

A computer Time Trade-Off: a feasible and reliable alternative for the interview Time Trade-Off in rheumatoid arthritis

L. Buitinga¹, L.M.A. Braakman-Jansen¹, E. Taal¹, M.A.F.J. van de Laar^{1,2}

¹IBR Research Institute for Social Sciences and Technology, University of Twente, Faculty of Behavioural Sciences, and Department of Psychology, Health and Technology, Enschede, The Netherlands;

²Department of Rheumatology and Clinical Immunology, Medisch Spectrum Twente, Enschede, The Netherlands.

Abstract

Objective

The Time Trade-Off (TTO) is an instrument used for valuing health-related quality of life. This study evaluated the test-retest reliability of a computer TTO in rheumatoid arthritis (RA), and compared the computer with the interview TTO regarding feasibility and agreement.

Methods

In study 1 using a cross-over design, thirty patients completed both TTOs. In study 2, twenty-nine other patients completed the computer TTO twice to examine test-retest reliability. Feasibility was measured by assessing actual and perceived time duration and general experience of the patient. Agreement between utility scores of both TTOs was measured by Bland-Altman analysis.

Results

Both TTOs were feasible. The computer TTO showed high test-retest reliability (ICC=0.88). Bland-Altman analysis showed a small mean difference (0.06, SD=0.14, effect size=0.30) between both TTOs. Limits of agreement were wide (-0.22 to 0.34). Differences between interview and computer TTO utilities did not vary over the range of scores.

Conclusion

The computer TTO was feasible and reliable, but did not provide similar results as the interview TTO. However, no systematic biases in the differences were found over the range of scores.

Key words

quality of life, patient preference, rheumatoid arthritis.

Laurien Buitinga, MSc

Louise M.A. Braakman-Jansen, PhD

Erik Taal, PhD

Mart A.F.J. van de Laar, MD, PhD

Please address correspondence
and reprint requests to:

Dr Laurien Buitinga,
IIBR Research Institute for Social
Sciences and Technology,
University of Twente,
Faculty of Behavioural Sciences,
Department of Psychology and
Health and Technology,
P.O. Box 217,
7500 AE Enschede, The Netherlands.
E-mail: l.buitinga@gw.utwente.nl

Received on November 26, 2010; accepted
in revised form on April 18, 2011.

© Copyright CLINICAL AND
EXPERIMENTAL RHEUMATOLOGY 2011.

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disease which has, as many chronic diseases, a great impact on the quality of life (1, 2). Beside disease activity, disability and quality of life are in particular important disease outcomes (3, 4). When the impact of health on an individual's functioning is examined, we speak of 'health-related quality of life' (HRQoL). HRQoL is a broad concept covering the overall impact of the illness and its treatment on patients, and also patients' reactions to this impact (5).

To measure health-related quality of life either descriptive methods or valuation methods can be used (6). The first type of method (*e.g.* SF-36) describes the health state of the patient, while the second type of method provides a value to the health state of the patient. A value, often termed 'utility', is typically scaled between 0.0 representing a health state judged equivalent to being dead and 1.0 representing perfect health. Utilities can be measured directly in a group of patients using the Standard Gamble (SG) (7) or Time Trade-Off (TTO) (8), whereby patients value their own health state. Questionnaires, such as the EQ-5D (9), SF-6D (10) and Health Utility Index (HUI) (11) measure utilities indirectly by attaching community derived utility weights to health states of patients.

The TTO has become a frequently used method. However, different variations in TTO methodology are used, which influence utilities. Aspects such as the way in which the TTO utilities are elicited (*i.e.* by computer, interview, or questionnaire), the way the TTO question is formulated or framed (12, 13), and the time frame that is chosen, differ between studies (14). In their review on the TTO method, Arnesen and Trommald concluded that there is no standardised TTO in its current use, because studies differed in application of TTO methodology (14).

Traditionally, TTO utilities are elicited by means of a personal interview. However, the administration of the TTO by an interviewer is time-consuming (15). Besides, the possibility of interviewer bias and social desirability can

be a threat to the validity of personal interviews (16). Computer-based utility assessment can offer a solution to such problems (16, 17). Computers are used increasingly to administer HRQoL questionnaires or preference-based instruments (18-22). Specific computer software programs for utility elicitation have been developed and used (*e.g.* U-Titer (18, 19, 23), or IMPACT (24)). However, no studies have compared the use of the computer TTO directly with the use of the interview TTO regarding feasibility and the extent to which both methods agree in their results.

The first aim of this study was to evaluate the test-retest reliability of a computer TTO in patients with Rheumatoid Arthritis. The second aim was to compare the computer TTO with the interview TTO regarding feasibility and agreement.

Methods

Patients and study design

Consecutive patients diagnosed with rheumatoid arthritis (RA) were recruited from the rheumatology outpatient clinic of a hospital in the Netherlands (Medisch Spectrum Twente, Enschede). Patients who did not understand the Dutch language were excluded. We aimed to recruit 60 patients: the first 30 for a study to compare the computer TTO with the interview TTO (study 1), the other 30 for a study to determine test-retest reliability of the computer test (study 2).

To compare both TTO measures, a cross-over design was used. Fifteen patients completed the computer TTO first, the other 15 completed the interview TTO first. For the assessment of test-retest reliability, a group of 30 patients were asked to conduct the computer TTO twice. In both studies there were approximately two weeks between the two TTO measures. We expected that this time interval would neither be too short nor too long to minimize the chance of respectively recall bias and significant health changes (25). Evaluation questionnaires were administered verbally after every test.

The TTO measures were administered either at the hospital or at the university, depending on the preference of the

Funding: This study was financially supported by a grant from the 'Stichting Onderzoek en Zorgontwikkeling Reumacentrum Twente'

Competing interests: none declared.

participant. Both measures were done at the same location. Randomisation for the first TTO to be either personal interview or computer test was done block wise with the use of sealed envelopes. Informed consent was obtained from all participants. According to local regulations in the Netherlands (WMO), no approval of the ethical review board was needed.

Measures

Pilot study

A pilot study was executed with five RA patients (aged between 48 and 81) of the research panel of the university to examine whether improvements in the TTOs were needed. Based on this study, some adaptations were made in the formulation of one question and in the graphical presentation of the computer TTO.

TTO interview

The Time Trade-Off question used in this study was formulated as follows:

“Imagine that a new treatment became available which helped you to recover fully. A side-effect of this treatment, however, is that you will die sooner. Would you opt for this treatment?”

For both the interview and the computer TTO a similarly formatted graphical aid was used for a good understanding of the question. When participants asked about the definition of being perfectly healthy, they were told to imagine being in perfect health without any disease or health-related complaints. The participants were visually shown two options: a life in current health according to their life expectancy as based on calculations of the Dutch Central Bureau of Statistics (CBS) (26) and a shorter life in perfect health. The remaining life expectancy was calculated by extracting the age of the participant from his or her expected age of dying according to the CBS. The trade-off started with setting the shorter life in perfect health on half of the remaining life expectancy. For example, a person with a remaining life expectancy of 20 years was first asked about his or her willingness to trade off 10 life years. If the person accepted the trade, a remaining life expectancy of 5 years in perfect health was presented. If the person did not accept the trade, a re-

maining life expectancy of 15 years was presented. This process continued until the patient was indifferent between his or her own current health state for his or her life expectancy and a shorter life in perfect health. The TTO score was calculated by the formula: 1-(number of life years given up/remaining expected life years).

Computer TTO

Touch screens were used to gather TTO utilities. The computer TTO was formulated in exactly the same way as the interview TTO as described above. Computer graphics supported the questions (Fig. 2). These graphics were made similar to the aid used in the interviews. During the computer assessment of the TTO, the researcher (LB) was present to start up the computer. The respondents completed the TTO independently. The researcher observed the patients during the assessment.

Feasibility questionnaire

Aspects of feasibility were measured by an evaluation questionnaire. For both methods the same questionnaire was used. Questions were about the duration ('good' / 'too long'), the clear-

ness of the introduction and conclusion ('yes' / 'no') and the general experience with the TTO ('unpleasant' / 'neutral' / 'pleasant'). For the computer test there was an additional question about the difficulty of using the touch screen ('difficult' / 'neutral' / 'easy'). Patients who completed both the personal interview and the computer test were asked for their preference between both TTOs ('personal interview' / 'computer test' / 'indifferent'). When participants did not finish the TTO, they were asked about the reason for not completing it. Participants were able to explain their answers by open questions.

Clinical characteristics

Demographic and clinical characteristics were gathered (Table I). Clinical characteristics were assessed by using a numerical rating scale measuring pain and the Health Assessment Questionnaire Disability Index (HAQ-DI) (27) measuring functional disability.

Statistical analysis

Non-parametric tests were used, because of the non-normal distribution of the data and because of the ordinal character of most variables.

Table I. Demographic and clinical characteristics of the RA patients attending study 1 (interview *versus* computer TTO) and study 2 (test-retest of computer TTO).

	Study 1 n (30)	Study 2 n (29)
Mean age (SD)	58 (13)	56 (11)
Gender (%)		
Men	27	34
Women	73	66
Marital status (%)		
Single	17	21
Married/Living together	83	79
Educational level (%)		
Low	50	59
Moderate	30	21
High	20	20
Work status (%)		
Paid work	40	42
Housekeeping	10	10
Retired/unemployed/disabled	50	48
Disease duration (median years) (IQR)	10 (19-3)	10 (18-6)
Mean pain (SD) (numerical rating scale 0-10)	4.1 (2.3)	3.1 (2.4)
Mean functional disability (SD) (HAQ-DI 0-3)	0.87 (.60)	0.52 (.53)

*HAQ-DI: Health Assessment Questionnaire Disability Index; a self-report measure of functional status (disability) consisting of eight categories: dressing and grooming, arising, eating, walking, hygiene, reach, grip, common daily activities.

To examine whether computer-based and interview-based utilities differed between participants who started with the interview or computer TTO, a Mann-Whitney U-test was executed. Differences in completion time between the interview TTO and computer TTO were tested with the Wilcoxon signed rank test.

Test-retest reliability was examined by determining the Intraclass Correlation Coefficient (ICC) in the subgroup of patients who administered the computer version twice.

To test whether the median utility score of the interview TTO differed from the median utility score of the computer TTO, a Wilcoxon signed rank test was executed.

To determine whether the computer TTO can replace the face-to-face interview, individual agreement between both versions was assessed by the Bland-Altman method (28). Individual differences between the interview TTO and the computer TTO were plotted against their mean to measure this agreement. Limits of agreement were calculated to examine whether the utility scores of both methods agreed sufficiently. The mean difference and limits of agreement were plotted as lines. Limits of agreement ($d-1.96$ SD and $d+1.96$ S.D.) were calculated using the mean difference (d) and the standard deviation of the differences (SD). Ninety-five percent of differences between measurements by the two methods would lie between these limits, if these differences were normally distributed. We checked the normal distribution by histogram. Furthermore, confidence intervals for the limits of agreement were calculated with the formula $\sqrt{3 \text{ SD}^2/n}$ (28). The data were analysed using SPSS version 16.0.

Results

Subjects

Consecutive patients were asked to participate. Eventually, complete data could be collected on 59 patients (mean age 57 years). One patient in the test-retest study did not complete the study, because of personal circumstances. From another person data were missing because of a temporary technical

Table II. Feasibility of the TTO (study 1: interview *versus* computer TTO).

	Interview TTO (n=30)	Computer TTO (n=30)
Median time duration (minutes)*	5.2 (IQR 4-7)	2.7 (IQR 1-3)
<i>Perceived time duration</i>		
Too long	2	1
Good	27	29
<i>Introduction clear?</i>		
Yes	30	30
No	0	0
<i>Conclusion understandable?</i>		
Yes	30	30
No	0	0
<i>General experience**</i>		
Unpleasant	1	3
Neutral	7	17
Pleasant	22	10

*($p<0.001$) (Wilcoxon signed rank test); **($p=0.049$) (Chi-square test; likelihood ratio).

computer malfunction. Demographic and clinical characteristics of the 59 patients attending study 1 (interview *versus* computer TTO) or study 2 (test-retest of computer TTO) are given in Table I.

Test order

An effect of test order was not found for utility scores on the interview ($U=89$, $p=0.73$) or the computer test ($U=102$, $p=0.37$).

Feasibility

The median duration for the completion of the computer TTO was significantly shorter than the median duration of the interview TTO (Table II). Eleven participants preferred the interview TTO, four participants preferred the computer TTO, and 11 were neutral. Preferences for the interview TTO were mainly attributed to the preference for personal contact. Nobody judged the computer program as difficult to use. Further results are shown in Table II. The only significant difference between the interview and computer TTO was found for general experience. The interview TTO was most often judged as pleasant, while people were more neutral about their experience with the computer TTO. Reasons for the feelings of unpleasantness were that it was unrealistic and confronting to trade off life years. During the computer assessment of the TTO, participants had difficulties with understanding the conclusion. The con-

clusion was framed in terms of losses (the number of life years people had to give up for perfect health), whereas the questions of the TTO were framed in terms of gains (the number of life years people still had in perfect health). The researcher explained that this loss in the conclusion corresponded to the life years people accepted to have in perfect health as stated in the questions.

Test-retest reliability

A high test-retest reliability of 0.88 (95% CI 0.75 to 0.94) was shown for the computer TTO, as indicated by the Intraclass Correlation Coefficient (ICC).

Agreement between the interview TTO and the computer TTO

The median utility score of the interview TTO (0.87 IQR 0.80–1.00) differed significantly from the median utility score of the computer TTO (0.85 IQR 0.58–1.00) ($p=0.04$).

The agreement between the interview and the computer TTO is shown in Fig. 1. Despite a high correlation between the interview and computer TTO (ICC=0.75; 95% CI 0.54 to 0.87), agreement was poor. The mean difference between the interview and computer TTO was a small difference (0.06, S.D.=0.14, effect size = 0.30) and limits of agreement were wide (−0.22 to 0.34). Differences between the two methods were approximately normally distributed. The 95% CI for the limits of agreement were −0.31 to −0.13 for the lower limit, and 0.25 to

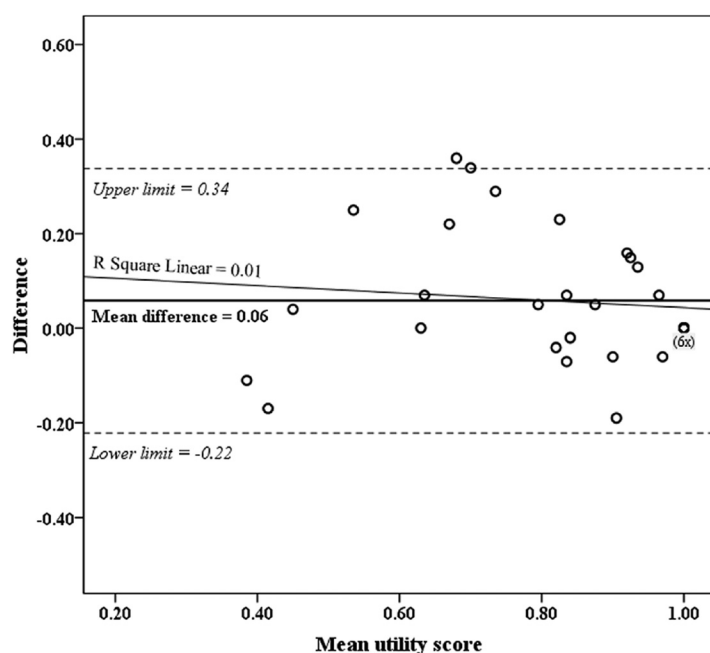


Fig. 1. Bland-Altman plot: agreement between the interview and computer TTO of the 30 participants attending study 1 (interview versus computer TTO).

and reliable instrument for measuring utilities. The computer TTO took significantly less time to complete. Although 11 participants preferred the interview compared with four who preferred the computer TTO, most participants indicated the computer TTO as pleasant or neutral. Only a few indicated it as unpleasant. Preference for personal contact was the main reason to prefer the interview TTO to the computer TTO.

The computer TTO did not give similar results as the interview TTO, because of a small mean difference between the interview and computer TTO and wide limits of agreement. This implies restraints in comparing TTO utilities obtained by interview with those obtained by computer if this difference is a relevant one. Assessing the minimal important difference can support the interpretation of this statistical difference. A study in laryngeal cancer (29) has attempted to establish the MID for the TTO ($\pm 5\%$ of the maximum instrument score = 0.05) and it is argued that

0.43 for the upper limit. The slight negative slope in the regression line showed that differences did not vary over the range of utility scores ($r^2 = 0.01$).

Discussion

This first study comparing a computer TTO with an interview TTO showed that the computer TTO was a feasible

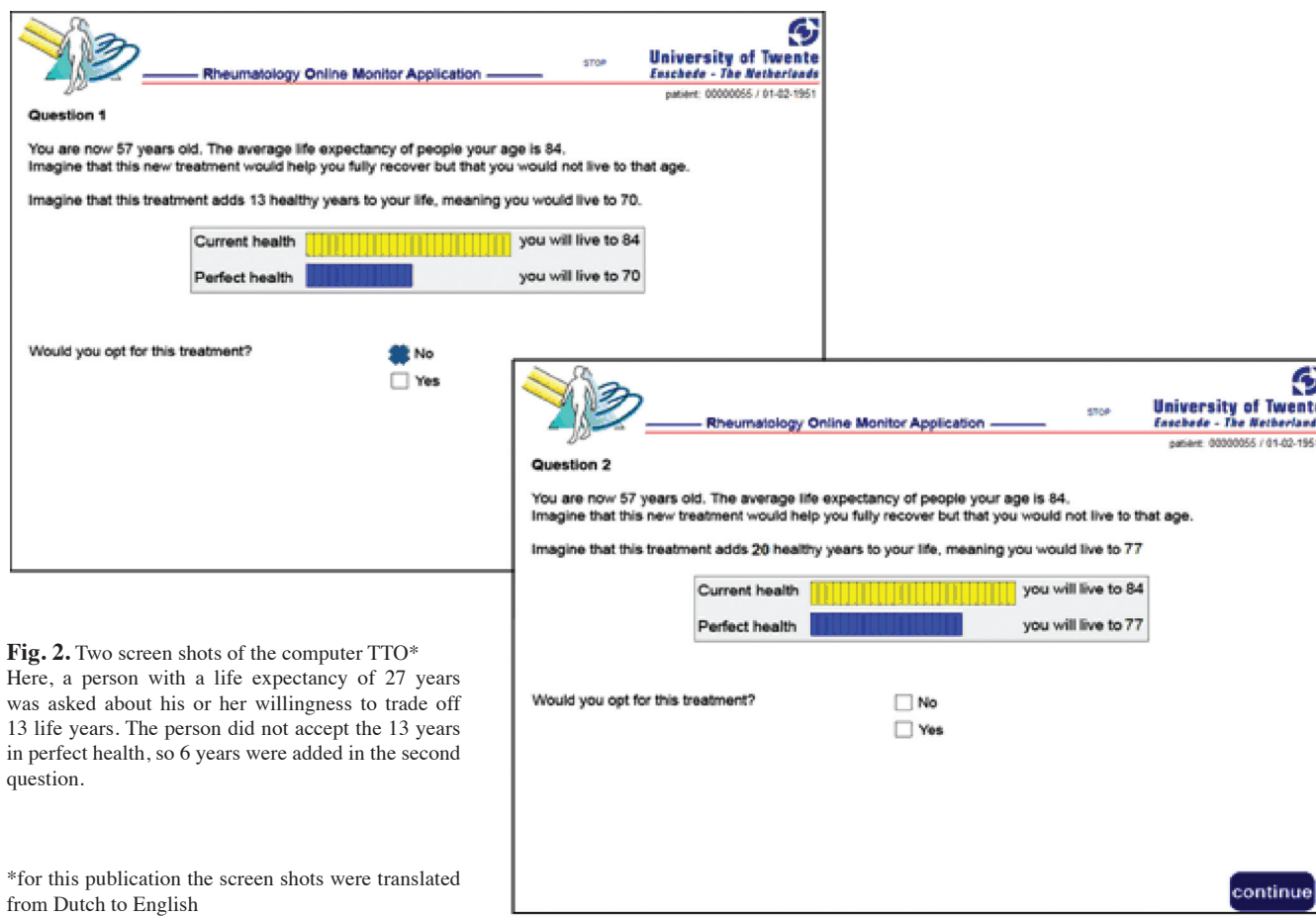


Fig. 2. Two screen shots of the computer TTO*. Here, a person with a life expectancy of 27 years was asked about his or her willingness to trade off 13 life years. The person did not accept the 13 years in perfect health, so 6 years were added in the second question.

*for this publication the screen shots were translated from Dutch to English

most published MID estimates range between 5% and 10% of the instrument range. So, although the mean difference of 0.06 seems to be small, it actually indicates a relevant difference using 0.05 as a benchmark. Unfortunately, the MID has not been established yet for the TTO in RA patients. However, no systematic biases in the differences between the interview and computer TTO were found over the range of scores. On average, the interview TTO provided slightly higher utility scores than the computer TTO. A possible, but speculative explanation for this difference is the presence of an interviewer bias. Possibly, patients were reluctant to trade off more life years in the presence of an interviewer, resulting in higher utilities on the interview TTO. In study 2, the high Intraclass Correlation between the two assessments with the computer TTO indicated good test-retest reliability.

Other studies using a computer TTO in the general public (22) or interview TTO in RA (30, 31) also showed a good feasibility. However, in these studies only completion rates and/or failure rates were assessed. Concerning test-retest reliability, we found in our study much higher test-retest reliability than was found by Lenert *et al.* (24), but their time interval between testing was longer than in our study. Other studies in RA also demonstrated good test-retest reliability for the interview TTO (period between measurements comprised one week (ICC=0.89) (30) or two weeks (ICC=0.85) (31)), comparable with the test-retest reliability of 0.88 for the computer TTO in this study.

A limitation of this study is the fact that although it was possible to compare the computer TTO with the interview TTO because of the similar formats used, it is difficult to compare our computer TTO with other computer programs for utility measurement, because programs such as IMPACT3 or U-Titer have different procedural options and representations available. Additionally, the use of a convenience sample limits the generalizability of the results. Besides, one needs to bear in mind that different variations in the procedure of the TTO are used to gather TTO utilities (14).

Our study applied just one of these procedures. Therefore, results can be different if other procedures are used.

One adaptation is needed in the conclusion of the TTO to enable people to complete the computer TTO without assistance. The conclusion that is framed in terms of losses should be framed both in terms of losses and in terms of gains. The reason to frame the conclusion only in terms of losses was to emphasize the trade. Although people would become perfectly healthy, they would also die earlier. Fortunately, this could not have affected the utilities obtained, because the trade was already made during the questions. With these adaptations in the presentation of the computer TTO, we expect that the computer TTO can be conducted without assistance. This would result in a less time-consuming and less costly administration of this utility instrument. However, further research is needed in larger samples of patients to verify our expectation.

This is the first study that compared the computer TTO directly with the interview TTO. In conclusion, the computer TTO is a feasible and reliable instrument for measuring utilities, which reduces the amount of administration time for the interviewer and for patients. The small mean difference between the interview TTO and computer TTO and the wide limits of agreement imply that utilities obtained with both TTOs are not interchangeable. However, no systematic biases in the differences were found over the range of scores.

Acknowledgements

We would like to thank the participants and the rheumatology department of Medisch Spectrum Twente in Enschede for their contribution to this study. We are also very grateful to André Brands, who programmed the computer TTO.

References

1. RUSSAK SM, CROFT JR JD, FURST DE *et al.*: The use of rheumatoid arthritis health-related quality of life patient questionnaires in clinical practice: Lessons learned. *Arthritis Care and Res* 2003; 49: 574-84.
2. STRAND V, KHANNA D: The impact of rheumatoid arthritis and treatment on patients' lives. *Clin Exp Rheumatol* 2010; 28 (Suppl. 59): S32-S40.
3. SMOLEN JS, ALETAHA D: The assessment of

disease activity in rheumatoid arthritis. *Clin Exp Rheumatol* 2010; 28 (Suppl. 59): S18-S27.

4. TAYLOR PC: The importance of the patients' experience of RA compared with clinical measures of disease activity. *Clin Exp Rheumatol* 2010; 28 (Suppl. 59): S28-S31.
5. KILTZ U, VAN DER HEIJDE D: Health-related quality of life in patients with rheumatoid arthritis and in patients with ankylosing spondylitis. *Clin Exp Rheumatol* 2010; 27 (Suppl. 55): S108-S111.
6. RUTTEN-VAN MÖLKEN MPMH, VAN BUSSCH-BACH JJ, RUTTEN FFH: *Van kosten tot effecten: een handleiding voor evaluatiestudies in de gezondheidszorg*. 1st ed., Maarssen, Elsevier gezondheidszorg, 2001.
7. TORRANCE GW: Toward a utility theory foundation for health status index models. *Health Serv Res* 1976; 11: 349-69.
8. TORRANCE GW, THOMAS WH, SACKETT DL: A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972; 7: 118-33.
9. BROOKS R, DE CHARRO F: EuroQol: The current state of play. *Health Policy* 1996; 37: 53-72.
10. BRAZIER J, ROBERTS J, DEVERILL M: The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; 21: 271-92.
11. HORSMAN J, FURLONG W, FEENY D, TORRANCE G: The Health Utilities Index (HUI®): Concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003; 1: 1-13.
12. BUCKINGHAM JK, BIRDSALL J, DOUGLAS JG: Comparing three versions of the time tradeoff: Time for a change? *Med Decis Making* 1996; 16: 335-47.
13. UBEL PA, LOEWENSTEIN G, SCANLON D, KAMLET M: Individual utilities are inconsistent with rationing choices: A partial explanation of why oregon's cost-effectiveness list failed. *Med Decis Making* 1996; 16: 108-16.
14. ARNESEN T, TROMMALD M: Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Econ* 2005; 14: 39-53.
15. LENERT LA, SHERBOURNE CD, REYNA V: Utility elicitation using single-item questions compared with a computerized interview. *Med Decis Making* 2001; 21: 97-104.
16. FOLEY KL, MANUEL J, VITOLINS M: The utility of self-report in medical outcomes research. *Evid Based Healthc Public Health* 2005; 9: 263-4.
17. SAIGAL CS, GORNBEIN J, REID K, LITWIN MS: Stability of time trade-off utilities for health states associated with the treatment of prostate cancer. *Qual Life Res* 2002; 11: 405-14.
18. TOSTESON ANA, KNEELAND TS, NEASE RF, SUMNER W: Automated current health time-trade-off assessments in women's health. *Value Health* 2002; 5: 98-105.
19. KERRIGAN CL, DALE COLLINS E, KNEELAND TS *et al.*: Measuring health state preferences in women with breast hypertrophy. *Plast Reconstr Surg* 2000; 106: 280-8.
20. VELIKOVA G, WRIGHT EP, SMITH AB *et al.*: Automated collection of quality-of-life data:

- A comparison of paper and computer touch-screen questionnaires. *J Clin Oncol* 1999; 17: 998-1007.
21. LEE SJ, KAVANAUGH A, LENERT L: Electronic and computer-generated patient questionnaires in standard care. *Best Pract Res Clin Rheumatol* 2007; 21: 637-47.
 22. NEASE JR RF, TSAI R, HYNES LM, LITTENBERG B: Automated utility assessment of global health. *Qual Life Res* 1996; 5: 175-82.
 23. SUMNER W, NEASE R, LITTENBERG B: Utiter: a utility assessment tool. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*. Washington D.C. 1991: 701-5.
 24. LENERT LA, STURLEY A, WATSON ME: iM-PACT3: Internet-based development and administration of utility elicitation protocols. *Med Decis Making* 2002; 22: 464-74.
 25. STREINER DL, NORMAN GR: *Health measurement scales: a practical guide to their development and use*. 3rd ed., Oxford, Oxford University Press, 2003.
 26. CENTRAAL BUREAU VOOR DE STATISTIEK: Overlevingstafels 2006. Available from: <http://www.cbs.nl>. Accessed 17 december 2007.
 27. FRIES JF, SPITZ P, KRAINES RG, HOLMAN HR: Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980; 23: 137-45.
 28. BLAND JM, ALTMAN DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10.
 29. RINGASH J, O' SULLIVAN B, BEZJAK A, REDELMEIER DA: Interpreting clinically significant changes in patient-reported outcomes. *Cancer* 2007; 110: 196-202.
 30. BEJIA I, SALEM KB, TOUZI M, BERGAOUI N: Measuring utilities by the time trade-off method in Tunisian rheumatoid arthritis patients. *Clin Rheumatol* 2006; 25: 38-41.
 31. TIJHUIS GJ, JANSEN SJT, STIGGELBOUT AM, ZWINDERMAN AH, HAZES JMW, VLIET VLIELAND TPM: Value of the time trade off method for measuring utilities in patients with rheumatoid arthritis. *Ann Rheum Dis* 2000; 59: 892-7.