

Using Surveys to Calculate Disability-Adjusted Life-Years

Wolfgang Wiedermann, Ph.D., and Ulrich Frick, Ph.D.

Wolfgang Wiedermann, Ph.D., is a junior researcher at the Carinthia University of Applied Sciences, Feldkirchen, Austria.

Ulrich Frick, Ph.D., is a professor of public health at the Carinthia University of Applied Sciences, Feldkirchen, Austria, and senior scientist at the Psychiatric University Hospital, University of Regensburg, Regensburg, Germany.

Mapping a certain disease into a system of disabling attributes allows researchers to compare diseases within a common framework. To quantify the total burden of morbidity (e.g., morbidity attributable to alcohol use), so-called disability weights (DWs) must be generated. General-population surveys can be used to derive DWs from health valuation tasks. This article describes the application of three psychometric methods (i.e., pairwise comparisons, ranking tasks, and visual analog scales) in general-population surveys and outlines their strengths and weaknesses. A recently proposed health valuation framework also is presented, which highlights the underlying cognitive processes from a social-judgment perspective and presents a structured data-collection procedure that seems promising in deriving DWs from general-population surveys.

To quantify the burden of a disease within a population, a health-gap measure is more useful than measures of health expectancy or quality-adjusted life-years (see Etches et al. 2006). Disability-adjusted life-years (DALYs), the most prominent of the health-gap measures, combine the burden attributable to early death and to morbidity into one single number. Alcohol affects a long list of diseases and disabilities in varying intensities, each of which can be described by a number of health-state attributes. Common measures of health outcomes include the EuroQol5D (EQ5D) (Brooks and EuroQol Group 1996), the Health Utilities Index III (HUI III) (Feeny et al. 2002), the Short-Form 36 Health Survey (SF36) (Ware and Sherbourne 1992), and the Classification and Measurement System of Functional Health (CLAMES) (McIntosh et al. 2007).

Mapping a certain disease into a system of disabling attributes (e.g., physical functioning, pain, memory and thinking, etc.) enables health researchers to compare qualitatively different diseases within a common framework. To quantify the total burden of alcohol-attributable morbidity, it is necessary to provide so-called disability weights (DWs) for each of these health states, which are bounded by the DWs of 0 (for complete health) and 1 (for death). It should be noted that health states are considered, rather than diseases with labels (and their psychological and/or medical implications), when DWs are determined.

How DWs can validly be measured, defined, or (more neutrally speaking) elicited is of equal importance for the results as the question, “Who is asked to provide the DWs?” Although elicitation methods will be discussed below, this article does not focus on the question of which sources (e.g., patients, clinical experts, etc.) should be consulted to quantify DWs. Rather, this article considers only general-population surveys (i.e., telephone, face to face, or mailed) as sources of information on the disabilities associated with different health states.

How Are DWs Elicited?

Three popular methods to construct DWs stem from econometric utility theory: standard gamble (SG), time tradeoff (TTO), and person tradeoff (PTO). They all share the central idea that a respondent's point of indifference, at which he or she cannot unequivocally decide on a certain judgmental task, enables researchers to measure utility differences via the traded “goods.” For example, in SG, respondents are given a choice between an outcome that is certain (i.e., remaining in ill health) and a

gamble with one better and one worse outcome (e.g., full health or death). Respondents are asked what probability of the better outcome would make them indifferent to remaining in the described state (ill health) for certain or choosing the risky option. Therefore, if they are indifferent to the ill-health state and gamble with a 0.8 probability of the better outcome (but 0.2 probability of the worse outcome), 0.8 represents the utility of the ill health.

In a TTO task, respondents are asked to consider the relative amounts of time (e.g., number of life-years) they would be willing to sacrifice to avoid a certain poorer health state (e.g., frequent headaches). Assuming a scenario of 10 years with frequent headaches, the respondent may be indifferent to this state and a shorter lifetime of 7 years, resulting in an estimated utility for the frequent-headaches health state of 0.7 (7 years divided by 10 years).

A typical PTO elicitation asks respondents to choose between two equally expensive health care treatment programs that improve quality of life or save lives for two groups of patients. The decisionmaker must choose to fund one of the two mutually exclusive programs, one of which has a fixed number of patients. Respondents are asked how many patients would need to be treated to make them indifferent to the two programs. For example, program A might extend the life of 100 healthy individuals for 1 year, whereas program B might cure 100 individuals of a chronic health condition.

All three methods are time consuming, require highly motivated respondents, and are hardly feasible without a trained interviewer or computer program. Whereas TTO has been used in face-to-face interviews in the general population quite often (e.g., Badia et al. 2001; Chevalier

and de Pouvourville 2011; Dolan 1997; Greiner et al. 2005; Jelsma et al. 2003; Jo et al. 2008; Lamers et al. 2006; Lee et al. 2009; Shaw et al. 2010; Tsuchiya et al. 2002; Witttrup-Jensen et al. 2009; Zarate et al. 2008), in mail surveys only two studies (Burström et al. 2006; Lundberg et al. 1999) used TTO to quantify respondents' own health states. SG and PTO have rarely been used for eliciting health-state preferences in mailed surveys (i.e., they are usually used in face-to-face or phone interviews), and they have only been used among former patients (i.e., not the general public) (Hammerschmidt et al. 2004).

Readers are referred to Rehm and Frick (2010) for an overview on the methodological problems associated with econometric elicitation methods in this context. Recently, Wittenberg and Prosser (2011) described two additional sources of bias or mistaken responses in preference measurement in surveys: ordering errors (i.e., illogical responses, which violate a naturally given order, whereas inconsistent responses contradict each other within a person), and objections/invariance (i.e., respondents may refuse to participate because of an unwillingness to trade time [in the TTO task] or risk [in the SG task]). Furthermore, the meaning of SG results has been criticized as rather a measurement of risk attitude than a representation of subjective utility (Lenert and Kaplan 2000). TTO results as a metric for utility have been shown to vary with respondents' age, education, and current health state (Ayalon and King-Kallimanis 2010; Meropol et al. 2008; Stiggelbout et al. 1996; Voogt et al. 2005). Feasibility of PTO frequently is hampered because people tend to refuse such tasks because of their desire to avoid prejudice and discrimination (Damschroder et al. 2005).

As alternatives to the methods described above, psychometric theory provides paired comparisons, ranking tasks, and visual analogue scales as tools to elicit health-state preferences. These tools are discussed below.

Paired Comparisons

In the context of health-state valuation, a paired comparison (PC) task simply means that respondents must choose which of two given states is more disabling, worse, or dominant in some way. Because measuring via PC seems quite simple and feasible (because it is only necessary to present all health states in a consistent descriptive system), it has been applied in various surveys in the general population (Bijlenga et al. 2009; Kind 1982, 2005; Prieto and Alonso 2000; Ratcliffe et al. 2009; Stolk et al. 2010). For a recent application of PCs among an expert panel see Rehm and Frick (2013). Deriving DWs from the resulting pattern of dominance relations, by contrast, constitutes a complex statistical task for which solutions have been formulated from the theory of Thurstone scaling (Thurstone 1927), conditional logistic regression (Hosmer and Lemeshow 2000), and loglinear modeling (Critchlow and Fligner 1991).

Methodological challenges associated with PC stem from logically inconsistent judgments (e.g., $A > B$ and $B > C$, but $C > A$) and from rapidly increasing burden of task when comparing larger numbers of health states (i.e., combinatorial explosion). Intransitive judgments (e.g., in comparing 10, 7, and 5, 10 is preferred to 7 and 7 is preferred to 5, but 5 is preferred to 10) may originate in unintended framing effects as well as in imperfect judgment (von Winterfeldt and Edwards 1986). Recently published experimental stud-

ies favor the position that excluding inconsistent ratings cannot improve the description of true preferences and therefore might to some degree be an inevitable consequence of the decisionmaking process itself (Linares 2009). To keep the number of judgments at manageable dimensions, several studies have used incomplete factorial designs (Bijlenga et al. 2009; Prieto and Alonso 2000; Ratcliffe et al. 2009).

Asking subjects to rank order several health states, but statistically analyzing rankings as PCs, was used as an alternative in several studies (Krabbe 2008; Ip et al. 2004). Rankings can be transformed into a series of PCs (Francis et al. 2002), which at first glance avoids inconsistent judgments.

Ranking Tasks

Health-state rankings (i.e., putting several health states into an ordinal sequence of disability), which also provide comparative information, require less cognitive effort for survey respondents. Furthermore, simultaneous comparisons of multiple health states might be less sensitive to biases (e.g., those provoked by arbitrarily labeled endpoints of rating scales) (Maydeu-Olivares and Böckenholt 2008). Although ranking exercises had been included in numerous valuation studies as an external comparison measure for TTO and SG, researchers had not used the resulting ordinal data (McCabe et al. 2006) for construction of DWs before the seminal article by Salomon (2003). Cardinal utilities derived from health-state rankings displayed high agreement to utilities from TTO or SG methods (Craig et al. 2009*a, b*; Kind 2005) and were more stable in a cross-cultural comparison than weights derived from SG (Ferreira et al. 2011).

From a more theoretical viewpoint, articles by Flynn and colleagues (2010) and Flynn (2010) have raised serious statistical concerns about the use of ranks as a substitution for econometric valuation tasks. Their critique focuses on modeling assumptions and thus seems beyond the scope of this article. Nevertheless, their argument suggests that it can be important to restrict the number of alternatives to be ranked and to pay special attention to how a respondent generates rankings. In addition, Lenert and colleagues (1998) have demonstrated that reported utilities are heavily influenced by the search process used to form a certain judgment. This matches the notion that preferences often are constructed (instead of merely obtained) in the elicitation process (Slovic 1995). Ranking tasks within self-administered questionnaires might be hampered by limited control of the mechanism respondents use to generate the rank order. This introduces at least two issues: First, it remains unclear which reference attributes the respondent uses to generate the rank order, which constricts intersubjective comparability and provokes primacy biases (i.e., the tendency to give more attention to items listed first) (Bowling 2005). Second, from a more technical perspective, statistical ranking models (such as the rank-ordered logit model) assume that rankings were obtained using a particular psychological mechanism (Flynn 2010). For free rankings, however, it remains unclear which statistical model is most appropriate to describe the ranking mechanism. Furthermore, it cannot be ensured that respondents using a self-administered questionnaire judge along repeated best/worst choices, a “ping-pong” method that was shown to produce reliable data (Louviere et al. 2008).

Visual Analog Scale

To use a visual analog scale (VAS), respondents are asked to specify their level of agreement to a statement by indicating a position along a continuous line between two endpoints. Numerous studies have used VAS responses to derive health-state values in the general population (Björk and Norinder 1999; Cleemput 2010; Devlin et al. 2003; Dolan and Kind 1996; Essink-Bot et al. 1993; Greiner et al. 2003; Johnson and Pickard 2000; Johnson et al. 1998; Leidl and Reitmeir 2011). Krabbe and colleagues (2007) proposed a methodology based on differences in VAS values, where the ranks of pairwise VAS differences are used in a multidimensional scaling analysis to estimate cardinal health-state values. However, other researchers have questioned the validity of VAS data as cardinal values (Bleichrodt and Johannesson 1997; Devlin et al. 2004; van Osch and Stiggelbout 2005) for various reasons. First, VAS tasks in which the top and the bottom endpoints are precisely defined (e.g., death versus perfect health) allow direct comparison between individuals, whereas vague labels such as “worst imaginable” and “best imaginable” hamper an interindividual comparison (Torrance et al. 2001). Second, VAS responses might be affected by a so-called end-aversion bias, the phenomenon of respondents tending to be reluctant to mark positions near the endpoints of the scale (Bleichrodt and Johannesson 1997; Robinson et al. 2001; Torrance et al. 2001). Third, a VAS score for a certain health state may depend on other states presented at the same time (i.e., context bias) (Torrance et al. 2001). Fourth, the accuracy of VAS responses may be influenced by hand preferences and which hand was used (McKechnie and Brodie 2008).

Finally, the orientation of the VAS scale (vertical versus horizontal) itself might affect the shape of the resulting score distribution (e.g., Lundqvist et al. 2009). Taken together, VAS responses therefore should be interpreted on an ordinal scale level only.

Health Valuation: A Social Judgment Perspective

Stiggelbout and de Vogel-Voogt (2008) presented a four-step framework describing respondents' cognitive processes while valuing health states: perspective/perception of the stimulus, interpretation, judgment, and formation of a manifest response (see also Rehm and Frick 2010). For each step, several mechanisms have been identified, which may affect the final response.

1. Perspective/perception of the stimulus. In a meta-analysis, Dolders and colleagues (2006) reported no significant differences in preferences when patient surveys were compared with those of the general public, whereas a more recent and more extensive meta-analysis by Peeters and Stiggelbout (2010) suggests that patients differ from the general public in their valuations. Frick and colleagues (2012) reported on the importance of social relationships as determinants of health valuation, especially for health professionals. Health states hampering social relationships are judged as more disabling. Ubel and colleagues (2003) described several factors that may contribute to these discrepancies: adaptation effects (i.e., affected patients often adapt physically and emotionally to their health state, resulting in a more positive valuation of the respective state), focusing illusion (i.e., healthy people focus on impaired attributes, largely ignoring unchanged attributes

of a certain disease), and contrast effects (i.e., severely ill patients may underestimate the impact of lenient diseases, while healthy people may overestimate this impact). Conducting a survey in the general public will result in a weighted mixture of affected and healthy valuation perspective.

2. Interpretation/primary appraisal.

The interpretation of a health state depends on a subject's values, goals, and beliefs, as well as on the cognitive framing (Kahneman and Tversky 1984) and/or context (Schwarz 1999) of the health-state description.

3. Judgments on health states.

Like human judgments in general, these are not formed to fulfill the criteria of an exhaustive information processing. By contrast, they serve as decision rules to govern behavior (e.g., giving an answer in a questionnaire) and follow the principles of parsimony and functional pragmatism rather than coherence and rationality. Stiggelbout and de Vogel-Voogt (2008) identified various sources of biases that might be relevant in the context of health valuation, such as focusing illusion (see step 1), status quo bias (i.e., respondents are more sensitive to changes in their own health state compared with imagined health states), loss aversion (see Tversky and Kahneman 1992), or failure to anticipate negative events (i.e., poor hedonic forecasting). In addition, affects and mood are known to be highly influential during judgmental processing.

4. A deliberate editing of the response. In this last step, for example, a respondent's attempt to be compatible with perceived norms (e.g., perceived fairness, political correctness, or ethical considerations) further biases a subjective valuation (Rehm and Frick 2010).

Conclusion

Econometric elicitation methods were not originally developed for self-administered questionnaires. Given the many methodological risks of using this data collection mode, TTO, PTO, or SG elicitation methods are not recommended for paper-and-pencil surveys. Understanding the introductory scenarios and autonomously and successively approaching the point of indifference seems too complicated a task for lay respondents. Though VAS scales were developed specifically for self-administered questionnaires, their validity and reliability are too weak to measure the utilities of complex health states on the interval level. Choosing between rankings and PC tasks would mean a tradeoff between economy and validity of the measurement procedure.

Among PCs presented to respondents from the general public, those with the following characteristics seem to be most promising: (1) The number of pairs of health states should be limited (to a number determined by pre-analysis) so that annoyance effects or reactance can mostly be precluded. (2) Cognitive complexity of the health state descriptions should not exceed seven (plus or minus two) judgmental attributes (Miller 1956). However, this does not necessarily mean that health-state descriptions should be limited to seven dimensions or attributes, as respondents tend to

organize redundant information into broader superconcepts. That being said, this ability should also be evaluated prior to the survey. Applying these principles would allow surveys to pose complex vignettes to respondents. (3) To avoid biases due to the direction of a comparison (e.g., A versus B is not the same as B versus A) (Wänke 1996), presentation of health states within one comparison should be randomly balanced. To avoid order effects or carryover effects, factorial design techniques that also preclude repetitive presentations of certain health states (A versus B followed by C versus D and not by A versus C, for instance) should be used in the assignment of comparison tasks to respondents. Complex survey designs like the one proposed here require adequate techniques for statistical analysis (Hox et al. 1991). ■

Financial Disclosure

The authors declare that they have no competing financial interests.

References

- AYALON, L., AND KING-KALLIMANIS, B.L. Trading years for perfect health: Results from the Health and Retirement Study. *Journal of Aging and Health* 22:1184–1197, 2010. PMID: 20660638
- BADIA, X.; ROSET, M.; HERDMAN, M.; AND KIND, P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making* 21:7–16, 2001. PMID: 11206949
- BIJLENGA, D.; BIRNIE, E.; AND BONSEL, G.J. Feasibility, reliability and validity of three health-state valuation methods using multiple-outcome vignettes on moderate-risk pregnancy at term. *Value in Health* 12:821–827, 2009. PMID: 19508667
- BJÖRK, S., AND NORINDER, A. A weighting exercise for the Swedish version of the EuroQol. *Health Economics* 8:117–126, 1999. PMID: 10342725
- BLEICHRODT, H., AND JOHANNESSEN, M. An experimental test of a theoretical foundation for rating-scale valuations. *Medical Decision Making* 17:208–216, 1997. PMID: 9107617
- BOWLING, A. Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health* 27:281–291, 2005. PMID: 15870099
- BROOKS, R. Euroqol: The current state of play. *Health Policy* 37:53–72, 1996. PMID: 10158943
- BURSTRÖM, K.; JOHANNESSEN, M.; AND DIDERICHSEN, F. A comparison of individual and social time trade-off values for health-states in the general population. *Health Policy* 76:359–370, 2006. PMID: 16214258
- CHEVALIER, J., AND DE POUVOURVILLE, G. Valuing EQ-5D using time trade-off in France. *European Journal of Health Economics* 14(1):57–66, 2013. PMID: 21935715
- CLEEMPUT, I. A social preference valuations set for EQ-5D health states in Flanders, Belgium. *European Journal of Health Economics* 11:205–213, 2010. PMID: 19582490
- CRAIG, B.M.; BUSSCHBACH, J.J.; AND SALOMON, J.A. Keep it simple: Ranking health states yields similar to cardinal measurement approaches. *Journal of Clinical Epidemiology* 62:296–305, 2009a. PMID: 18945585
- CRAIG, B.M.; BUSSCHBACH, J.J.; AND SALOMON, J.A. Modeling ranking, time trade-off, and visual analogue scale values for EQ-5D health states: A review and comparison of methods. *Medical Care* 47:634–641, 2009b. PMID: 19433996
- CRITCHLOW, D.E., AND FLIGNER, M.A. Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation in GLIM. *Psychometrika* 56:517–533, 1991.
- DAMSCHRODER, L.J.; ROBERTS, T.R.; GOLDSTEIN, C.C.; ET AL. Trading people versus trading time: What is the difference? *Population Health Metrics* 3:10, 2005. PMID: 16281982
- DEVLIN, N.J.; HANSEN, P.; KIND, P.; AND WILLIAMS, A. Logical inconsistencies in survey respondents' health state valuations: A methodological challenge for estimating social tariffs. *Health Economics* 12:529–544, 2003. PMID: 12825206
- DEVLIN, N.J.; HANSEN, P.; AND SELAI, C. Understanding health state valuations: A qualitative analysis of respondents' comments. *Quality of Life Research* 13:1265–1277, 2004. PMID: 15473505
- DOLAN, P. Modeling valuations for EuroQol health states. *Medical Care* 35:1095–1108, 1997. PMID: 9366889
- DOLAN, P., AND KIND, P. Inconsistency and health state valuations. *Social Science & Medicine* 42:609–615, 1996. PMID: 8643985
- DOLDERS, M.G.; ZEEGERS, M.P.; GROOT, W.; AND AMENT, A. A meta-analysis demonstrates no significant differences between patient and population preferences. *Journal of Clinical Epidemiology* 59:653–664, 2006. PMID: 16765267
- ESSINK-BOT, M.L.; STOUTHARD, M.E.; AND BONSEL, G.J. Generalizability of valuations on health states collected with the EuroQol questionnaire. *Health Economics* 2:237–246, 1993. PMID: 8275169
- ETCHES, V.; FRANK, J.; DI RUGGIERO, E.; AND MANUEL, D. Measuring population health: A review of indicators. *Annual Review of Public Health* 27:29–55, 2006. PMID: 16533108
- FEENY, D.; FURLONG, W.; TORRANCE, G.W.; ET AL. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care* 40:113–128, 2002. PMID: 11802084
- FERREIRA, L.N.; FERREIRA, P.L.; ROWEN, D.; AND BRAZIER, J.E. Do Portuguese and UK health state values differ across valuation methods? *Quality of Life Research* 20:609–619, 2011. PMID: 21061071
- FLYNN, T.N. Using conjoint analysis and choice experiments to estimate QALY values: Issues to consider. *Pharmacoeconomics* 28:711–722, 2010. PMID: 20568837
- FLYNN, T.N.; LOUIERE, J.J.; PETERS, T.J.; AND COAST, J. Using discrete choice experiments to understand preferences for quality of life: Variance-scale heterogeneity matters. *Social Science & Medicine* 70:1957–1965, 2010. PMID: 20382460
- FRANCIS, B.; DITTRICH, R.; HATZINGER, R.; AND PENN, R. Analysing partial ranks by using smoothed paired comparison methods: An investigation of value orientation in Europe. *Applied Statistics* 51:319–336, 2002. PMID: 21633877
- FRICK, U.; IRVING, H.; AND REHM, J. Social relationships as a major determinant in the valuation of health states. *Quality of Life Research* 21:209–213, 2012. PMID: 21633877
- GREINER, W.; WEIJNEN, T.; NIEUWENHUIZEN, M.; ET AL. A single European currency for EQ-5D health states: Results from a six-country study. *European Journal of Health Economics* 4:222–231, 2003. PMID: 15609189
- GREINER, W.; CLAES, C.; BUSSCHBACH, J.J.; AND VON DER SCHULENBURG, J.M. Validating the EQ-5D with time trade off for the German population. *European Journal of Health Economics* 6:124–130, 2005. PMID: 19787848
- HAMMERSCHMIDT, T.; ZEITLER, H.P.; GULICH, M.; AND LEIDL, R. A comparison of different strategies to collect standard gamble utilities. *Medical Decision Making* 24:493–503, 2004. PMID: 15358998
- HOSMER, D.W., AND LEMESHOW, S. *Applied Logistic Regression* (2nd ed.). New York: Wiley and Sons, 2000.
- HOX, J.J.; KREFT, I.G.G.; AND HERMKENS, P.L.J. The analysis of factorial surveys. *Sociological Methods & Research* 19:493–510, 1991.
- IP, W.C.; CHIU, L.L.; AND KWAN, Y.K. Construction of health indices using paired comparisons. *Social Indicators Research* 67:353–373, 2004.
- JELSMAN, J.; HANSEN, K.; DE WEERTD, W.; ET AL. How do Zimbabweans value health states? *Population Health Metrics* 1:11, 2003. PMID: 14678566
- JO, M.W.; YUN, S.C.; AND LEE, S.I. Estimating quality weights for EQ-5D health states with the time trade-

- off method in South Korea. *Value in Health* 11:1186–1189, 2008. PMID: 18489498
- JOHNSON, J.A., AND PICKARD, A.S. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care* 38:115–121, 2000. PMID: 10630726
- JOHNSON, J.A.; COONS, S.J.; ERGO, A.; AND SZAVA-KOVATS, G. Valuation of EuroQol (EQ-5D) health states in an adult US sample. *Pharmacoeconomics* 13:421–433, 1998. PMID: 10178666
- KAHNEMAN, D., AND TVERSKY, A. Choices, values, and frames. *American Psychologist* 39:341–350, 1984.
- KIND, P. A comparison of two models for scaling health indicators. *International Journal of Epidemiology* 11:271–275, 1982. PMID: 7129741
- KIND, P. Applying paired comparisons models to EQ-5D valuations: Deriving TTO utilities from ordinal preference data. In: Kind, P., Brooks, R., and Rabin R., Eds. *EQ-5D Concepts and Methods: A Developmental History*. Dordrecht, The Netherlands: Springer, 2005, pp. 201–220.
- KRABBE, P.F. Thurstone scaling as a measurement method to quantify subjective health outcomes. *Medical Care* 46:357–365, 2008. PMID: 18362814
- KRABBE, P.F.; SALOMON, J.A., AND MURRAY, C.J. Quantification of health states with rank-based non-metric multidimensional scaling. *Medical Decision Making* 27:395–405, 2007. PMID: 17761959
- LAMERS, L.M.; MCDONNELL, J.; STALMEIER, P.F.; ET AL. The Dutch tariff: Results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* 15:1121–1132, 2006. PMID: 16786549
- LEE, Y.K.; NAM, H.S.; CHUANG, L.H.; ET AL. South Korean time trade-off values for EQ-5D health states: Modeling with observed values for 101 health states. *Value in Health* 12:1187–1193, 2009. PMID: 19659703
- LEIDL, R., AND REITMEIR, P. A value set for the EQ-5D based on experienced health states: Development and testing for the German population. *Pharmacoeconomics* 29:521–534, 2011. PMID: 21247225
- LENERT, L., AND KAPLAN, R.M. Validity and interpretation of preference-based measures of health-related quality of life. *Medical Care* 38(9 Suppl.):II 138–II 150, 2000. PMID: 10982099
- LENERT, L.A.; CHER, D.J.; GOLDSTEIN, M.K.; ET AL. The effect of search procedures on utility elicitation. *Medical Decision Making* 18:76–83, 1998. PMID: 9456212
- LINARES, P. Are inconsistent decisions better? An experiment with pairwise comparisons. *European Journal of Operational Research* 193:492–498, 2009.
- LOUVIERE, J.J.; STREET, D.; BURGESS, L.; ET AL. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1:128–163, 2008.
- LUNDBERG, L.; JOHANNESSON, M.; ISACSON, D.G.; AND BORGGQUIST, L. The relationship between health-state utilities and SF-12 in a general population. *Medical Decision Making* 19:128–140, 1999. PMID: 10231075
- LUNDQVIST, C.; BENTH, J.S.; GRANDE, R.B.; ET AL. A vertical VAS is a valid instrument for monitoring headache pain intensity. *Cephalalgia* 29:1034–1041, 2009. PMID: 19735531
- MAYDEU-OLIVARES, A., AND BÖCKENHOLT, U. Modeling subjective health outcomes: Top 10 reasons to use Thurstone's method. *Medical Care* 46:346–348, 2008. PMID: 18362812
- MCCABE C.; BRAZIER, J.; GILKS, P.; ET AL. Using rank data to estimate health state utility models. *Journal of Health Economics* 25:418–431, 2006. PMID: 16499981
- MCINTOSH, C.N.; GORBER, S.C.; BERNIER, J.; AND BERTHELOT, J.M. Eliciting Canadian population preferences for health states using Classification and Measurement System of Functional Health (CLAMES). *Chronic Diseases in Canada* 28:29–41, 2007. PMID: 17953796
- McKECHNIE, J.G., AND BRODIE, E.E. Hand and hand preferences in use of a visual analogue scale. *Perceptual and Motor Skills* 107:643–650, 2008. PMID: 19235396
- MEROPOL, N.J.; EGGLESTON, B.L.; BUZAGLO, J.S.; ET AL. Cancer patient preferences for quality and length of life. *Cancer* 113:3459–3466, 2008. PMID: 18988231
- MILLER, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97, 1956. PMID: 13310704
- PEETERS, Y., AND STIGGELBOUT, A.M. Health state valuations of patients and the general public analytically compared: A meta-analytical comparison of patient and population health state utilities. *Value in Health* 13:306–309, 2010. PMID: 19744288
- PRIETO, L., AND ALONSO, J. Exploring health preferences in sociodemographic and health related groups through the paired comparison of the items of the Nottingham Health Profile. *Journal of Epidemiology and Community Health* 54:537–543, 2000. PMID: 10846197
- RATCLIFFE, J.; BRAZIER, J.; TSUCHIYA, A.; ET AL. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Economics* 18:1261–1276, 2009. PMID: 19142985
- REHM, J., AND FRICK, U. Valuation of health states in the US study to establish disability weights: Lessons from the literature. *International Journal of Methods in Psychiatric Research* 19:18–33, 2010. PMID: 20191661
- REHM, J., AND FRICK, U. Establishing disability weights from pairwise comparisons for a US burden of disease study. *International Journal of Methods in Psychiatric Research* 22(2):144–154, 2013.
- ROBINSON A.; LOOMES G.; AND JONES-LEE, M. Visual analog scales, standard gambles, and relative risk aversion. *Medical Decision Making* 21:17–27, 2001. PMID: 11206943
- SALOMON, J.A. Reconsidering the use of rankings in the valuation of health states: A model estimating cardinal values from ordinal data. *Population Health Metrics* 1:12, 2003. PMID: 14687419
- SCHWARZ, N. Self-reports: How the question shape the answers. *American Psychologist* 54:93–105, 1999.
- SHAW, J.W.; PICKARD, A.S.; YU, S.; ET AL. A median model for predicting United States population-based EQ-5D health state preferences. *Value in Health* 13:278–288, 2010. PMID: 19961566
- SLOVIC, P. The construction of preference. *American Psychologist* 50:364–371, 1995.
- STIGGELBOUT, A.M.; DE HAES, J.C.; KIEBERT, G.M.; ET AL. Tradeoffs between quality and quantity of life: Development of the QQ questionnaire for cancer patient attitudes. *Medical Decision Making* 16:184–192, 1996. PMID: 8778537
- STIGGELBOUT, A.M., AND DE VOGEL-VOOGT, E. Health state utilities: A framework for studying the gap between the imagined and the real. *Value in Health* 11:76–87, 2008. PMID: 18237362
- STOLK, E.A.; OPPE, M.; SCALONE, L.; AND KRABBE, P.F. Discrete choice modeling for the quantification of health states: The case of EQ-5D. *Value in Health* 13:1005–1013, 2010. PMID: 20825618
- THURSTONE, L.L. A law of comparative judgment. *Psychological Review* 34:273–286, 1927.
- TORRANCE, G.W.; FEENEY, D.; AND FURLONG, W. Visual analog scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making* 21:329–334, 2001. PMID: 11475389
- TSUCHIYA, A.; IKEDA, S.; IKEGAMI, N.; ET AL. Estimating an EQ-5D population value set: The case of Japan. *Health Economics* 11:341–353, 2002. PMID: 12007165
- TVERSKY, A., AND KAHNEMAN, D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5:297–323, 1992.
- UBEL, P.A.; LOEWENSTEIN, G.; AND JEPSON, C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research* 12:599–607, 2003. PMID: 14516169
- VAN OSCH, S.M.C., AND STIGGELBOUT, A.M. Understanding VAS valuations: Qualitative data on the cognitive process. *Quality of Life Research* 14:2171–2175, 2005. PMID: 16328897
- VOOGT, E.; VAN DER HEIDE, A.; RIETJENS, J.A.; ET AL. Attitudes of patients with incurable cancer toward medical treatment in the last phase of life. *Journal of Clinical Oncology* 23:2012–2019, 2005. PMID: 15774792