stage evaluated previously, i.e. the initial elicitation for a disease stage will provide the by default accessible knowledge to be used in subsequent target evaluation. We tested our hypothesis in two-by-two factorial design. A convenience sample of 137 nurses and student nurses was randomly assigned to the elicitation of: (1) either a mild or a severe disease stage, and then (2) a disease stage of moderate severity, i.e. "moderate persistent asthma". The initial mild and severe disease stage had either similarities (physical limitations only) or dissimilarities (physical and mental limitations) with "moderate persistent asthma" depending on the group (nurses or student nurses). Accordingly, tests for an order effect in multiple TTO valuations of disease stages were performed overall, and depending on similarities or dissimilarities between the first disease stage and "moderate persistent asthma". Finally, the results on order effect were discussed in light of the selective accessibility model and debriefing records.

# Material and Methods

## Study design

Respondents were randomly assigned to the valuation of an initial disease stage of low or high severity, and then valued the same disease stage of moderate severity, i.e. "moderate persistent asthma". In 73 nurses, the initial disease stage of low severity ("chronic low back pain") had

similarities with “moderate persistent asthma”, whether the initial disease stage of high severity had salient dissimilarities on the mental axe (“moderate permanent impairments after stroke”). In 90 student nurses, the initial disease stage of high severity (“quadriplegia”) had similarities with “moderate persistent asthma”, whether the initial disease stage of low severity had salient dissimilarities on the mental axe (“mild depression”). In the follow-up, two-third of respondents were proposed closed-ended questions on their knowledge of each condition either from self, relatives, or medical departments, and one-third of respondents were proposed an open-ended question on their thought processes at stake when eliciting a disease stage. Finally, the experiment was reproduced in one-third of student nurses with disease stages related to the same condition (“mild, moderate, severe persistent asthma”). The order of presentation of disease stages to each respondent was the same in the first and follow-up experiments.

## Population

We selected purposely a homogenous group to limit confounding factors, i.e. knowledge of conditions. We recruited a convenience sample of nurses (n=73) graduating for sister (Ecole Montsouris, Paris), and student nurses (n=90) during their summer training (Hôpital Henri Mondor, Créteil).

## Disease stages' descriptions

Disease stage’s descriptions were taken from the European Disability Weights study [21,22]. In this study, external health care professionals and public health experts participated in both the subdivision of prevalent diseases into homogenous disease stages with respect to functional status, treatment and prognosis, and the elaboration of a brief clinical description for each disease stage. All disease stages were described on a separate sheet with the name of the disease at the top, the position of the disease stage among other stages of the condition, a brief clinical description and a health state profile, i.e. EQ-5D extended with a cognitive dimension [23-25].

We selected five disease stages based on both their Visual Analogue Scale mean (SD) score of relative severity (from 0=worst imaginable health state to 100=best imaginable health state) in the European Disability Weights study and their physical and mental limitations. The initial disease stages of low severity were either “chronic low back pain” (68 (18)) or “mild depression”

(65 (18)). The initial disease stages of high severity were either “moderate permanent impairments after stroke” (32 (16)) or “paraplegia” (29 (19)). The subsequent disease stage of moderate severity was “moderate persistent asthma” (54 (19)).

## Valuation method

Face-to-face interviews were conducted by two interviewers (LF, MS) following a standardized elicitation protocol assisted by a laptop. Utilities were elicited by the time trade-off technique with the following features: a fixed 40 years life-expectancy; a fixed gain formulation (e.g. “do you prefer to live 30 years in full health over living 40 years in the described health state?”); and a “ping-pong” elicitation procedure starting by the extreme anchors (i.e., full health and then death). The Excel® program: (1) allocated randomly respondents to the order of presentation of disease stages (initial disease stage of low or high severity); (2) helped the respondent in the “ping-pong” elicitation procedure by means of a chart (one horizontal bar of 40 years life expectancy fixed in the specified disease stage, and another parallel varying life expectancy in full health according to respondent’s answers); (3) indicated when respondent provided an inconsistent answer in the “ping-pong” elicitation procedure, e.g. respondent provided a trade-off number outside the interval defined by previous answers, to allow respondent to reconsider her choice.

## Statistical analysis

Disease stage’s utility was computed on a scale from zero (death) to unity (full health) by dividing by 40 the trade-off number of years for which respondent was indifferent. Normality of utilities was rejected at the 0.001 level for all disease stages (Kolmogorov-Smirnov test). We used non-parametric Wilcoxon two-sample tests to compare utilities. Binary data were compared by a chi-square test or Fisher's exact test, and normal data were compared by the Student t-test. Finally, explanatory factors for an assimilation effect as strictly measured by an equal utility elicited for the first disease stage and the subsequent “moderate persistent asthma”, were tested in logistic regression. Significance was attributed at the 10% level, and data were analyzed with SAS 8.2 (SAS Institute, Cary NC).

# Results

## Manipulation checks

Seventy-three nurses and ninety student nurses were approached to participate in the elicitation experiment. After presentation of the study purpose, the first disease stage, and the TTO method, 26 (16%) respondents refused to go through the experiment. The main reason provided was the lack of time and/or incomprehension of the elicitation task. Acceptance rate among nurses depended on the first disease stage presented (low severity: 91% vs. high severity: 66%, $p<0.10$). For the 50 nurses and 87 student nurses who went through the experiment, mean age (23 (SD 4) and 39 (SD 7), respectively) and gender (80% and 80% females, respectively) did not differ significantly according to the first disease stage presented. Overall, median utilities decreased significantly from disease stages of low severity ("chronic low back pain": 0.9, and "mild depression": 0.88) to disease stages of high severity ("moderate permanent impairments after stroke": 0.5, and "paraplegia": 0.38) (**Table 1**).

## Order effect with different conditions

Overall, the median utility of the subsequent disease stage ("moderate persistent asthma") was significantly higher in respondents having first elicited a disease stage of low severity than in those having first elicited a disease stage of high severity (0.88 vs. 0.75, respectively, $p<0.02$) (**Table 1**). In the order of presentation from low to moderate severity, median utilities of the first disease stage and "moderate persistent asthma" were very similar and showed a strong assimilation effect, even if the first disease stage, i.e. "mild depression" had salient mental limitations. In the order of presentation from high to moderate severity, median utilities of the first disease stage and "moderate persistent asthma" were very dissimilar, and this contrast effect was even more pronounced when the first disease stage, i.e. "stroke" had salient mental limitations. As a result, median utility for "moderate persistent asthma" was significantly higher in the low to moderate severity order than in the high to moderate severity order if the severe disease stage, i.e. "quadriplegia", shared disease's similarities with "moderate persistent asthma" ($p<0.06$), whether it was not anymore if the severe disease stage, i.e. "stroke", had dissimilarities with "moderate persistent asthma" ($p=0.14$).

As shown in **Table 2**, an assimilation effect strictly measured by an equal utility elicited for the first disease stage and the subsequent "moderate persistent asthma", was present in 26 (39%) respondents in the order of presentation from low to moderate severity vs. 14 (21%) in the order of presentation from high to moderate severity ($p<0.02$). Of these respondents, 16 (62%) and 5 (36%) gave a utility equal to unity for both disease stages, respectively. On the other hand, 74% of respondents contrasted "moderate persistent asthma" from a disease stage of high severity.

We performed a multivariate logistic regression to look for possible explanatory factors of an assimilation effect as it may provide the best explanation for an order effect. As shown in **Table 3**, a higher utility elicited for the first disease stage was the only significant explanatory factor for an assimilation effect ($p<0.001$). Respondents who gave a utility above 0.85 to the first disease stage were 21 times more likely to assimilate "moderate persistent asthma" to the first disease stage than those who gave a utility below 0.44. Similarities between the first disease stage and "moderate persistent asthma", the lack of knowledge of the first disease stage, and knowledge of "asthma" were not significantly associated to an assimilation effect. When model 1 was performed according to order of disease stages' presentation (low to moderate severity or high to moderate severity), a higher utility elicited for the first disease stage was still significantly associated with an assimilation effect ($p<0.01$ in each model).

Debriefing questions in 44 respondents on the thought processes at stake indicated that respondents showing an assimilation effect were more unwilling to trade-off any year of life expectancy or below a threshold (13 (72%) vs. 15 (58%), $p=0.32$). The main reasons (non exclusive) included the willingness to enjoy life anyway in younger respondents, the need to take care of children in older respondents, and curative hopes from medical research in the next 40 years. On the other hand, respondents that did not show an assimilation effect insisted more on the relative severity and the level of social support of disease stages (17 (65%) vs. 7(39%), $p<0.08$).

## Order effect related to the same condition

When disease stages belonged to the same condition, there was no significant order effect anymore in 33 student nurses having participated in this follow-up experiment (**Table 1**). An

assimilation effect was still present in 27% of respondents without significant difference according to order of disease stages' presentation (p=0.70). Two-third of respondents showing an assimilation effect when eliciting disease stages related to the same condition had previously shown an assimilation effect when eliciting disease stages from different conditions.

# Discussion

QALYs are recommended for cost-effectiveness analysis since 1996[1], and they are indeed increasingly used in cost-effectiveness analysis [19]. However, the question remains whether the 'Q' is simply revealed by elicitation exercises or actually constructed during the process of elicitation. The former approach relies on strong assumptions, i.e. individual preferences are well formed, complete, and stable [26,27]. As shown by this study, TTO utility assessments using multiple ratings may be subject to substantial bias, i.e. the utility of a subsequent disease stage elicited after a disease stage of low severity was significantly higher than after a disease stage of high severity.

The order effect evidenced in this study was mainly related to an assimilation effect between the first and the subsequent disease stages' elicitations. Respondents exposed to an initial disease stage of low severity were prone to take into consideration life expectancy *per se* and their life expectations. About 40% of those respondents gave the same utility to the subsequent disease stage of an expected higher severity ("moderate persistent asthma"), revealing a low or null willingness to trade-off years of life expectancy for both disease stages. On the other hand, respondents exposed to an initial disease stage of high severity grasped on the relative severity and the level of social support of the condition. Three-quarters of respondents contrasted the utility of the subsequent disease stage of an expected lower severity ("moderate persistent asthma"), whether one-quarter showed an assimilation effect. Contrary to our expectations, similarities between the first and subsequent conditions (physical limitations only in "chronic low back pain" or "quadriplegia") did not reduce the order effect. The main reason for an order effect was the proportion of respondents showing an assimilation effect between the first and subsequent disease stages' utilities, as predicted by the utility given to the first disease stage.

To the best of our knowledge, this study is the first to show an order effect in multiple utility assessments of health states. The only study we are aware off dealing with order effects in health state valuation actually controlled for order effects rather than studied them [29]. Fifty-five patients with early-stage breast cancer had to value their actual health state and a radiotherapy scenario at different times of their disease (before, during and after radiotherapy). There was no significant difference in utilities of the two health states depending on which one was elicited first. In light of this study, it may be explained by the clear overlap of the two health state utilities (e.g. 0.94 and 0.89 before radiotherapy, respectively). Order effects were shown however with other valuation methods than TTO or SG [28,30]. In all, order effects reflect the informational underpinnings of comparison in any subjective evaluation. Evaluation is relative in that it refers to a comparison of the evaluated target to a given context, norm or standard [18]. The elicitation of a first disease stage of low or high severity calls for different psychological mechanisms and provides a standard for a subsequent disease stage elicitation. The persistence of an order effect when the first disease stage had either similarities (physical limitations only) or dissimilarities (both physical and mental limitations) with the subsequent "moderate persistent asthma" is in favor of different psychological mechanisms mainly driven by the interaction of the relative severity of disease stage and the time trade-off task itself.

This study has some limitations. The refusal rate was about 16% which compares to usual refusal rates in valuation experiments conducted in the general population. However, refusal rate was significantly higher in nurses (37%) than student nurses (3%) ($p<0.0001$). This discrepancy may be truly related to the available time to participate in this 20 minutes' experiment (i.e., nurses were mostly interviewed in their medical department during the busy summer period as compared to nurse students interviewed during lunch breaks during their summer courses). However, many refusing respondents said they did not understand the valuation task and the refusal rate depended surprisingly on the first disease stage presented (low severity: 9% vs. high severity: 34%). Since time pressure impacts on comparison processes and consequences [31], it may be argued that those respondents did not find a standard for disease stage elicitation within their own time constraint, and that standard selection is more time-consuming in a TTO for a severe disease stage to the extent that a real trade-off is even more expected.

The generalisability of the results depends on two factors: 1) the sample selection; 2) the choice of stimuli. First, we selected a convenience sample of French nurses and student nurses. The between-subject experimental study design allows some generalization on the order effect in multiple TTO assessments for disease stages. Moreover, the knowledge of the conditions was not associated to the assimilation effect supposedly responsible for the order effect. However, it will be worth replicating the experiment in other countries to look for differential comparison processes and consequences, even if a reasonably high level of agreement was recently supported for TTO utilities of disease stages in Western European countries [22].

Second, the choice of disease stages could affect the results. The order effect was significant with disease stages related to different conditions. However, the follow-up experiment in student nurses did not find anymore a significant order effect when disease stages were related to the same condition (i.e. asthma). The small sample size (n=33) and the rather high utility of the most severe disease stage (median 0.75) could explain it. It might be worth replicating the experiment with disease stages related to the same condition and covering a broader range of utilities.

What are the implications of this study for multiple TTO assessments? Firstly, the study brings into question the assumptions that are typically made in health state valuation. The order effect could be added to procedure effects [10-13], framing effects [4-8], logical inconsistencies [32,33], and internal inconsistency [34] found in health state valuation methods that show that health states utilities are poorly constructed values. Secondly, an order effect in multiple TTO assessments is the likely result of the interaction between the relative severity of the first health state presented for valuation and the valuation method itself. Specific debiasing strategies should be looked for. Previous experiments showed the robustness of anchoring effects in other economic areas when respondents were paid or warnings were provided [35,36]. On the other hand, the initial valuation of the disease stages of the lowest and highest severity could reduce the order effect when eliciting an in-between disease stage [37,38]. Finally, the recognition of these failures of valuation method should contribute to a new conception of judgment and choice in which preferences are often constructed, not merely revealed, in the process of elicitation. One urgent task is to define a gold standard valuation method including the number of health states to be valued and in which order.

Otherwise, the comparison and use in league tables of (constructed) utilities from various valuation studies may be meaningless [39,40].

# Acknowledgments

We are grateful to the student nurses and nurses that accepted to participate in this valuation experiment. We are also indebted to the EDW project for providing the background valuation material of this study. We would like to thank Pr Arthur Elstein for very useful suggestions.

# References


1. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
2. Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med* 2000;**132**(12)**:**964-72.
3. Nord E. Methods for quality adjustment of life years. *Soc Sci Med* 1992;**34**(5)**:**559-69.
4. McNeil BJ, Pauker SG, Sox HC, Jr., Tversky A. On the elicitation of preferences for alternative therapies. *N Engl J Med* 1982;**306**(21)**:**1259-62.
5. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. Describing health states. Methodologic issues in obtaining values for health states. *Med Care* 1984;**22**(6)**:**543-52.
6. Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. *J Clin Epidemiol* 1993;**46**(1)**:**77-84.
7. Bernstein LM, Chapman GB, Elstein AS. Framing effects in choices between multioutcome life-expectancy lotteries. *Med Decis Making* 1999;**19**(3)**:**324-38.
8. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;**211**(4481)**:**453-8.
9. Blumenschein K, Johannesson M. An experimental test of question framing in health state utility assessment. *Health Policy* 1998;**45**(3)**:**187-93.
10. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;**15**(2)**:**209-31.
11. Lenert LA, Sherbourne CD, Reyna V. Utility elicitation using single-item questions compared with a computerized interview. *Med Decis Making* 2001;**21**(2)**:**97-104.
12. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitations. *Med Decis Making* 1998;**18**(1)**:**76-83.
13. Spencer A. The TTO method and procedural invariance. *Health Econ* 2003;**12**(8)**:**655-68.
14. Dolan P. Whose preferences count? *Med Decis Making* 1999;**19**(4)**:**482-6.
15. Rabin R, Rosser RM, Butler C. Impact of diagnosis on utilities assigned to states of illness. *J R Soc Med* 1993;**86**(8)**:**444-8.

16. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *J Chronic Dis* 1978;**31**(11)**:**697-704.
17. Hawthorne G, Richardson J. Measuring the value of program outcomes: a review of multiattribute utility measures. *Expert Rev Pharmacoeconomics Outcomes Res* 2001;**1**(2)**:**215–28.
18. Mussweiler T. Comparison processes in social judgment: mechanisms and consequences. *Psychol Rev* 2003;**110**(3)**:**472-89.
19. Neumann PJ, Zinner DE, Wright JC. Are methods for estimating QALYs in cost-effectiveness analyses improving? *Med Decis Making* 1997;**17**(4)**:**402-8.
20. Murray CJ. Rethinking DALYs. In: Murray CJ, Lopez AD, eds. The Global Burden of Disease. Vol 1: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge: Harvard University Press, 1996: 1-99.
21. Essink-Bot ML, Pereira J, Packer C, Schwarzinger M, Burström K, and the European Disability Weights group. Cross-national comparability of burden of disease estimates: the European Disability Weights project. *Bull World Health Organ* 2002;**80:**644-52.
22. Schwarzinger M, Stouthard MEA, Burström K, Nord E, on behalf of the European Disability Weights Group. Cross-national agreement on disability weights: the European Disability Weights Project. *Popul Health Metr* 2003**:**in press.
23. EuroQol--A new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy* 1990;**16**(3)**:**199-208.
24. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**(1)**:**53-72.
25. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol* 1999;**52**(4)**:**293-301.
26. Shiell A, Hawe P, Seymour J. Values and preferences are not necessarily the same. *Health Econ* 1997;**6**(5)**:**515-8.
27. Treadwell JR, Lenert LA. Health values and prospect theory. *Med Decis Making* 1999;**19**(3)**:**344-52.
28. Ubel PA, Richardson J, Baron J. Exploring the role of order effects in person trade-off elicitations. *Health Policy* 2002;**61**(2)**:**189-99.

29. Jansen SJ, Stiggelbout AM, Wakker PP, Nooij MA, Noordijk EM, Kievit J. Unstable preferences: a shift in valuation or an effect of the elicitation procedure? *Med Decis Making* 2000;**20**(1)**:**62-71.
30. Moore DA. Order effects in preference judgments: evidence for context dependence in the generation of preferences. *Organ Behav Hum Decis Process* 1999;**78**(2)**:**146-165.
31. Mussweiler T, Strack F. Hypothesis-consistent testing and semantic priming in the anchoring paradigm: a selective accessibility model. *Journal of Experimental Social Psychology* 1999;**35:**136-64.
32. Souchek J, Stacks JR, Brody B, et al. A trial for comparing methods for eliciting treatment preferences from men with advanced prostate cancer: results from the initial visit. *Med Care* 2000;**38**(10)**:**1040-50.
33. Devlin NJ, Hanse P, Kind P, Williams A. Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Econ* 2002**:**in press.
34. Stalmeier PF. Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. *Med Decis Making* 2002;**22**(1)**:**53-64.
35. Wilson TD, Brekke N. Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychol Bull* 1994;**116**(1)**:**117-42.
36. Wilson TD, Houston CE, Etling KM, Brekke N. A new look at anchoring effects: basic anchoring and its antecedents. *J Exp Psychol Gen* 1996;**125**(4)**:**387-402.
37. Chapman GB, Johnson EJ. The limits of anchoring. *J Behav Dec Making* 1994;**7:**223-42.
38. Mussweiler T. The malleability of anchoring effects. *Exp Psychol* 2002;**49**(1)**:**67-72.
39. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000;**38**(6)**:**583-637.
40. Bell CM, Chapman RH, Stone PW, Sandberg EA, Neumann PJ. An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses. *Med Decis Making* 2001;**21**(4)**:**288-94.

Table 1. Summary statistics of utilities of the first disease stage and subsequent "moderate persistent asthma" according to order of disease stage's presentation and similarities between diseases stages.

| Disease stage | Order of disease stages' presentation: Low to moderate severity | | | | Order of disease stages' presentation: High to moderate severity | | | | Wilcoxon two-sample test | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | Median | Mean | SD | n | Median | Mean | SD | Overall | |
| Disease stage other than asthma | 66 | 0.88 | 0.76 | 0.30 | 66 | 0.48 | 0.43 | 0.29 | <0.0001 | |
| Moderate persistent asthma | 71 | 0.88 | 0.81 | 0.24 | 66 | 0.75 | 0.71 | 0.26 | <0.02 | |
| | | | | | | | | | P*P | P*(P&M) |
| Disease stage other than asthma with similarities (physical limitations) (P) | 27 | 0.90 | 0.80 | 0.27 | 44 | 0.38 | 0.41 | 0.31 | <0.0001 | <0.001 |
| Moderate persistent asthma (P) | 28 | 0.88 | 0.82 | 0.22 | 44 | 0.75 | 0.71 | 0.27 | <0.06 | 0.14 |
| | | | | | | | | | (P&M)*(P&M) | (P&M)*P |
| Disease stage other than asthma with dissimilarities (physical and mental limitations) (P&M) | 39 | 0.88 | 0.73 | 0.31 | 22 | 0.50 | 0.47 | 0.22 | <0.01 | <0.0001 |
| Moderate persistent asthma (P&M) | 43 | 0.88 | 0.80 | 0.25 | 22 | 0.81 | 0.72 | 0.24 | 0.13 | <0.06 |
| | | | | | | | | | Overall | |
| Disease stage related to asthma | 15 | 1 | 0.91 | 0.25 | 20 | 0.75 | 0.61 | 0.30 | <0.001 | |
| Moderate persistent asthma | 18 | 0.81 | 0.69 | 0.28 | 19 | 0.88 | 0.77 | 0.29 | 0.41 | |

P (P&M) related to utilities of disease stages when the first disease stage has similarities (dissimilarities) with "moderate persistent asthma".

P*(P&M): Wilcoxon two-sample tests to compare utilities of either the first disease stage or "moderate persistent asthma" elicited in the order low to moderate severity (left letter, here first disease stage has similarities) and high to moderate severity (right letter, here first disease stage has dissimilarities).

Table 2. Assimilation and contrast effects between utilities of the first disease stage and subsequent "moderate persistent asthma" according to the relative severity of the first disease stage and similarities between the first and subsequent disease stages.

| Relative severity of the first disease stage and similarities with "moderate asthma" | Utility of "moderate persistent asthma" compared to that of the first disease stage elicited | | |
|---|---|---|---|
| | = | lower | higher |
| Low severity and similarities (physical limitations only) (n=27) | 48% | 22% | 30% |
| Low severity and dissimilarities (physical and mental limitations) (n=39) | 33% | 31% | 36% |
| High severity and similarities (physical limitations only) (n=44) | 20% | 5% | 75% |
| High severity and dissimilarities (physical and mental limitations) (n=22) | 23% | 5% | 73% |

Chi-square test of the distribution of equal utilities according to order (Chi-2, 1 DF=5.17, $p<0.02$)

Chi-square test of the distribution of equal utilities according to order when first disease stage has physical limitations only (Chi-2, 1 DF=6.00, $p<0.02$)

Chi-square test of the distribution of equal utilities according to order when first disease stage has both physical and mental limitations only (Chi-2, 1 DF=0.76, ns)

Table 3. Logistic regression models to predict an assimilation effect.

| Explanatory factors | Assimilation effect=1 n (%) | Model 1 (n=131) OR (CI 95%) | p value | Assimilation effect=1 n (%) | Model 2 (n=96)* OR (CI 95%) | p value |
|---|---|---|---|---|---|---|
| Utility of the first disease | | | | | | |
| Upper third (0.85-1) | 26 (57) | 21.2 (5.2-86.5) | | 19 (59) | 27.5 (5.1-148.6) | |
| Middle third (0.45-0.84) | 11 (26) | 5.4 (1.3-2.3) | <0.0001 | 6 (21) | 2.4 (0.5-12.1) | <0.001 |
| Lower third (0-0.44) | 3 (7) | 1 | | 3 (8) | 1 | |
| First disease stage with physical limitations only | | | | | | |
| Yes | 22 (31) | 1.5 (0.6-3.5) | 0.38 | 13 (25) | 1.7 (0.4-7.8) | 0.47 |
| No | 18 (30) | 1 | | 15 (33) | 1 | |
| Age of respondent | | | | | | |
| Upper third (>32 y)) | 14 (31) | 2.0 (0.4-10.3) | | 3 (25) | 2.7 (0.5-15.8) | |
| Middle third (22-32 y) | 13 (33) | 1.1 (0.4-3.1) | 0.69 | 13 (33) | 1.1 (0.3-3.8) | 0.52 |
| Lower third (<22 y) | 12 (26) | 1 | | 12 (26) | 1 | |
| Sex of respondent | | | | | | |
| Male | 9 (35) | 1.15 (0.41-3.24) | 0.79 | 8 (42) | 1.47 (0.41-5.35) | 0.56 |
| Female | 31 (29) | 1 | | 20 (26) | 1 | |
| Personal knowledge of first disease stage* | | | | | | |
| No | -- | -- | | 20 (29) | 1.2 (0.3-5.0) | 0.78 |
| Yes | -- | -- | | 7 (25) | 1 | |
| Knowledge of first disease stage from medical departments* | | | | | | |
| No | -- | -- | | 9 (31) | 1.7 (0.4-8.4) | 0.49 |
| Yes | -- | -- | | 19 (28) | 1 | |
| Personal knowledge of asthma* | | | | | | |
| Yes | -- | -- | | 16 (32) | 1.5 (0.5-4.3) | 0.49 |
| No | -- | -- | | 12 (26) | 1 | |
| Knowledge of asthma from medical departments* | | | | | | |
| Yes | -- | -- | | 16 (28) | 1.8 (0.4-7.9) | 0.46 |
| No | -- | -- | | 11 (29) | 1 | |
| Interviewer* | | | | | | |
| LF | 28 (29) | 2.2 (0.4-11.6) | 0.34 | -- | -- | |
| MS | 12 (34) | 1 | | -- | -- | |

OR (CI 95%): Odds-ratio (Confidence interval 95%)

* Information on knowledge of disease stages was available in 96 respondents interviewed by LF (model 2)