

VALUATION OF QUALITY OF LIFE IN RHEUMATOID ARTHRITIS

LAURIEN BUTTINGA



**VALUATION OF QUALITY OF LIFE IN
RHEUMATOID ARTHRITIS**

Laurien Buitinga



Thesis, University of Twente, 2012

© Laurien Buitinga

ISBN: 978-90-365-3467-3

Printed by Gildeprint Drukkerijen, Enschede, The Netherlands

The studies in this thesis were performed at the department of Psychology, Health and Technology (PHT) of the University of Twente (Enschede) and the Arthritis Center Twente at Medisch Spectrum Twente (Enschede). The rheumatology research program of PHT is financially supported by the Dutch Arthritis Foundation (het Reumafonds). This project is financially supported by the foundation 'Stichting Onderzoek en Zorgontwikkeling Reumacentrum Twente'.

Publication of this thesis was financially supported by the Dutch Arthritis Foundation (het Reumafonds).

VALUATION OF QUALITY OF LIFE IN RHEUMATOID ARTHRITIS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 20 december 2012 om 16.45 uur

door

Laurien Buitinga
geboren op 9 maart 1985
te Hengelo (OV)

Dit proefschrift is goedgekeurd door de promotor prof. dr. M.A.F.J. van de Laar en de assistent-promotoren dr. E. Taal en dr. L.M.A. Braakman-Jansen.

Samenstelling promotiecommissie

Promotor: prof. dr. M.A.F.J. van de Laar

Assistent-promotoren: dr. E. Taal

dr. L.M.A. Braakman-Jansen

Leden: dr. G.J. Westerhof
(Universiteit Twente)

prof. dr. K.M.G. Schreurs
(Universiteit Twente)

prof. dr. M.J. IJzerman
(Universiteit Twente)

dr. A.E.R.H.C. Boonen
(Maastricht Universitair Medisch Centrum)

dr. P.F.M. Krabbe
(Universitair Medisch Centrum Groningen)


prof. dr. J.L. Severens
(Erasmus Universiteit Rotterdam)

Contents

Chapter 1	Objective & outline	10
Chapter 2	Theoretical background	16
Chapter 3	Comparative responsiveness of the EuroQol-5D and Short Form 6D to improvement in patients with rheumatoid arthritis treated with Tumor Necrosis Factor Blockers: results of the Dutch Rheumatoid Arthritis Monitoring Registry	38
Chapter 4	A computer Time Trade-Off: a feasible and reliable alternative for the interview Time Trade-Off in rheumatoid arthritis	54
Chapter 5	Construct validity of the interview Time Trade-Off and computer Time Trade-Off in patients with rheumatoid arthritis: a cross-sectional observational pilot study	68
Chapter 6	Future expectations and worst-case future scenarios of patients with rheumatoid arthritis: a focus group study	84
Chapter 7	Worst-case future scenarios of patients with rheumatoid arthritis: a cross-sectional study	98
Chapter 8	Chapter summary & discussion	112
	Samenvatting (Dutch summary)	124
	Dankwoord (Acknowledgements)	130
	Curriculum Vitae	134
	List of publications	138

Chapter

Objective & outline



Patient-reported outcomes (PROs) including health-related quality of life (HRQoL) instruments have gained attention in the medical world and in health economics. An increased awareness of the importance of using PROs alongside clinical measures has led to their expanded use in clinical trials. With the rising societal expenditures in health-care and the relatively limited resources, it is inevitable that choices must be made with respect to reimbursement of health technologies. Within the field of rheumatology, the economic burden of rheumatoid arthritis (RA) on society is high in terms of health-care costs and costs due to loss in work productivity (1,2). The expanding indications for the use of effective but very expensive biological therapies and the development of novel biological therapies warrant economic evaluations in which the expected costs of these therapies are compared with the expected benefits. One important form of economic analysis, namely cost-utility analysis (CUA), incorporates HRQoL instruments into the assessment of benefits. Although these HRQoL instruments are frequently carried out within an economic analysis, the performance of some of them in studies with RA patients has not been examined extensively. In addition, different variants of some of these instruments are used, which makes comparison of results between studies difficult. Furthermore, research has shown that the measurement of HRQoL can be biased (3-7). Therefore, additional research is required into the psychometric performance of these instruments and whether they can be improved for patients with RA. The objective of this thesis was to examine the psychometric properties of some HRQoL instruments used in economic analysis, and to examine alternative anchors for one HRQoL instrument in RA patients, which could improve the measurement of HRQoL representing the patient perspective.

The responsiveness of HRQoL questionnaires in rheumatoid arthritis

For CUAs of health technologies, utilities are commonly measured with the EuroQol-5D (EQ-5D) or Short Form-6D (SF-6D). A review of Harrison et al. (8) has shown that there is no conclusive evidence to favour any one measure (EQ-5D, SF-6D, Health Utility Index (HUI-2 or HUI-3)) with respect to responsiveness to improvement, i.e. their ability to detect clinical improvements. The authors conclude that more head-to-head comparisons are needed in longitudinal studies across the spectrum of RA disease severity. Therefore, they performed another study comparing the EQ-5D and SF-6D (9) and demonstrated in accordance with most previous studies into RA that the SF-6D was most responsive to improvement. One study of RA showed, however, that the EQ-5D was more responsive than the SF-6D (10). Nevertheless, these studies examined different aspects of responsiveness, i.e. internal or external responsiveness. To contribute to the evidence of both internal and external responsiveness in RA patients with active disease, in **chapter 3** the responsiveness of the EQ-5D and SF-6D to improvement was compared in patients with active RA treated with TNF-blockers, using both aspects of responsiveness.

Computer versus interview administration of an HRQoL instrument in RA

The Time Trade-Off (TTO) and Standard Gamble (SG) are frequently used instruments for measuring HRQoL by obtaining preferences for health states ('utilities') directly in patients. The SG has good theoretical underpinnings (11), but is a complex instrument that requires respondents to incorporate risk assessment into their responses. The TTO has been developed to be an easy alternative for the SG with similar empirical properties as the SG, and overcomes the problems of explaining probabilities to respondents (12-14). The TTO asks respondents to indicate how many years of life they are willing to trade in order to become perfectly healthy, i.e. to avoid being in their current health state. It is assumed



that the more life years respondents are prepared to trade, the worse their health state is. Evidence exists that TTO utilities better reflect individual preferences than SG utilities (15,16). Nevertheless, great variations in the TTO methodology are applied in research (17), which limits the comparability between studies, and has implications for the use of this instrument in economic evaluations. Traditionally, the TTO is administered in the form of an interview. An interview, however, has disadvantages including a long administration time, possible interview bias and social desirability that can be a threat to the validity of personal interviews. Computer-based TTOs can offer a solution to these problems. Although computer-based TTOs have been applied, no study has compared the use of the computer-based TTO directly with the use of the interview format of the TTO. In **chapter 4** a computer TTO has been developed and examined in RA patients with respect to its test-retest reliability and compared with the interview TTO regarding feasibility and level of agreement.

Validity of an HRQoL instrument in RA

Although the TTO is frequently used, only a few studies have examined the validity of the TTO in patients with RA. The few studies that reported on the construct validity showed poor to moderate correlations between the TTO and other disease-related patient-reported outcomes (PROs), also referred to as self-reported health outcomes (18-20). In addition, contradicting results regarding its discriminative ability were found (18-20). **Chapter 5** reports about results concerning the construct validity of both the interview and the computer TTO in RA patients, using a standardized TTO procedure.

Adaptation of an HRQoL instrument for RA

Previous studies, including our own study described in chapter 5, have shown that there is much scope for improving the construct validity of the TTO in RA patients. One possible explanation for the poor results could be the fact that trading off life years, i.e. 'dying earlier', is not perceived by RA patients as a realistic consequence of RA. Therefore, patients may not be willing to relinquish any years of their life in return for perfect health, giving the impression that their health is perfect. This will lead to an overestimation of patients' HRQoL, thereby limiting the validity and responsiveness of the instrument. Studies have shown that people do not always trade life years for perfect health ('zero-traders'). A range in the proportions of zero-traders from 1.7% to 58% of participants has been found (17). Zero-traders can lead to ceiling effects as they score the highest possible TTO score (1.00). Replacing the current anchor 'death' by an anchor that is perceived as a realistic consequence of RA from the perspective of patients may reduce the number of zero-traders, which could improve the validity and responsiveness of the TTO in patients with RA. In **chapter 6** the results of a focus group study are reported examining patients' future expectations and worst-case future scenarios. In **chapter 7** a quantitative study showed which future scenario was the most appropriate for RA patients.

Possibilities for improving HRQoL measurement

Finally, in **chapter 8**, this thesis ends with a summary and discussion of the findings and with a general discussion focused on the possibilities for improving the measurement of HRQoL.

REFERENCES

1. Boonen A, Severens JL. The burden of illness of rheumatoid arthritis. *Clin Rheumatol* 2011;30(Suppl. 1):S3-S8.
2. Zhang W, Anis AH. The economic burden of rheumatoid arthritis: beyond health care costs. *Clin Rheumatol* 2011; 30(Suppl. 1):S25-S32.
3. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002;11:447-56.
4. van Osch SMC, Wakker PP, van Den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making* 2004;24:511-7.
5. Bleichrodt H, Pinto JL. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Manage Sci* 2000;46:1485-96.
6. Gonzalez R, Wu G. On the shape of the probability weighting function. *Cogn Psychol* 1999;38:129-66.
7. Delquie P. Inconsistent trade-offs between attributes: new evidence in preference assessment biases. *Manage Sci* 1993;39:1382-95.
8. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DPM. The validity and responsiveness of generic utility measures in rheumatoid arthritis: A review. *J Rheumatol* 2008;35:592-602.
9. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Verstappen SMM, Watson K, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Qual Life Res* 2009;18:1195-1205.
10. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care* 2003;41:791-801.
11. von Neumann J, Morgenstern, O. *Theory of games and economic behavior*. Princeton: Princeton University Press; 1944.
12. Drummond MF, Sculpher MJ, Torrance GW. *Methods for the economic evaluation of health care programmes*. New York: Oxford University Press; 2005.
13. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;3:1-164.
14. Buckingham K, Devlin N. A theoretical framework for TTO valuations of health. *Health Econ* 2006;15:1149-54.
15. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;15:209-31.
16. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: Experimental results on the ranking properties of QALYs. *J Health Econ* 1997; 16:155-75.
17. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Econ* 2005;14:39-53.
18. 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.
19. 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.
20. Witney AG, Treharne GJ, Tavakoli M, Lyons AC, Vincent K, Scott DL, et al. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology* 2006;45:975-81.

Chapter 2

Theoretical background

Rheumatoid arthritis

Rheumatoid arthritis (RA) is a systemic inflammatory disease that manifests in the synovial membrane, resulting in the structural damage of cartilage, bone and ligaments (1). Its prevalence is around 0.5% to 1.0% of the adult population in developed countries, with a higher prevalence in women than in men (2,3). The course of RA differs between patients, and its severity can range from self-limiting disease to severe destruction and systemic complications (1,4). RA affects patients physically, psychologically and socially (5-7). Patients experience pain, joint swelling, stiffness (1), functional limitations (8) and fatigue (9-11). In addition, they report anxiety and depressive symptoms (12-14) and concerns e.g. about increased physical limitations and the loss of independence (15-17). Furthermore, patients experience a loss of independence (18-20) and restrictions in participation, i.e. a decrease in socialising and restrictions in remunerative employment (21). RA does not only have a significant impact on the individual affected by the disease, but also on society as a whole in terms of societal health care costs and costs due to loss in work productivity (5,22).

Treatment of RA

A few decades ago, treatments of RA were neither effective nor well tolerated, causing ongoing inflammation that resulted in joint deformity (23). RA was considered to be a ruthless disease with devastating progressive consequences and with little hope for a positive long-term prognosis (24). Nevertheless, as a result of advances in the treatment and management of RA, disease activity can now be controlled and structural damage can be prevented, resulting in less joint deformity. Novel treatments and treatment strategies have improved patients' symptoms, and can elevate the clinical status of a significant proportion of patients to a state of remission (23,25). In this state, patients do not have apparent disease activity, joint damage and disease progression (23). Requirements for preventing irreversible damage, and thus for enabling patients to achieve remission, are early diagnosis and treatment (23). Despite considerable improvements in the treatment of the majority of RA patients, in some patients the disease remains progressive (26).

Patient-reported outcomes

Clinicians tend to rely solely on biological, objective clinical measures revealing a patient's physiological status to measure the disease activity in RA and evaluate responses to treatment. However, more recently, attention has been broadened to include the use of outcomes that incorporate the patient's perspective alongside these clinical measures. Evidence demonstrating differences between patients and clinicians regarding their ratings of various symptoms and their beliefs about the importance of symptoms underlines the importance of using patient-reported outcome measures (PROs) (27). Obviously, it is not enough to cure diseases in patients, but restoration of all aspects of health is the ultimate goal of health-care. The use of the patient's perspective corresponds to the definition of health by the World Health Organization (WHO) in 1948 in which health is defined as a 'state of complete physical, mental and social well-being and not merely the absence of disease or infirmity' (28). The definition includes the concept of well-being that requires subjective evaluation by the people concerned, and emphasizes that health is more than simply the absence of disease.

Since the recognition of the patient as a valuable source of information for assessing the disease's impact and effect of treatment, the number of PROs has expanded tremendously. Patients' perceptions about their pain level, functional status and health status are frequently measured within the field of rheumatology (29). Pain level and functional status are more narrow domains of health, whereas health

status is broader, referring to patients' subjective assessment of their physical and psychological state (30). The term 'health status' is often used interchangeably with the term 'physical status' and 'health-related quality of life' (HRQoL) (31). Formally, these terms are not similar, but because of the absence of a clear conceptualization of HRQoL and lack of corresponding instruments, the term seems to cover all possible aspects related to health (31). At a conceptual level, HRQoL is often considered to include valuations or judgements with respect to domains of physical, psychological and social functioning (30), for example in terms of 'satisfaction' or 'annoyance' (31). When considering, however, the use of the term HRQoL, it can be observed that the term is frequently used to denote descriptive instruments that assess patients' level of functioning in various domains. Within descriptive instruments, generic HRQoL and disease-specific HRQoL instruments can be distinguished. Generic instruments of HRQoL measure broad aspects and, therefore, they enable comparisons of HRQoL across diseases (e.g. Short Form (SF)-36 (32)). Disease-specific instruments of HRQoL are only designed for specific diseases (e.g. the Rheumatoid Arthritis Quality of Life Scale (RAQoL) (33)). Besides descriptive HRQoL instruments, instruments have also been developed that incorporate importance weights or preferences into the measurement of HRQoL. A measure of QoL that incorporates importance weights is the Schedule for the Evaluation of Individual Quality of Life (SEIQoL) (34). The SEIQoL is an individualized measure that allows individuals to identify factors that are important to them and to indicate the relative importance of each factor to overall QoL. Instruments that incorporate preferences into the measurement of HRQoL (so-called preference-based or utility measures) are used in health economic analysis. These preferences can be obtained in two ways. First, preferences can be measured directly in a group of patients using the Standard Gamble (SG) or Time Trade-Off (TTO) instrument (35,36), whereby patients value their own health state. Second, preferences can be measured indirectly by attaching societal preferences derived from the general population to the health states of patients using questionnaires, such as the EuroQol-5D (EQ-5D) (37), Short Form-6D (SF-6D) (38) and Health Utility Index (HUI) (39,40), allowing members of the public to determine the severity of patients' health states. Further on in this introduction, these instruments will be discussed in more detail.

Although an unambiguous definition of HRQoL is still missing, HRQoL has become increasingly important for the measurement of disease impact and the effects of treatments. The increase in the prevalence of chronic diseases, the more active role played by patients when receiving medical care, and the awareness that many treatments for chronic diseases cannot actually cure the disease, have all contributed to the increased importance of HRQoL assessment (41).

Psychometric properties of PROs in RA

In this paragraph, the psychometric properties of important PROs are evaluated. Only descriptive instruments are discussed. The preference-based instruments used in economic analysis will be discussed in the next paragraph as they play a central role in this thesis.

For the evaluation of the performance of outcome measures, it is important to assess the psychometric properties of instruments. Within the field of psychometrics, a set of criteria has been developed for assessing the performance of outcome measures. The most commonly used are practicability, reliability, validity and responsiveness (42). First, an instrument must be acceptable to the patient and to health-care providers, whereby the length of use of the instrument, the feasibility, cost and the burden to the respondent should all be taken into account. Second, an instrument must be reliable, i.e. its outcomes must reflect a certain degree of consistency both over time when the concept being measured is stable (test-retest reliability) and among the items within a scale (internal consistency). Third, an instrument must be valid, i.e. it measures the concept it is supposed to measure.



Ideally, an instrument would be tested against a 'gold standard'. In the absence of such a standard, other ways have been developed to establish validity. Important types of validity are content validity (appropriateness of items for the construct being measured) and construct validity (correlation with other measures in a hypothesized way; and the degree to which the scores are an adequate reflection of the dimensionality of the construct, i.e. structural validity) (43). Finally, an instrument must be responsive. A critical review of the literature (44) implies there are two major aspects of responsiveness: internal responsiveness and external responsiveness. Internal responsiveness refers to the ability of a measure to detect change over a pre-specified time frame. External responsiveness reflects the extent to which changes detected by a measure over a specified time frame relate to corresponding changes in a reference measure of health status.

Frequently used PROs that have been evaluated repeatedly with respect to their psychometric performance in rheumatology are the Health Assessment Questionnaire (HAQ) (45), pain-rating scales (e.g. Visual Analogue Scale, Numerical Rating Scale), the Rheumatoid Arthritis Quality of Life Scale (RAQoL) (33), the Arthritis Impact Measurement Scales 2 (AIMS2) (46) and the Medical Outcomes Study Short Form-36 (SF-36) (32).

The HAQ was initially developed to measure the functional status of RA patients, but has also been applied to other diseases. The short HAQ is the most frequently used version and is commonly referred to as 'the HAQ'. The instrument consists of the HAQ Disability Index (HAQ-DI), HAQ Pain Scale and the Patient Global Visual Analogue Scale (VAS). The HAQ has demonstrated good reliability and validity in RA (47-50). In a recent review of physical function scales in RA (51), it was concluded that the HAQ was currently one of the best available measures of physical function, with good reliability, validity and responsiveness. Nevertheless, limitations with respect to the presence of ceiling effects (best possible scores) in the HAQ have been reported (51,52). Furthermore, the instrument has been criticized for its nonlinearity and for problems among patients with understanding and correctly answering some items (52).

The pain Visual Analogue Scale (VAS) and the Numeric Rating Scale (NRS) for Pain are unidimensional measures of pain intensity that have been shown to be reliable and valid measures in rheumatic diseases and to be sensitive to change (53). A disadvantage of these pain scales is the fact that they only capture one single component of the pain experience, namely pain intensity. Other multidimensional measures of pain have been developed. Unfortunately, because of the variability in purpose, content, administration and evidence to support their psychometric properties, there is no single measure that can be recommended for all clinical and research purposes (53).

The RAQoL, AIMS2 and SF-36 are measures of HRQoL, incorporating aspects of physical, psychological and social functioning. The first two instruments are only applicable to RA, whereas the SF-36 is a generic instrument that can be used to compare the burden of different diseases. Evidence regarding reliability, validity and responsiveness (particularly to improvement) in RA has been found for all three measures (33,48,50,54-63). A limitation of the AIMS2 as well as the SF-36 is the presence of marked floor and ceiling effects in subscales of the instruments (48,54,57,64). Although studies have not reported a fully consistent pattern of floor and ceiling effects in the SF-36, the subscales physical functioning, role limitations and social functioning in particular are affected (48,54,57,64). Minimal ceiling and floor effects were reported for the RAQoL (33,48).

Health-related quality of life in health economics

Economic evaluation and the role of Quality-Adjusted Life-Years

Since the early 1960s, economic evaluation has become an essential part of policy decisions regarding the reimbursement of health-care technologies (65). Because of rising societal expenditures in health-care and the relatively limited resources currently available to us today, choices must be made with respect to reimbursement. Within the field of rheumatology, the economic burden of RA on society is high in terms of health-care costs and costs due to loss in work productivity. The expanding indications for the use of effective, but very expensive, biological therapies and the development of novel biological therapies warrant economic evaluations (66).

Economic analysis can be defined as ‘the comparative analysis of alternative courses of actions in terms of both costs and consequences’ (67). An important form of economic evaluation is cost-utility analysis (CUA). This form was developed to allow decision-makers to compare different interventions across diseases (67). In a CUA the expected costs of a health technology are compared with the expected benefits whereby the benefits of a health technology are expressed in the number of quality-adjusted life-years (QALYs) gained by that specific health technology. The QALY combines both quality of life (QoL) and life years into a single measure (68). More precisely, QALYs are calculated by multiplying the duration of time spent in a health state (duration of the health benefit) by the valuation of that health state (‘utility’). The formula for calculating QALYs is as follows: $U_1T_1 + U_2T_2 + U_3T_3$ for three health states with utilities U_1 , U_2 and U_3 with corresponding time periods (T) spent in the states (Figure 1). Thus, QoL or more specifically HRQoL is measured by a utility score that indicates the relative preference for that health state relative to death (value of 0) and perfect health (value of 1); the more desirable a health state, the higher the utility associated with it. To explain the QALY principle: a novel treatment of RA provides two QALYs. Because the time spent in the health state is multiplied by the valuation of that state, it is possible that different combinations will lead to these two QALYs: four years in a health state with a utility score of 0.5, but also eight years in a health state with a utility score of 0.25. In this way, the summation of life years is adjusted for the quality of those years (69). The results of a CUA are expressed in terms of cost per QALY gained by the adoption of one health technology compared with the adoption of another technology. Thus, economic evaluation is concerned with a choice between competing alternatives or between a novel technology and the ‘status quo’ (i.e. usual care) (67). Therefore, the incremental cost-utility ratio (ICUR) is calculated, in which the difference in the expected cost of two health technologies is divided by the difference in the expected QALYs produced by the two health technologies.



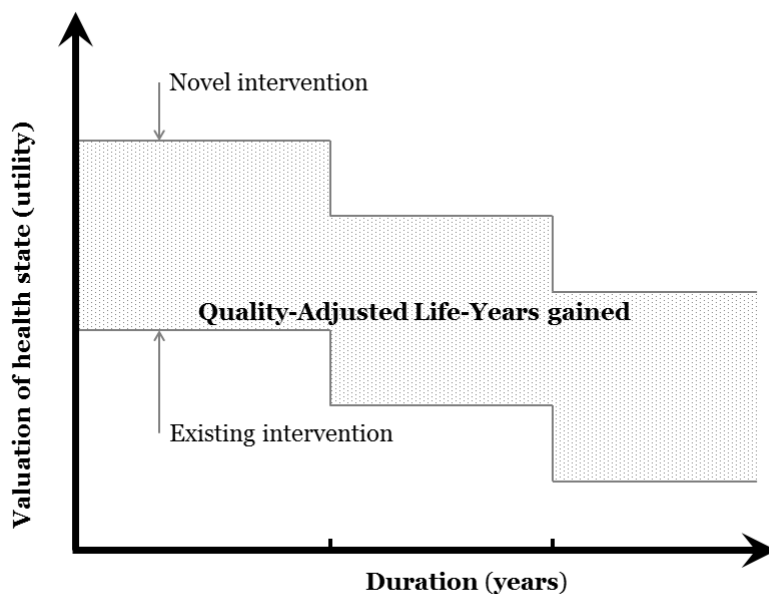


Figure 1 The QALY combines both quality of life (QoL) and life years into a single measure. QALYs are calculated by multiplying the duration of time spent in a health state (duration) by the valuation of that health state (utility)

Utility measurement

In CUA, utilities are used and they represent a particular approach to the measurement of HRQoL that is founded on the Von Neumann-Morgenstern utility theory, often referred to as ‘expected utility theory’ (EUT) (70). EUT is a theory of rational decision-making under uncertainty and incorporates a normative model by prescribing how a rational individual ‘ought’ to make decisions when confronted with uncertain outcomes. Therefore, a number of axioms were formulated to define rational behaviour (67). Although the axioms have been widely applied in varying fields for several decades, they are often criticized (71). Much evidence exists that individuals do not follow the model (72-74); that is, individuals do not behave in a rational manner. Nevertheless, EUT has remained the dominant model of rational behaviour under uncertainty for over half a century (67).

The objective of the utility measurement process is to determine the score (desirability or preference) for health states on a scale ranging between death (0) and perfect health (1) (75). Before continuing with the description of the instruments, it is relevant to remark that officially the term ‘utility’ is used only for preferences based on uncertain outcomes; that is, preferences based on EUT. To clarify this: the term is only used when there is a risk or probability involved in the preference measurement question (75). Nevertheless, it has been argued that the term ‘utility’ can be used in a broader sense as an index of the strength of preference both under conditions of certainty or uncertainty (76). In literature, the term has been frequently used to indicate preferences obtained by SG as well as TTO (77-80), thus irrespective of whether there was risk or probabilities involved in the measurement

questions. In subsequent paragraphs of this introduction and in the other chapters of this thesis, this broader meaning of the term ‘utility’ will be used to indicate valuations of health states.

Two approaches towards measuring utilities can be distinguished. First, utilities can be measured directly in patients using direct utility instruments. Second, utilities can be measured indirectly by attaching societal preferences derived from the general population to the health states of patients using indirect utility instruments, allowing members of the general public to determine the severity of the patients’ health states. Important direct utility instruments to elicit preferences for health states are the Visual Analogue Scale (VAS), Standard Gamble (SG) and Time Trade-Off (TTO) (36, 81). The VAS is a type of rating scale that asks a respondent to indicate his or her current health state on a scale with end points death (0) and perfect health (100). The SG requires the respondent to make a choice between alternative outcomes, where one of them involves uncertainty (for instance ‘risk of death’) (82) (Figure 2). The respondent has to indicate how much risk of death (expressed as percentage chance of death) he or she is prepared to accept in order to avoid being in a particular health state (e.g. own current health). The SG is based directly on the axioms of EUT. The TTO requires the respondent to choose between two certain alternative outcomes (Figure 3). The respondent has to indicate how many life years he or she is prepared to trade off in order to become perfectly healthy, i.e. to avoid being in a particular state (e.g. own current health). The theoretical underpinnings of the TTO are less clear. The instrument was developed to be an easy-to-administer instrument that generates preferences with similar empirical properties as the SG, and overcomes the problems of explaining probabilities to respondents (67,82,83). For being choice-based, the SG and TTO are preferred to rating scales based on psychometric scaling, because choosing is a natural task for people and observable and verifiable (67).

Obtaining utilities directly from patients can be costly and time-consuming. Indirect utility instruments collect information about patients’ health state rapidly through a self-administered, generic, HRQoL questionnaire. Subsequently, societal preferences derived from the general population can be attached to the health states of patients, allowing the estimation of utility scores. Examples of indirect instruments are the EQ-5D (37), SF-6D (38) and HUI2 and 3 (39,40). These instruments include a number of domains to define patients’ health states, and each domain has between three to six levels indicating the severity of problems experienced. Utility values are attached by either a preference-based scoring function based on regression techniques, in the case of the EQ-5D and SF-6D, or a multiplicative scoring formula, in the case of the HUI2 and 3. The domains of the indirect utility instruments that have been examined in this thesis (EQ-5D and SF-6D) are shown in Table 1. There is a difference in the possible range of utility scores between the SF-6D and EQ-5D, with SF-6D utility scores ranging between the conventional death (0) and perfect health (1) endpoints, whereas the EQ-5D utility scores can range between -1 and 1, allowing for states being worse than death. By way of illustration, in the study that determined the British value set of societal preferences (84), people who considered states as being worse than death in the TTO were subsequently asked to choose between dying immediately or living for a number of life years in the target state (x) followed by (10-x) years in the best EQ-5D state (i.e. the highest score on all domains). The more time they wanted to spend in the best EQ-5D state to compensate for the shorter time in the target state, the worse the target state was for them.



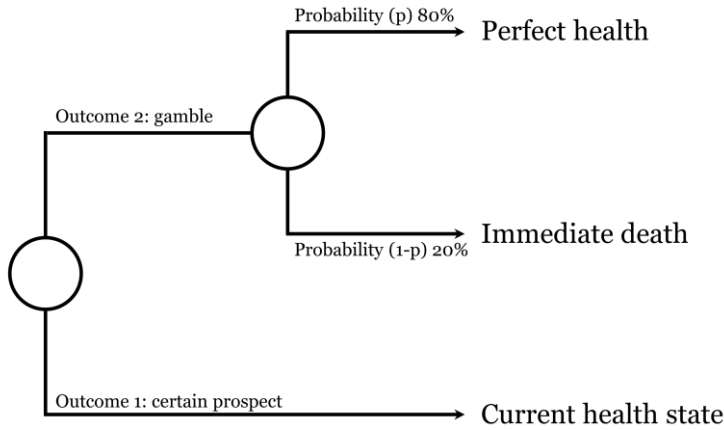


Figure 2 The Standard Gamble instrument: the first outcome is the certainty of living for a number of life years in a particular health state (e.g. own current health). The second outcome is the gamble between living for a number of life years in perfect health with probability p (e.g. 80%) or dying immediately $(1-p)$ (e.g. 20%). The probability is varied until the respondent is indifferent between the two choices (based on Drummond et al. (67))

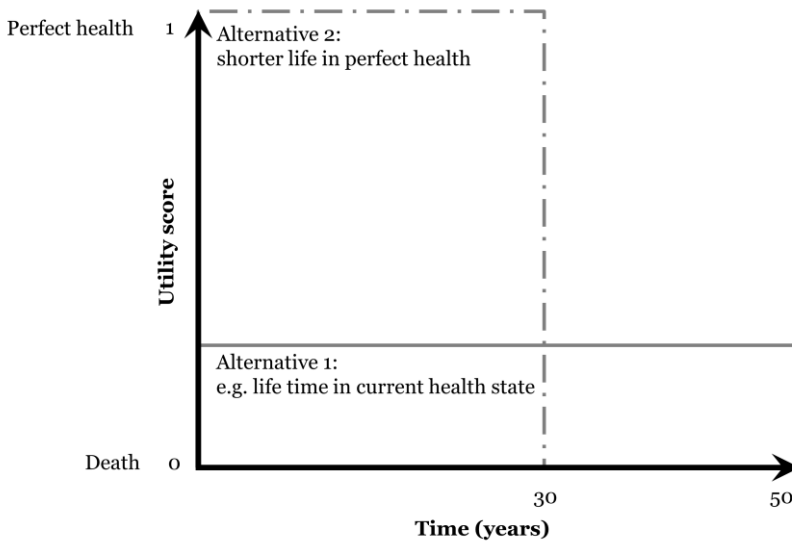


Figure 3 The Time Trade-Off instrument: alternative 1 is the current health state of the respondent for a number of life years (e.g. life expectancy of the respondent) followed by death. Alternative 2 is a shorter life in perfect health followed by death. The time in perfect health is varied until the respondent is indifferent between the two alternatives (based on Drummond et al. (67))

Table 1 Domains of instruments

EQ-5D	mobility self-care usual activities pain/discomfort anxiety/depression
SF-6D	pain mental health physical functioning social functioning role limitations vitality

Valuations of health states from a societal perspective: the norm in cost-utility analysis

Utilities can be obtained from patients, people with past experience, close relatives, health-care professionals and from members of the general public. Different disciplines have different perspectives on the issue of whose values should be used (85). The patient perspective is the preferred perspective in medical decision-making and clinical research because of the direct influence of medical interventions on patients (85). Conversely, the health economists' perspective uses a societal perspective in which members of the general public generate utilities in the context of decisions about the alternative allocation of resources. Arguments include the fact that it is not patients who bear the costs associated with health-care decisions but rather the general public (86), and that members of the general public are therefore the best judges of health states because they can at least partially act under 'a veil of ignorance', i.e. they are ignorant of their own future health states and needs, and they are blind to their own self-interest (85,87). Nevertheless, there are also strong arguments in favour of using a patient perspective in the context of resource allocation, because patients have full knowledge of, and experience with, their health condition (86-88) and are directly affected by interventions (89). In addition, even when providing members of the general public with as much information as possible, these descriptions of health states remain incomplete (87,90), leaving the general public with a 'lack of scope' (91). Furthermore, public preferences may reflect the fear and shock associated with health states they have never experienced rather than the valuation of the true health state. One explanation for this is the fact that people tend to focus more on the negative aspects associated with the disease, e.g. impairments in activities, while ignoring the positive aspects such as unaffected or even enhanced activities; a phenomenon which is defined as focusing illusion (86,90,92,93). In addition, people fail to take account of the fact that they will adapt to health changes (86,90,92,93), which can be mediated by a phenomenon called response shift. Response shift refers to the fact that people redefine their internal standards of what comprises health, change their values and priorities, and/or may redefine the concept of QoL, as a result of a diminishing health state (94). Nevertheless, the process of adaptation can also be used as an argument against the patient perspective. Adaptation will result in higher values for patients compared with those of the general public, leaving less room for improvement, i.e. less QALY gains.

So far, this debate about perspectives remains unsolved. Currently, the societal perspective is the dominant one for valuing health states in CUAs.



Critiques of the QALY model

CUA is originally founded on the optimization theory and the maximization of QALYs, which has been labeled as ‘extra welfare foundation’ (95). Practically, this means that within the distribution of health-care resources priority should be given to those patients and activities where these resources produce the largest health gains (96). To achieve this, an implicit assumption is being made, namely that all QALYs are of equal social value (‘A QALY is a QALY is a QALY’) (97). This implies that the QALY approach is blind to issues relevant for providing equity in resource allocation, such as the severity of the disease, socioeconomic status, age, the initial or end health state (98-100). Ethical theory and studies of public preferences are reluctant to prioritize only according to the size of the benefits achieved (101-104). An argument is that people with a lower potential for health are placed in a disadvantageous position (96,101,102). Evidence has shown that people give priority to severely ill patients in the case of equal QALY gains (101,105); that is, they do want to take the patients’ initial health state into account. Furthermore, people who have more treatable conditions are favoured (100), because they have an increased capacity to benefit.

Much critique of the traditional QALY model is related to the valuation of health states. The use of simplifying assumptions, such as the fact that the value of being in a health state for two years is twice the value of being in that health state for one year, has been debated frequently (106). In addition, empirical evidence has shown that people’s valuations of health states are influenced by the duration of health states and their sequence (107). The QALY model, however, assumes that preferences for health states are independent of the duration and sequences of health states. A more general critique of the QALY concerns the assumption that valuations should be preference-based. It has been argued that utilities, irrespective of whose perspective is used to obtain them, will be biased estimates (92,108). During the elicitation task people have to make choices whereby they have to predict how much they will enjoy the consequences of their choice or how much happiness the choice will give them; that is, people have to be capable of predicting how they will feel in the future (so-called ‘affective forecasting’) (92). Evidence exists concerning people’s failures in affective forecasting (92,93,108). It can be suggested that members of the general public overestimate the transitional loss in utility that results from a change in the health state of a patient because they do not take into account the adaptation of a patient to changes in his or her health state, whereas patients underestimate this loss in utility because they have adapted to changes in health (108). Therefore, according to Dolan and Kahneman (108), ‘moment-based happiness’ or ‘the quality of the hedonic experience’ that patients are having moment by moment during their lives should be captured rather than people’s predictions of how they will feel in the future in a changed health state.

Obviously, there are doubts about the validity of the QALY model because of its simplifying assumptions. The focus on efficiency concerns rather than equity concerns has been the subject of considerable discussion. Suggestions for incorporating these concerns into the QALY approach have been proposed (102,109,110).

Limitations of direct utility instruments

A key assumption of the QALY model, already briefly discussed in the previous paragraph, is that utilities are not independent of the duration of the health state. Here, we will discuss more elaborately the fact that the utility instruments, SG and TTO, both violate this assumption. For the SG it means that utilities are independent of the choice of the duration in the SG task. For example, someone who is willing to take a 70% risk to become perfectly healthy for a duration of five years, is also willing to take a risk of 70% to become perfectly health for a duration of e.g. 10, 12 or 20 years. For the SG, this

phenomenon is called 'utility independence'. In their review (107), Tsuchiya & Dolan concluded that utility dependence is violated, although there is no clear pattern in the responses of people who violate utility independence. For the TTO, this phenomenon is called 'constant proportional time trade-off'. It means that the same proportional amount of time is traded off independently of the absolute duration in the TTO task (107). For example, someone who is willing to trade five years out of 10 years is also willing to trade 10 years out of 20 years. Some authors showed that people satisfy a constant proportional time trade-off, at least at the aggregate level (107,111), whereas other authors showed that people do not satisfy a constant proportional time trade-off (112,113). In addition, it has been shown for longer durations that a phenomenon called 'maximal endurable time' can be observed, especially for severe health states (113); that is, after a certain threshold duration people do not tolerate this health state anymore and prefer to die. This will result in lower utility scores.

Research has shown that the SG and TTO produce systematically different utility scores (114,115), whereby SG utilities have been found to be higher than TTO utilities. Empirical evidence exists which shows that TTO utilities better reflect individual preferences than SG utilities (76,115). As previously discussed in the paragraph about utility measurement, individuals have been found to violate EUT, i.e. they do not behave in a rational manner. An implication is that SG utilities are biased upwards. Three reasons why people deviate from EU are 'loss aversion', 'probability weighting' and 'scale compatibility' (116). A recent systematic review on health utility bias has suggested that the biases in the TTO may be cancelled out (117), which might explain why TTO utilities are more consistent with individual preferences than the SG. Nevertheless, a net upward bias in TTO utilities has also been suggested (118).

Variations in utility scores have been found between the SG and TTO, but also within these instruments. In a review of empirical studies applying the TTO in diagnostic groups, Arnesen and Trommald (119) showed that the utility scores bore no relation to the severity of the disease (e.g. the median utility score for metastatic, symptomatic cancer was 1.00) and that there were great variations in the utility scores of the same diagnostic group. Subsequently, these authors performed another review to explore two explanations for their findings, namely the variation within the TTO methodology and the unrepresentativeness of samples for their diagnostic groups (120). Large differences within the TTO methodology were shown between studies, e.g. with respect to data collection strategies, formulations and framing of the TTO question, and the time frame chosen. Furthermore, the different diagnostic groups differed with respect to demographic characteristics. It was shown that these characteristics influenced the results and that the results were robust, implying a decrease of utility scores with age and somewhat lower scores for women than for men. Nevertheless, it was also shown that median utility scores varied substantially between studies with similar characteristics (large studies among the general population) and that there were large differences in the proportion of people who did not want to trade any life years at all in order to become perfectly healthy. These people are called 'zero-traders' and their proportion ranged from 1.7 to 58%. These results indicate that different TTO approaches cannot be compared directly, and they have implications for the use of the TTO in QALYs, whereby different diseases are compared. Studies have also compared different data collection strategies of the SG (76,121) and different search procedures to find the point at which people are indifferent between the two outcomes (121,122). The studies showed that different data collection strategies and search procedures yielded different results.



Limitations of indirect utility instruments

Research has shown that utilities for health states can vary considerably, depending on the instrument used. Kopec et al. (123) compared four indirect utility instruments (the Quality of Well-Being Scale, HUI, EQ-5D and SF-6D) with respect to three health states and found considerable differences between these instruments. In general, the SF-6D utility scores were higher than those of the other instruments. A study by Brazier et al. (124) compared the EQ-5D and SF-6D across seven patient groups, and showed by plotting EQ-5D scores against SF-6D scores that negative EQ-5D utility scores could be related to SF-6D utility scores up to 0.75. In addition, it was found that the majority of patients reporting full health on the EQ-5D did not report full health on the SF-6D. A wide range in SF-6D utility scores was observed. Furthermore, a significant proportion of patients reporting the worst health state on the SF-6D were shown to score across a large range of EQ-5D values. These findings imply that indirect instruments cannot be used interchangeably. A possible explanation for the finding that different indirect utility instruments generate different utility scores for similar health states is the fact that different direct utility instruments and different procedures are used: SF-6D utilities were obtained by using a SG with a chained procedure whereby the health state of interest is not directly compared with death but indirectly with the aid of an intermediate anchor health state, whereas the original UK EQ-5D utilities were obtained by using a conventional TTO protocol.

With respect to the two indirect instruments evaluated in this thesis, it has been observed that the distribution of SF-6D utility scores is more narrow compared with the EQ-5D, whereby the scores are clustered around the milder end of the scale (124). The EQ-5D has a bimodal distribution in RA (125) and in other patient groups (124) due to the scoring algorithm used, resulting in a pattern of clustered scores. The descriptive ability of the EQ-5D has been suggested to be limited because of the three-level scaling of the domains (125). Consequently, patients may underestimate or overestimate their health in a certain domain (126). Evidence has been found for the existence of marked floor effects in the SF-6D utility scores of different patient groups (124). Although no marked floor effects were found in the SF-6D utility scores of RA patients (127,128), the fact that the SF-6D has a high floor suggests that the instrument is not good at discriminating among those people who are in poor health (125). On the other hand, the observation of marked ceiling effects in the EQ-5D (123) suggests that this instrument has difficulty with discriminating among those people who are in very good health (123-125,129). The fact that these instruments provide different indices over the range of ill health can affect the QALY gains estimated. It is, therefore, important to examine whether limitations in these instruments can be removed by revising them.

Psychometric properties of direct and indirect utility instruments

It has been stated that the TTO is considered to be the practical gold standard because of its better feasibility, higher discriminative power and better face validity compared with the SG (118). Nevertheless, two earlier reviews of health-state valuation techniques suggested that there was very limited evidence concerning the empirical validity of both instruments and that, although evidence slightly advanced the TTO over the SG, this was insufficient to conclude that the TTO was more valid than the SG (82,130). To date, empirical evidence that compares both instruments with respect to their psychometric properties is scarce. One study in patients with IBS showed, on the contrary, that the SG had better construct validity than the TTO (131).

When considering the psychometric properties of both instruments specifically in different patients groups, the results are generally disappointing with respect to construct validity, discriminative ability and responsiveness. Both the SG and TTO are feasible direct utility instruments (78,79,82,130),

although both instruments have been suggested to be complex. In varying patient groups, such as patients with renal disease who are receiving dialysis (132), Crohn's disease (77), chronic stable angina (133) and rheumatoid arthritis (78,79), the TTO was found to be a reliable instrument. The SG also demonstrated reliability (77,134). The results regarding construct validity of the TTO and SG were generally poor, with both instruments showing poor to moderate correlations with disease- or health-related outcome measures (78-80,132,134-139). Poor results were also found regarding their responsiveness (77,138-142). Evidence for and against the discriminative ability of the TTO was also found (78,79,133,135). Contradictory findings were found for pain and disease activity scores in patients with RA (78,79) and for functional status scores in patients with cardiovascular disease (133,135). A poor to moderate discriminative ability was found for the SG in patients with amyotrophic lateral sclerosis (ALS) (143).

When evaluating the instrument considered to be the practical 'gold standard', i.e. TTO, specifically in RA, it can be observed that only a few studies have examined the psychometric properties of the TTO. The few studies that examined construct validity showed poor correlations of the TTO with measures of HRQoL, functional status, disease activity and pain (78-80). Evidence for and against its discriminative ability was found (78-80). It has to be remarked that the results above comprise different variants of the TTO (e.g. they differ in data collection strategy and/or time frame). Therefore, it is difficult to interpret the results. In addition, it is formally incorrect to speak of 'the TTO', because the TTO methodology is not standardized. Nevertheless, the majority of the results indicate that the psychometric properties are poor. A possible explanation is the presence of zero-traders who give the impression that their health is perfect because they do not want to trade any life years for perfect health. Possibly, patients with chronic diseases, such as RA, do not want to trade any life years at all because dying earlier is not perceived as a realistic consequence of their disease.

The indirect utility instruments EQ-5D, SF-6D and HUI have shown themselves to be feasible as measured by completion time and/or missing data (58,125,144,145), although a recent review in rheumatic diseases (63) showed a high variability in the reports of missing data (from 1 to as much as 40%). Two reviews (63,146) demonstrated a moderate to high test-retest reliability for the EQ-5D. Evidence for its construct validity and responsiveness has been found (63,146), although Gignac et al. (63) showed conflicting poor or good results regarding its ability to detect change in different rheumatic diseases. Another review of general health and HRQoL measures in rheumatic diseases found a good as well as a moderate reliability for the SF-6D (58). In addition, they demonstrated a good convergent validity of moderate correlations with other HRQoL scales and a good discriminative validity for the SF-6D. With respect to its ability to detect change, mixed results were found in rheumatic diseases. Some found that the SF-6D performed well, others that it performed poorly. The same review showed diverse results with respect to the reliability, construct validity and responsiveness of the HUI3, whereby positive results as well as negative results were found.

When only considering RA, there is extensive evidence that the EQ-5D and HUI3 are feasible, valid and responsive (125). The HUI3 demonstrated a higher reliability (ICC 0.72-0.89) than the EQ-5D (ICC 0.46-0.66) in studies using the same patients. The SF-6D was also found to be feasible and reliable. Although evidence has been found to support its construct validity, Harrison et al. (125) concluded that the SF-6D needs further validation as it is a relatively new measure. Although there is evidence for the responsiveness of all three measures, there is conflicting evidence about which instrument is most responsive to improvement. Irrespective of which aspect of responsiveness was measured (internal or external), most studies demonstrated that the SF-6D was most responsive to improvement (55,129,147), whereas one study demonstrated that the EQ-5D, compared with the SF-6D, was most responsive to improvement (127). Therefore, further research into the 'responsiveness to improvement' of instruments is required. Because of the crude three-level response system of the EQ-5D and ceiling and floor effects



in the EQ-5D and SF-6D, respectively, these instruments might not adequately reflect the range and distribution of health states within an illness, which compromises their validity. As mentioned in the previous paragraph, the EQ-5D might not be appropriate in patients with milder disease levels, whereas the SF-6D might not be appropriate in patients with worse disease levels.

Opportunity for improvements in the measurement of HRQoL

Previously, we have argued that there is much room for improvement in the measurement of HRQoL. More research is needed with respect to the empirical validity of these instruments in RA patients. There is insufficient research with respect to the performance of the TTO and SG in RA patients. In addition, as concluded by Harrison et al. in their review (125), further research is required into the responsiveness of both the SF-6D and EQ-5D to improvement in RA patients. Within the debate about which perspective should be used in health economic analysis, arguments for both the patient and the societal perspective are reasonable. We believe that it is essential to obtain patients' valuations, because patients have a unique and full knowledge of, and experience with, their health condition and are directly affected by interventions. Nevertheless, extensive research is required to improve the measurement of HRQoL that represents the patient perspective.



REFERENCES


1. Klippel JH. *Primer on the Rheumatic Diseases*, 12th edn. Atlanta: Arthritis Foundation; 2001.
2. Gabriel SE, Michaud K. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther* 2009;11:229.
3. El-Gabalawy H, Guenther LC, Bernstein CN. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. *J Rheumatol* 2010;37(Suppl. 85):S2-S10.
4. Scott DL, Steer S. The course of established rheumatoid arthritis. *Best Practice and Research. Clin Rheumatol* 2007;21:943-67.
5. Boonen A, Severens JL. The burden of illness of rheumatoid arthritis. *Clin Rheumatol* 2011;30(Suppl. 1):S3-S8.
6. Uhlig T, Loge JH, Kristiansen IS, Kvien TK. Quantification of reduced health-related quality of life in patients with rheumatoid arthritis compared to the general population. *J Rheumatol* 2007;34:1241-7.
7. Pollard L, Choy EH, Scott DL. The consequences of rheumatoid arthritis: quality of life measures in the individual patient. *Clin Exp Rheumatol* 2005;23(Suppl. 39):S43-S52.
8. Welsing PMJ, van Gestel AM, Swinkels HL, Kiemeny LALM, van Riel PLCM. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis Rheum* 2001;44:2009-17.
9. Hewlett S, Cockshott Z, Byron M, Kitchen K, Tipler S, Pope D, Hehir M. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis Care Res* 2005;53:697-702.
10. Repping-Wuuts H, Uitterhoeve R, van Riel P, van Achterberg T. Fatigue as experienced by patients with rheumatoid arthritis (RA): a qualitative study. *Int J Nurs Stud* 2008;45:995-1002.
11. Nikolaus S, Bode C, Taal E, van de Laar MAFJ. New insights into the experience of fatigue among patients with rheumatoid arthritis: a qualitative study. *Ann Rheum Dis* 2010;69:895-7.
12. Pouchot J, Le Parc JM, Queffelec L, Sichère P, Flinois A. Perceptions in 7700 patients with rheumatoid arthritis compared to their families and physicians. *Joint Bone Spine* 2007;74:622-6.
13. Sheehy C, Murphy E, Barry M. Depression in rheumatoid arthritis - underscoring the problem. *Rheumatology* 2006;45:1325-7.
14. VanDyke MM, Parker JC, Smarr KL, Hewett JE, Johnson GE, Slaughter JR, et al. Anxiety in rheumatoid arthritis. *Arthritis Care Res* 2004;51:408-12.
15. Lütze U, Archenholtz B. The impact of arthritis on daily life with the patient perspective in focus. *Scand J Caring Sci* 2007;21:64-70.
16. Lempp H, Scott D, Kingsley G. The personal impact of rheumatoid arthritis on patients' identity: a qualitative study. *Chronic Illn* 2006;2:109-20.
17. Bath J, Hooper J, Giles M, Steel D, Reed E, Woodland J. Patient perceptions of rheumatoid arthritis. *Nurs stand* 1999;14:35-8.
18. van Lankveld W, Naring G, van der Staak C, van 't Pad Bosch P, van de Putte L. Stress caused by rheumatoid arthritis: relation among subjective stressors of the disease, disease status, and well-being. *J Behav Med* 1993;16:309-21.
19. Taal E, Rasker JJ, Seydel ER, Wiegman O. Health status, adherence with health recommendations, self-efficacy and social support in patients with rheumatoid arthritis. *Patient Educ Couns* 1993;20:63-76.
20. Westhoff G, Listing J, Zink A. Loss of physical independence in rheumatoid arthritis: interview data from a representative sample of patients in rheumatologic care. *Arthritis Care Res* 2000;13:11-22.
21. Geuskens GA, Burdorf A, Hazes JMW. Consequences of rheumatoid arthritis for performance of social roles - a literature review. *J Rheumatol* 2007;34:1248-60.
22. Zhang W, Anis AH. The economic burden of rheumatoid arthritis: beyond health care costs. *Clin Rheumatol* 2011;30(Suppl. 1):S25-S32.
23. Villeneuve E, Emery P. Rheumatoid arthritis: what has changed? *Skeletal Radiol* 2009;38:109-12.
24. Smolen JS, Aletaha D. What should be our treatment goal in rheumatoid arthritis today? *Clin Exp Rheumatol* 2006;24(Suppl. 43):S7-S13.
25. Vermeer M, Kuper HH, Hoekstra M, Haagsma CJ, Posthumus MD, Brus HLM, et al. Implementation of a treat to target strategy in very early rheumatoid arthritis: results of the DREAM remission induction cohort. *Arthritis Rheum* 2011;63:2865-72.



26. Verstappen SMM, Symmons DPM. What is the outcome of RA in 2011 and can we predict it? *Best Pract Res Clin Rheumatol* 2011;25:485-96.
27. Hewlett S, Carr M, Ryan S, Kirwan J, Richards P, Carr A, et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care* 2005;3:131-42.
28. World Health Organization. Constitution of the World Health Organization. October 2006. Available at: http://www.who.int/governance/eb/who_constitution_en.pdf. Accessed February 22, 2012.
29. Kalyoncu U, Dougados M, Daurès JP, Gossec L. Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. *Ann Rheum Dis* 2009;68:183-90.
30. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000;9:887-900.
31. van Heck G. Verwarring rondom Kwaliteit van Leven: beter ten halve gekeerd, dan ten hele gedwaald. *Psychologie & Gezondheid* 2008; 36:72-78. In Dutch.
32. Ware Jr JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Med Care* 1992;30:473-83.
33. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. *Rheumatology* 1997;36:878-83.
34. O'Boyle C, McGee H, Hickey A, Joyce C, Browne J, O'Malley K, et al. The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). Administration Manual. Dublin: Royal College of Surgeons in Ireland; 1993.
35. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7:118-33.
36. Torrance GW: Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1-30.
37. Brooks R, De Charro F. EuroQol: The current state of play. *Health Policy* 1996;37:53-72.
38. 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.
39. 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.
40. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001;33:375-84.
41. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
42. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ* 1999;8:41-51.
43. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19:539-49.
44. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459-68.
45. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
46. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2: The content and properties of a revised and expanded arthritis impact measurement scales health status questionnaire. *Arthritis Rheum* 1992;35:1-10.
47. Bruce B, Fries J. The stanford health assessment questionnaire: dimensions and practical applications. *Health Qual Life Outcomes* 2003;1:20.
48. Linde L, Sørensen J, Østergaard M, Hørslev-Petersen K, Hetland ML. Health-related quality of life: validity reliability, and responsiveness of SF-36, EQ-15D, EQ-5D, RAQoL, and HAQ in patients with rheumatoid arthritis. *J Rheumatol* 2008;35:1528-37.
49. White DK, Wilson JC, Keysor JJ. Measures of adult general functional status: SF-36 Physical Functioning Subscale (PF-10), Health Assessment Questionnaire (HAQ), Modified Health Assessment Questionnaire (MHAQ), Katz Index of Independence in Activities of Daily Living, Functional Independence Measure (FIM),


- and Osteoarthritis- Function-Computer Adaptive Test (OA-Function-CAT). *Arthritis Care Res* 2011;63(Suppl. 11):S297-S307.
50. Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). *Arthritis Care Res* 2011;63(Suppl. 11):S4-S13.
 51. Oude Voshaar MAH, ten Klooster PM, Taal E, van de Laar MAFJ. Measurement properties of physical function scales validated for use in patients with rheumatoid arthritis: a systematic review of the literature. *Health Qual Life Outcomes* 2011;9:99.
 52. Wolfe F, Michaud K, Pincus T. Development and validation of the Health Assessment Questionnaire II: a revised version of the Health Assessment Questionnaire. *Arthritis Rheum* 2004;50:3296-305.
 53. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res* 2011;63(Suppl. 11):S240-S252.
 54. Salaffi F, Stancati A, Carotti M. Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. *Clin Rheumatol* 2002;21:478-87.
 55. Marra CA, Rashidi AA, Guh D, Kopec JA, Abrahamowicz M, Esdaile JM, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Qual Life Res* 2005;14:1333-44.
 56. Kvien TK, Kaasa S, Smedstad LM. Performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. II. A comparison of the SF-36 with disease-specific measures. *J Clin Epidemiol* 1998;51:1077-86.
 57. Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). *Br J Rheumatol* 1998;37:425-36.
 58. Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQ O L). *Arthritis Care Res* 2011;63(Suppl. 11):S383-S412.
 59. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. *Br J Rheumatol* 1996;35:755-60.
 60. Taal E, Rasker JJ, Riemsma RP. Psychometric properties of a Dutch short form of the Arthritis Impact Measurement Scales 2 (Dutch-AIMS2-SF). *Rheumatology* 2003;42:427-34.
 61. Taal E, Rasker JJ, Riemsma RP: Sensitivity to change of AIMS2 and AIMS2-SF components in comparison to M-HAQ and VAS-pain. *Ann Rheum Dis* 2004;63:1655-8.
 62. Veehof MM, ten Klooster PM, Taal E, van Riel PLCM, van de Laar MAFJ. Comparison of internal and external responsiveness of the generic medical outcome study short form-36 (SF-36) with disease-specific measures in rheumatoid arthritis. *J Rheumatol* 2008;35:610-7.
 63. Gignac MAM, Cao X, McAlpine J, Badley EM. Measures of disability: Arthritis Impact Measurement Scales 2 (AIMS2), Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF), the Organization for Economic Cooperation and Development (OECD) Long-Term Disability (LTD) Questionnaire, EQ-5D, World Health Organization Disability Assessment Schedule II (WHODASII), Late-Life Function and. *Arthritis Care Res* 2011;63(Suppl. 11):S308-S324.
 64. Aaronson NK, Muller M, Cohen PDA, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;51:1055-68.
 65. Meltzer MI. Introduction to health economics for physicians. *Lancet* 2001;358:993-8.
 66. Tan MCY, Regier DA, Esdaile JM, Lynd LD, Anis AH, Marra CA. Health economic evaluation: a primer for the practicing rheumatologist. *Arthritis Care Res* 2006;55:648-56.



- 
67. Drummond MF, Sculpher MJ, Torrance GW. *Methods for the economic evaluation of health care programmes*. New York: Oxford University Press; 2005.
 68. Smith MD, Drummond M, Brixner D. Moving the QALY forward: rationale for change. *Value Health* 2009;12:S1-S4.
 69. Gafni A: The Standard Gamble method: what is being measured and how it is interpreted. *Health Serv Res* 1994;29:207-23.
 70. von Neumann J, Morgenstern, O. *Theory of games and economic behavior*. Princeton: Princeton University Press; 1944.
 71. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertain* 1992;5:297-323.
 72. Loomes G. Evidence of a new violation of the independence axiom. *J Risk Uncertain* 1991;4:91-108.
 73. Luce RD. Where does subjective expected utility fail descriptively? *J Risk Uncertain* 1992;5:5-27.
 74. Oliver A. A quantitative and qualitative test of the Allais paradox using health outcomes. *J Econ Psychol* 2003;24:35-48.
 75. Bennett K, Torrance G. Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven Publishers;1996.253-265.
 76. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;15:209-31.
 77. Gregor JC, McDonald JWD, Klar N, Wall R, Atkinson K, Lamba B, et al. An evaluation of utility measurement in Crohn's disease. *Inflamm Bowel Dis* 1997;3:265-76.
 78. 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.
 79. Tjihuis 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.
 80. Witney AG, Treharne GJ, Tavakoli M, Lyons AC, Vincent K, Scott DL, et al. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology* 2006;45:975-81.
 81. Kaplan RM, Feeny D, Revicki D. Methods for assessing relative importance in preference based outcome measures. *Qual Life Res* 1993;2:467-75.
 82. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;3:1-164.
 83. Buckingham K, Devlin N. A theoretical framework for TTO valuations of health. *Health Econ* 2006;15:1149-54.
 84. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095-1108.
 85. De Wit GA, Busschbach JJV, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* 2000;9:109-26.
 86. Stamuli E. Health outcomes in economic evaluation: who should value health? *Br Med Bull* 2011;97:197-210.
 87. Menzel P: How should what economists call "Social Values" be measured? *J Ethics* 1999;3:249-73.
 88. Menzel P, Dolan P, Richardson J, Olsen JA. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Soc Sci Med* 2002;55:2149-58.
 89. Happich M, Von Lengerke T: Valuing the health state 'tinnitus': differences between patients and the general public. *Hear Res* 2005;207:50-8.
 90. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual Life Res* 2003;12:599-607.
 91. Insinga RP, Fryback DG. Understanding differences between self-ratings and population ratings for health in the EuroQOL. *Qual Life Res* 2003;12:611-9.
 92. Kahneman D, Sugden R. Experienced utility as a standard of policy evaluation. *Environ Resour Econ* 2005;32:161-81.
 93. Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychol Sci* 1998;9:340-6.
 94. Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychol Health* 2004;19:51-69.

95. Torrance GW. Preferences for health outcomes and cost-utility analysis. *Am J Manag Care* 1997;3(Suppl. 1):S8-S20.
96. Nord E, Enge AU, Gundersen V. QALYS: Is the value of treatment proportional to the size of the health gain? *Health Econ* 2010;19:596-607.
97. Weinstein MC. A QALY is a QALY is a QALY - Or is it? *J Health Econ* 1988;7:289-90.
98. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull* 2010;96:5-21.
99. Schwappach DLB. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expect* 2002;5:210-22.
100. Nord E, Daniels N, Kamlet M. QALYS: some challenges. *Value Health* 2009;12(Suppl. 1):S10-S15.
101. Ubel PA, Richardson J, Prades JLP. Life-saving treatments and disabilities: are all QALYS created equal? *Int J Technol Assess Health Care* 1999;15:738-48.
102. Nord E, Pinto JL, Richardson J, Menzel P, Ubel P. Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Econ* 1999;8:25-39.
103. Abellan-Perpiñan JM, Pinto-Prades JL. Health state after treatment: a reason for discrimination? *Health Econ* 1999;8:701-7.
104. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ* 2005;14:197-208.
105. 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.
106. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value Health* 2009;12(Suppl. 1):S18-S26.
107. Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decis Making* 2005;25:460-7.
108. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008;118:215-34.
109. Bleichrodt H, Diecidue E, Quiggin J. Equity weights in the allocation of health care: the rank-dependent QALY model. *J Health Econ* 2004;23:157-71.
110. Dolan P. The measurement of individual utility and social welfare. *J Health Econ* 1998;17:39-52.
111. Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYS: constant proportional time trade-off versus the proportional heuristic. *J Health Econ* 2003;22:445-58.
112. Buckingham JK, Birdsall J, Douglas JG. Comparing three versions of the time tradeoff: time for a change? *Med Decis Making* 1996;16:335-47.
113. Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Habbema JDF, De Haes JCJM. The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *J Clin Epidemiol* 1995;48:1207-14.
114. Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Stoter G, De Haes JCJM. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making* 1994;14:82-90.
115. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYS. *J Health Econ* 1997;16:155-75.
116. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002;11:447-56.
117. Doctor JN, Bleichrodt H, Lin HJ. Health utility bias: a systematic review and meta-analytic evaluation. *Med Decis Making* 2010;30:58-67.
118. Van Osch SMC, Wakker PP, Van Den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making* 2004;24:511-7.
119. Arnesen T, Trommald M. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *J Health Serv Res Policy* 2004;9:43-50.
120. Arnesen T, Trommald M. Are QALYS based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Econ* 2005;14:39-53.
121. Hammerschmidt T, Zeitler HP, Gulich M, Leidl R. A comparison of different strategies to collect standard gamble utilities. *Med Decis Making* 2004;24:493.



- 
122. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitation. *Med Decis Making* 1998;18:76-83.
 123. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* 2003;56:317-25.
 124. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873-84.
 125. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DPM. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *J Rheumatol* 2008;35:592-602.
 126. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Farragher TM, Verstappen SMM, et al. Why do patients with inflammatory arthritis often score states "worse than death" on the EQ-5D? An investigation of the EQ-5D classification system. *Value Health* 2009;12:1026-34.
 127. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care* 2003;41:791-801.
 128. Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care* 2004;42:1125-31.
 129. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Verstappen SMM, Watson K, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Qual Life Res* 2009;18:1195-1205.
 130. Green C, Brazier J, Deverill M. Valuing health-related quality of life: A review of health state valuation techniques. *Pharmacoeconomics* 2000;17:151-65.
 131. Puhan MA, Schünemann HJ, Wong E, Griffith L, Guyatt GH. The standard gamble showed better construct validity than the time trade-off. *J Clin Epidemiol* 2007;60:1029-33.
 132. Maor Y, King M, Olmer L, Mozes B. A comparison of three measures: the time trade-off technique, global health-related quality of life and the SF-36 in dialysis patients. *J Clin Epidemiol* 2001;54:565-70.
 133. Nease Jr RF, Kneeland T, O'Connor GT, Sumner W, Lumpkins C, Shaw L, et al. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. *JAMA* 1995;273:1185-90.
 134. Cranney A, Coyle D, Pham BA, Tetroe J, Wells G, Jolly E, Tugwell P. The psychometric properties of patient preferences in osteoporosis. *J Rheumatol* 2001;28:132-37.
 135. Lalonde L, Clarke AE, Joseph L, Mackenzie T, Grover SA. Comparing the psychometric properties of preference-based and nonpreference-based health-related quality of life in coronary heart disease. *Qual Life Res* 1999;8:399-409.
 136. Khanna D, Ahmed M, Furst DE, Ginsburg SS, Park GS, Hornung R, et al. Health values of patients with systemic sclerosis. *Arthritis Rheum* 2007;57:86-93.
 137. Goossens MEJB, Vlaeyen JWS, Ruten- van Mólken MPMH, van der Linden SMJP. Patient utilities in chronic musculoskeletal pain: how useful is the standard gamble method? *Pain* 1999;80:365-75.
 138. Juniper EF, Norman GR, Cox FM, Roberts JN. Comparison of the standard gamble, rating scale, AQLQ and SF-36 for measuring quality of life in asthma. *Eur Respir J* 2001;18:38-44.
 139. Juniper EF, Thompson AK, Roberts JN. Can the standard gamble and rating scale be used to measure quality of life in rhinoconjunctivitis? Comparison with the RQLQ and SF-36. *Allergy* 2002;57:201-6.
 140. Feeny D, Wu L, Eng K. Comparing Short Form 6D, Standard Gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. *Qual Life Res* 2004;13:1659-70.
 141. Puhan MA, Guyatt GH, Goldstein R, Mador J, McKim D, Stahl E, et al. Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respir Med* 2007;101:308-16.
 142. Brozek JL, Guyatt GH, Heels-Ansdell D, Degl'Innocenti A, Armstrong D, Fallone CA, et al. Specific HRQL instruments and symptom scores were more responsive than preference-based generic instruments in patients with GERD. *J Clin Epidemiol* 2009;62:102-10.
 143. Kiebert GM, Green C, Murphy C, Mitchell JD, O'Brien M, Burrell A, et al. Patients' health-related quality of life and utilities associated with different stages of amyotrophic lateral sclerosis. *J Neurol Sci* 2001;191:87-93.
 144. Fisk JD, Brown MG, Sketris IS, Metz LM, Murray TJ, Stadnyk KJ. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *J Neurol Neurosurg Psychiatry* 2005;76:58-63.

145. Nilsson E, Wenemark M, Bendtsen P, Kristenson M. Respondent satisfaction regarding SF-36 and EQ-5D, and patients' perspectives concerning health outcome assessment within routine health care. *Qual Life Res* 2007;16:1647-54.
146. Carr A. Adult Measures of Quality of Life: The Arthritis Impact Measurement Scales (AIMS/AIMS2), Disease Repercussion Profile (DRP), EuroQoL, Nottingham Health Profile (NHP), Patient Generated Index (PGI), Quality of Well-Being Scale (QWB), RAQoL, Short Form-36 (SF-36), Sickness Impact Profile (SIP), SIP-RA, and World Health Organization's Quality of Life Instruments (WHOQoL, WHOQoL-100, WHOQoL-Bref). *Arthritis Care Res* 2003;49(Suppl. 5):S113-S133.
147. Russell AS, Conner-Spady B, Mintz A. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *J Rheumatol* 2003;30:941-7.



Chapter



Comparative responsiveness of the EuroQol-5D and Short Form 6D to improvement in patients with rheumatoid arthritis treated with Tumor Necrosis Factor Blockers: results of the Dutch Rheumatoid Arthritis Monitoring Registry

L. Buitinga
L.M.A. Braakman-Jansen
E. Taal
W. Kievit
H. Visser
P.L.C.M. van Riel
M.A.F.J. van de Laar

ABSTRACT

Objective For cost-utility analyses of health technologies, utilities are commonly measured with the EuroQol-5D (EQ-5D) or the Short Form 6D (SF-6D). Although most studies in rheumatoid arthritis (RA) found the SF-6D to be more responsive than the EQ-5D, evidence is not convincing. The aim of this study was to compare the responsiveness of the EQ-5D and SF-6D to improvement in RA patients treated with tumor necrosis factor (TNF) blockers.

Methods Data from 278 RA patients included in the Dutch Rheumatoid Arthritis Monitoring registry were used. Internal responsiveness over 1 year was evaluated by using standardized response means (SRMs). External responsiveness was evaluated by using receiver operating characteristic curves based on perceived health change (self-reported health transition item Short Form 36) and change in disease activity (European League Against Rheumatism response criteria based on the Disease Activity Score in 28 joints).

Results The scores of the EQ-5D and SF-6D changed moderately over 1 year (SRMs 0.50 and 0.67, respectively). The SF-6D was significantly more responsive to treatment than the EQ-5D. The EQ-5D and SF-6D were moderately able to correctly classify patients according to health transition (areas under the curve [AUCs] 0.67 and 0.72, respectively) and change in disease activity (AUCs 0.71 and 0.65, respectively).

Conclusion The EQ-5D and SF-6D were only moderately responsive to improvement in RA patients treated with TNF blockers. Overall, the SF-6D was more responsive than the EQ-5D.



INTRODUCTION

Clinical trials in patients with rheumatoid arthritis (RA) have demonstrated that relatively new treatments, such as tumor necrosis factor (TNF) blockers, are able to provide benefits with respect to clinical, functional (1–3), radiographic (1,2), and health-related quality of life outcomes (2). Clinical and functional benefits of TNF blockers have also been shown in routine clinical practice, although the relative improvement due to TNF blockers was often smaller due to differences in treatment criteria (4,5). The ongoing innovations in medicine and the costs related to novel interventions warrant economic evaluations in which the expected costs and health benefits of health technologies are compared. For economic evaluations, it is recommended to express the benefits obtained from health technologies in quality-adjusted life years (QALYs) (6). The QALY combines both life years and quality of life into a single measure (7). QALYs are assessed by multiplying the time spent in a health state by the preference for a health state (“utility”) relative to perfect health (1) and death (0). Because the QALY is applicable to “all individuals and all diseases,” it can be used to compare interventions across diseases (8). Analyses using this approach are called cost-utility analyses (CUAs). In CUAs, health technologies are compared using the incremental cost-effectiveness ratio (ICER), also known as the cost per QALY. The ICER is calculated by dividing the difference in the expected cost of 2 health technologies by the difference in the expected QALYs produced by the 2 health technologies. Based on the ICER, health care resource allocation decisions can be made.

Utilities can be obtained in 2 ways. First, patients can value their own health state directly, allowing the use of instruments such as the visual analog rating scale, standard gamble, and time trade-off (7). These methods use the length of a line, probabilities, and time, respectively, to indicate the magnitude of the preference for the health state (9). Second, health states of patients can be mapped to societal preferences using preference-based measures of health-related quality of life (HRQOL) such as the EuroQol-5D (EQ-5D) (10), Short Form 6D (SF-6D) (11), and Health Utility Index (HUI) (12). Patients have to complete a descriptive questionnaire to define their health state. Subsequently, corresponding values elicited from a general population sample are attached to patients’ health states in order to obtain societal preferences for these health states. The use of societal preferences is recommended for use in CUAs (13) because decisions about reimbursement of new technologies in health care are made from a societal perspective.

In order to evaluate the effectiveness of a health technology, e.g., of a novel RA treatment, utility measures should be reliable, valid, and responsive. Responsiveness is the ability of an instrument to detect small but important clinical changes (14). Therefore, if a novel RA treatment causes changes in patients’ health, preference-based measures of HRQOL should be able to detect these changes. Research has shown inconsistencies with respect to the psychometric performance of commonly used preference-based measures like the EQ-5D and SF-6D (6). The review by Harrison et al (6) showed that there was no conclusive evidence to favor any one measure (EQ-5D, SF-6D, HUI-2, or HUI-3) regarding responsiveness to improvement in RA patients. The authors concluded that more head-to-head comparisons are required in longitudinal studies across the spectrum of RA disease severity.

Therefore, Harrison et al compared the responsiveness of the EQ-5D and SF-6D across a range of arthritis severity (15). They showed that the SF-6D was more responsive to improvement than the EQ-5D. This finding was consistent with previous studies of patients with RA when external responsiveness was examined based on the patient global assessment of disease activity (16) or when internal responsiveness (responsiveness to treatment) was examined (17). Nevertheless, in the study by Conner-Spady and Suarez-Almazor (18), the EQ-5D was found to be the most responsive to improvement over 1 year when the external criterion to define improvement was self-perceived health change. The patients



in this study were, however, consecutive patients with musculoskeletal disease, of which 51% were diagnosed as having RA. Because more head-to-head comparisons are required across the spectrum of RA disease severity, we would like to contribute to the evidence of both internal and external responsiveness in RA patients with active disease. In our study, external responsiveness was based on a patient-reported criterion (self-perceived health change) as well as a clinically meaningful criterion (European League Against Rheumatism [EULAR] response criteria based on the Disease Activity Score in 28 joints [DAS28]).

The aim of this study was to compare the responsiveness of the EQ-5D and SF-6D to improvement in patients with active RA treated with TNF blockers. RA is a well-known example of a serious chronic condition for which TNF blockers make an impressive and relevant difference.

PATIENTS AND METHODS

Patient selection

Data were from the prospective Dutch Rheumatoid Arthritis Monitoring (DREAM) registry. In the DREAM registry, 13 Dutch hospitals monitor and evaluate the use of anti-TNF in daily clinical practice. After signing informed consent forms, RA patients who are starting TNF for the first time are included in the DREAM registry. Patients are included in the registry when they are diagnosed with RA by the rheumatologist according to the 1987 American College of Rheumatology (ACR) criteria (19). Further selection criteria include moderate to high disease activity (DAS28 > 3.2), previous treatment with at least 2 other disease-modifying antirheumatic drugs, including methotrexate, and adequate treatment (25 mg per week for methotrexate) for 2 to 3 months without sufficient effect unless the treatment is ended due to serious side effects. The DREAM registry prospectively captures data on disease activity and outcome every 3 months.

For this study, eligible patients were selected post hoc from the DREAM registry. All patients with available data at baseline (range used: 6 weeks before starting the first TNF blocker to 1 week after starting) and 1 year (range used: 1 year \pm 6 weeks) were included in this study.

Measures

EQ-5D

The EQ-5D is a preference-based instrument to measure HRQOL (20). The instrument consists of a descriptive system and a valuation of health states. The EQ-5D has 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and each dimension has 3 levels (no problems, some problems, and extreme problems). Respondents have to indicate the level of problems they have on the 5 dimensions. Their responses correspond to 1 of the 243 possible EQ-5D health states, ranging between the best health state of having no problems on any dimension (11111) and the worst state of having extreme problems on all 5 dimensions (33333). A preference-based scoring function that relies on regression techniques is used to assign a utility score to respondents' health states. Different value sets obtained in different countries are available, including The Netherlands (21). In this study, we used the British value set based on the time trade-off to calculate EQ-5D utility scores (20). In this way, we ensured a fair comparison with the SF-6D, for which no Dutch value set is available. This algorithm ranges from -0.59 to +1.00, with negative values indicating states worse than death.

SF-6D

The SF-6D (11) is a preference-based instrument derived from the Short Form 36 (SF-36) (22), a widely used instrument to measure HRQOL. Like the EQ-5D, the SF-6D consists of a descriptive system and a valuation of health states. The SF-6D has 6 dimensions (pain, mental health, physical functioning, social functioning, role limitations, and vitality) and each dimension has 4 to 6 levels. The SF-6D was constructed from 11 items of the SF-36 and defines 18,000 health states. In our study, we used the British value set based on the standard gamble to assign SF-6D utility scores to respondents' health states (11), because no Dutch value set is available. The scoring range of the British value set varies from -0.29 to +1.00.

Self-reported health transition item

In the SF-36, the self-reported health transition item measures the perceived change in general health over the past year on a 5-point scale ranging from "much better" to "much worse." This item validly measures perceived change in health status in RA patients (23).

DAS28

The DAS28 is a measure of disease activity developed for RA that combines the number of swollen and tender joints using 28 joint counts, the erythrocyte sedimentation rate, and the patient's general health on a visual analog scale in a single index score (24).

Statistical analysis

Based on the suggestion of Husted et al (25), internal and external responsiveness were measured. The internal responsiveness of a measure refers to its ability to detect change over a prespecified timeframe (25). The internal responsiveness of a measure depends on the effectiveness of the treatment that is provided (in this case, TNF blockers) as well as the choice of outcome measures to determine treatment efficacy (in this case, the EQ-5D and SF-6D) (25). External responsiveness reflects the extent to which changes detected by a measure over a specified timeframe relate to corresponding changes in a reference measure of health status (25). External responsiveness only depends on the choice of the external criteria used and not on the treatment (25).

To analyze internal responsiveness over 1 year, Wilcoxon's signed rank test and the standardized response mean (SRM) were used. Wilcoxon's signed rank test alone tested whether the utility scores of the EQ-5D and the SF-6D changed statistically significantly between baseline and 1 year. The SRM is calculated as the mean change score between baseline and 1 year divided by the SD of this difference (25). Using the criteria of Cohen (26), an SRM between 0.20 and 0.49 is indicative of a small effect, between 0.50 and 0.79 is indicative of a moderate effect, and an SRM of 0.80 and greater is indicative of a large effect. Ninety-five percent confidence intervals (95% CIs) for the SRMs were estimated based on bootstrap resampling, which was repeated 1,000 times (27). We chose to use the SRM to calculate internal responsiveness because the SRM value is not dependent on either the sample size (like the outcome of a paired t-test) or the variance of baseline scores (like the effect size). With the latter 2 measures, it is difficult to determine whether the observed value reflects true change or simply the size of the sample or the variance of baseline scores. Furthermore, other studies that examined the responsiveness of the EQ-5D and SF-6D in RA mostly used the SRM. Therefore, using the SRM makes it easier to compare study results. We expected to find a small or moderate SRM (≥ 0.4) for the EQ-5D



and a moderate or high SRM (≥ 0.6) for the SF-6D. Other studies in patients with RA receiving TNF α found small and moderate SRMs for the EQ-5D and SF-6D (0.44 and 0.64, respectively) (15) and moderate and high SRMs (0.64 and 0.87, respectively) (17). The comparative responsiveness of the EQ-5D and SF-6D was determined by comparing the SRMs and calculating a 95% CI for the difference in SRMs using the 1,000 bootstrap samples (27). SRMs were considered to be significantly different if the 95% CI of the difference in SRM did not contain the value zero.

To analyze external responsiveness, the relationship was assessed between changes in the measures (EQ-5D and SF-6D) and external criteria that were self-perceived health change (item SF-36) and disease activity (DAS28). Therefore, Spearman's rank correlation coefficients, receiver operating characteristic (ROC) curves, and SRMs were used. Spearman's rank correlation coefficients were computed between changes in scores on the measures (EQ-5D and SF-6D) and self-perceived health change scores, as well as between changes in EQ-5D or SF-6D scores and changes in the DAS28. ROC curves were plotted to demonstrate how successfully the change in EQ-5D or SF-6D scores over 1 year could discriminate between patients who improved on external criteria and those who did not improve. For the ROC curves, both external criteria had to be dichotomized. Patients were classified as "responders" (good and moderate responders) and "nonresponders," based on the EULAR response criteria based on the DAS28, which take into account the individual change in DAS28 and the level of DAS28 reached (28). Based on the self-perceived health change score, patients were classified as "improved patients" (self-perceived health much better or somewhat better than a year ago) or "nonimproved patients" (self-perceived health the same as a year ago, somewhat worse, or much worse than a year ago). The areas under the ROC curves (AUCs) were calculated to determine the accuracy to discriminate good responders from nonresponders and improved patients from nonimproved patients. The AUC range is from 0.5 (no discriminative accuracy) to 1.0 (perfect discriminative accuracy). An AUC between 0.5 and 0.7 indicates low accuracy, between 0.7 and 0.9 indicates moderate accuracy, and higher than 0.9 indicates high accuracy (29). A 95% CI for the difference in AUCs between the EQ-5D and SF-6D was computed; the AUCs were considered to be significantly different if this 95% CI did not contain the value zero.

SRMs were calculated separately for patients who improved and for patients who did not improve on the external criteria. A 95% CI for the difference in SRMs between the EQ-5D and SF-6D was computed using the bootstrap procedure (27); the SRMs were considered to be significantly different if this 95% CI did not contain the value zero. Analyses were performed in SPSS, version 17, MedCalc, version 11.2, and S-Plus, version 6.2.

RESULTS

Patient characteristics

Of the 397 patients selected, 119 (30%) did not have complete data on all 3 measures (EQ-5D, SF-36, and DAS28) at baseline and 1 year later. These patients could not be included for analysis. Of 278 patients, data were complete, and these patients were included for analysis. No significant differences in age, sex, disease duration, and disease activity were found between the included patients ($n=278$) and the nonincluded patients ($n=119$) from the DREAM registry (data not shown). Demographic and clinical characteristics are shown in Table 1. Patients had a high mean \pm SD disease activity at baseline (DAS28 5.3 ± 1.2). After 1 year, the mean \pm SD disease activity had declined to 3.6 ± 1.4 , which demonstrated improvements in patients' clinical status.

Table 1 Demographic and clinical characteristics at baseline (and after 1 year for pain and disease activity scores, n=278)*

	Baseline	1 year
Age, mean \pm SD years	55 \pm 12	
Sex, %		
Men	32	
Women	68	
Work status, %[†]		
Paid	34	
Unpaid	7	
Retired	29	
Disabled	30	
Disease duration, median (IQR) years	7 (2.5 - 15.4)	
Pain (VAS, range 0-100), median (IQR)	65 (45 - 75)	35 (20 - 55)
Disease activity (DAS28, range 0-10), mean \pm SD	5.3 \pm 1.2	3.6 \pm 1.4

* IQR = interquartile range; VAS = visual analogue scale; DAS28 = Disease Activity Score in 28 joints.

[†] Known for 188 patients.

Internal responsiveness

The EQ-5D and SF-6D scores changed significantly ($Z = -8.48$ and -9.75 , respectively; $P < 0.01$) over 1 year (median score of EQ-5D changed from 0.59 to 0.69 and median score of SF-6D changed from 0.60 to 0.67) (Table 2). In Table 3, the SRMs and 95% CIs are shown. The EQ-5D and SF-6D scores changed moderately over 1 year. The SF-6D was significantly more responsive to treatment than the EQ-5D (difference in SRM 0.16; 95% CI 0.05, 0.28). Neither significant floor nor ceiling effects ($\leq 15\%$ of participants scored the lowest or highest possible score [30]) were found in the utility scores of both instruments at baseline and after 1 year (Table 2).

Table 2 Median values and floor and ceiling values for the EQ-5D and SF-6D (n=278)*

	At baseline		After 1 year			
	Median (IQR)	Mean \pm SD	Floor/ ceiling, no. (%)	Median (IQR)	Mean \pm SD	Floor/ ceiling, no. (%)
EQ-5D	0.59 (0.16 - 0.69)	0.46 \pm 0.29	0/2 (1)	0.69 (0.59 - 0.76) [†]	0.61 \pm 0.25	0/17 (6)
SF-6D	0.60 (0.56 - 0.65)	0.60 \pm 0.09	0/0	0.67 (0.59 - 0.75) [†]	0.67 \pm 0.10	0/0

* EQ-5D = EuroQol-5D; SF-6D = Short Form 6D; IQR = interquartile range.

[†] Significant difference between baseline and 1 year ($P < 0.01$).



Table 3 Internal responsiveness statistics for the EQ-5D and SF-6D over 1 year (n=278)*

	SRM (95% CI)
EQ-5D	0.50 (0.36, 0.62)
SF-6D	0.67 (0.52, 0.79)
Difference	0.16 (0.05, 0.28)†

* EQ-5D = EuroQol-5D; SF-6D = Short Form 6D; SRM = standardized response mean; 95% CI = 95% confidence interval.

† Significant difference between the SRMs of both instruments.

External responsiveness

According to the EULAR response criteria, 40% of the patients were good responders, 38% were moderate responders, and 22% were nonresponders. Therefore, 78% were indicated as responders and 22% as nonresponders. The self-reported health transition item showed that 34% of the patients perceived their health as much better after 1 year of treatment, followed by 28% of the patients who perceived their health as somewhat better. The other patients perceived their health approximately the same (21%), somewhat worse (13%), or much worse (3%). Therefore, 63% of the patients had improved and 37% had not.

Moderate significant correlations ($P < 0.01$) were found between the change in either EQ-5D or SF-6D utility scores and self-perceived health change and change in disease activity (Table 4). The EQ-5D and SF-6D were moderately able to discriminate between improved and nonimproved patients and responders and nonresponders, respectively (EQ-5D AUCs 0.67 and 0.71, respectively; SF-6D AUCs 0.72 and 0.65, respectively) (Table 4 and Figure 1). No significant differences were found between the EQ-5D and SF-6D regarding the accuracy to discriminate between improved patients and nonimproved patients and between responders and nonresponders. The improvers and responders had moderate improvements (SRMs < 0.80) on the EQ-5D and large improvements (SRMs > 0.80) on the SF-6D. The SF-6D was significantly more responsive than the EQ-5D to improved patients and to responders (Table 4). The nonimprovers and nonresponders had no relevant improvements (SRMs ≤ 0.20) on both the EQ-5D and SF-6D (Table 4), and no significant differences were found between both instruments.



Table 4 External responsiveness statistics for the EQ-5D and SF-6D (95% confidence interval) over 1 year (n=278)*

	AUC	SRM improvers [†]	SRM nonimprovers [†]	SRM responders [‡]	SRM nonresponders [‡]	Spearman's rho
Self-perceived health change						
(self-reported health transition item)						
EQ-5D	0.67 (0.60, 0.73)	0.76 (0.62, 0.89)	0.18 (-0.02, 0.38)			0.32 (0.21, 0.42)§
SF-6D	0.72 (0.66, 0.78)	1.03 (0.86, 1.19)	0.19 (-0.02, 0.39)			0.44 (0.34, 0.53)§
Difference	0.05 (-0.01, 0.12)	0.27 (0.45, 0.12)¶	0.002 (-0.17, 0.20)			
EULAR response (DAS28)						
EQ-5D	0.71 (0.63, 0.79)			0.68 (0.52, 0.80)	0.02 (-0.23, 0.28)	0.37 (0.27, 0.47)§
SF-6D	0.65 (0.57, 0.73)			0.83 (0.69, 0.98)	0.20 (-0.09, 0.53)	0.30 (0.19, 0.40)§
Difference	0.06 (-0.01, 0.14)			0.15 (0.13, 0.02)¶	0.19 (-0.04, 0.46)	

* EQ-5D = EuroQol-5D; SF-6D = Short Form 6D; AUC = area under the curve; SRM = standardized response mean; EULAR = European League Against Rheumatism; DAS28 = Disease Activity Score in 28 joints.

† Improvers and nonimprovers based on dichotomized self-perceived health change scores.

‡ Responders and nonresponders based on EULAR response criteria.

§ Significant correlation between change in the external criteria and outcome measures (EQ-5D and SF-6D; $P < 0.01$).

¶ Significant difference between the SRMs of both instruments.



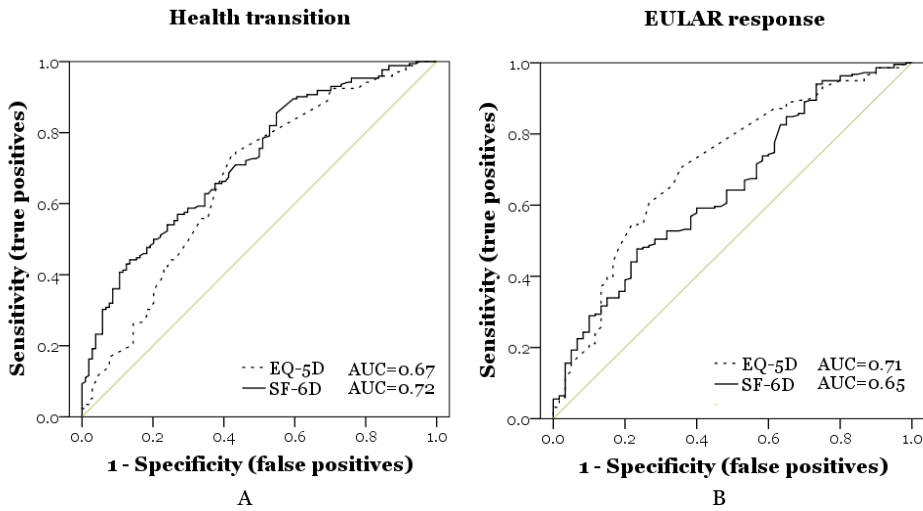


Figure 1 Receiver operating characteristic curves for differences in EuroQol-5D (EQ-5D) and Short Form 6D (SF-6D) utility scores between baseline and 1-year followup using **A**, self-perceived health change (self-reported health transition item Short Form 36) and **B**, European League Against Rheumatism (EULAR) response criteria based on disease activity (Disease Activity Score in 28 joints), respectively, as external criteria. AUC = area under the curve

DISCUSSION

This study of patients with RA treated with TNF blockers with impressive clinical improvement after 1 year demonstrated that the EQ-5D and SF-6D were moderately responsive to improvement. The SF-6D was significantly more responsive than the EQ-5D with respect to internal responsiveness. Regarding external responsiveness, we did not find significant differences between the EQ-5D and SF-6D with respect to the accuracy to discriminate between improved patients and nonimproved patients and between responders and nonresponders. When specifically considering patients who reported improved health (measured with the self-reported health transition item of the SF-36), the SF-6D was more responsive than the EQ-5D. This also applied to patients that were responders according to the EULAR response criteria. No significant differences were found between both instruments for nonimprovers and nonresponders.

In our study, both internal and external responsiveness were analyzed based on the distinction made by Husted et al (25). When SRMs are used to analyze internal responsiveness (responsiveness to treatment), the SRM values are independent of the sample size, but dependent on the treatment evaluated. In our sample, the effective treatment used provided impressive clinical improvement. Therefore, the SRMs were expected to be high. Unfortunately, these were only found to be moderate for both the EQ-5D and SF-6D, indicating a less than expected ability of these instruments to change. In most previous RA studies (15–17), the SF-6D was more responsive to improvement than the EQ-5D. In our study, the SF-6D was also significantly more responsive to improvement than the EQ-5D based on the internal responsiveness calculations. When comparing our results regarding internal responsiveness of the EQ-5D and SF-6D (0.50 and 0.67, respectively) with those of Harrison et al (0.44 and 0.64, respectively) (15), the SRM for the EQ-5D was slightly higher and the SF-6D was slightly similar to the SRM found by Harrison et al over 6 months. Nevertheless, they found a larger mean improvement in disease activity (2.2) than we did (1.7). According to the EULAR response criteria, 82% of the patients

improved in the study by Harrison et al compared with 78% in our study. Comparing our results with those of Russell et al (17) is difficult because the latter study did not report the improvement in disease activity.

A disadvantage of using the SRM to analyze internal responsiveness is that it does not account for the nonspecific score changes (often improvement) that may occur in clinically stable patients. The SRM only measures the extent of change over 2 occasions. The observed change may not reflect important changes in patients' clinical status (25,31). This can be overcome by using an external criterion. This external criterion should be a well-accepted indication of change in the condition of the patient; clinicians should regard change in this standard as clinically meaningful (25). This was the case in the present study because the Disease Activity Score is used by the rheumatologist to monitor treatment effects. In addition, response criteria such as the EULAR response criteria and the ACR response criteria (32) have been frequently used as criteria to measure the efficacy (1,2) and effectiveness of treatments (33).

Regarding the use of self-perceived health change, it might be difficult for people to accurately indicate the extent to which they have changed because of a possible interference of the present health state or psychological factors such as mood (34). Nevertheless, self-reported health transition questions have been validated in RA (23) and have been used as criteria for the measurement of responsiveness (16,18). Two studies performed in patients with RA (16) and musculoskeletal disease (18) examined external responsiveness by calculating SRMs for subgroups based on responses to a self-reported health transition question. When comparing our SRMs with those of Marra et al (16), we found a moderate to high external responsiveness of the EQ-5D and SF-6D, whereas they found a low external responsiveness of both instruments to self-reported improvement. Conner-Spady and Suarez-Almazor (18) only reported effect sizes. Based on their report of the mean \pm SD change, we could calculate SRMs for both instruments. When comparing our SRMs with the SRMs calculated on their data, we found a much higher responsiveness of both instruments in our study than in their study (0.76 versus 0.45 for the EQ-5D and 1.03 versus 0.42 for the SF-6D). Other studies in RA did not plot ROC curves, so our AUCs cannot be compared with other studies.

The potential of an instrument to detect change can be influenced by floor or ceiling effects in the overall utility scores of the instruments. In accordance with the review by Harrison et al (6), we found no clustering on the floor level of the SF-6D. However, the already high floor level (0.30) possibly limits the use of this instrument in patients with more severe RA. In addition, less change is possible in the utility scores when the floor level is 0.30. We did not find significant ceiling effects in the overall utility scores, which is in accordance with Hurst et al (35), Conner-Spady and Suarez-Almazor (18), and Harrison et al (15).

An important remark is that mean changes in utility scores are used to calculate the number of QALYs gained by a novel health technology. In our study, the mean change in utility scores of the EQ-5D was larger than the mean change in utility scores of the SF-6D. As a consequence, it is likelier that an intervention is considered as cost effective when using an EQ-5D as an SF-6D. Nevertheless, the SF-6D was more responsive to treatment than the EQ-5D. This underlines the larger variance around the EQ-5D (15,17). However, there might be less uncertainty regarding the relative cost-effectiveness of 2 treatments using the SF-6D because of its smaller variance (15).

Strengths of this study included the use of a large cohort of patients with RA in which an impressive clinically relevant improvement was reached, a variety of responsiveness analyses, and 2 clinically relevant criteria to analyze external responsiveness.

In conclusion, both the EQ-5D and SF-6D were only moderately responsive to improvement in RA patients treated with TNF blockers. In general, the SF-6D was more responsive than the EQ-5D.



ACKNOWLEDGMENTS

We would like to thank the DREAM registry and the participating hospitals for their contribution to the data collection.



REFERENCES

1. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;363:675–81.
2. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 2004;50:1400–11.
3. Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E, et al. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet* 2008;371:987–97.
4. Zink A, Strangfeld A, Schneider M, Herzer P, Hierse F, Stoyanova-Scholz M, et al. Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: comparison of patients according to their eligibility for major randomized clinical trials. *Arthritis Rheum* 2006;54: 3399–407.
5. Kievit W, Fransen J, Oerlemans AJ, Kuper HH, van der Laar MA, de Rooij DJ, et al. The efficacy of anti-TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. *Ann Rheum Dis* 2007;66:1473–8.
6. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DP. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *J Rheumatol* 2008;35:592–602.
7. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1–30.
8. Smith MD, Drummond M, Brixner D. Moving the QALY forward: rationale for change. *Value Health* 2009;12(Suppl. 1):S1–S4.
9. Cook KF, Ashton CM, Byrne MM, Brody B, Geraci J, Giesler RB, et al. A psychometric analysis of the measurement level of the rating scale, time trade-off, and standard gamble. *Soc Sci Med* 2001;53:1275–85.
10. Brooks R, de Charro F. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
11. 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.
12. 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.
13. Gold MR, Siegel JE, Russel LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
14. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171–8.
15. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Verstappen SM, Watson K, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Qual Life Res* 2009;18:1195–205.
16. Marra CA, Rashidi AA, Guh D, Kopec JA, Abrahamowicz M, Esdaile JM, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Qual Life Res* 2005;14:1333–44.
17. Russell AS, Conner-Spady B, Mintz A. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *J Rheumatol* 2003;30:941–7.
18. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care* 2003;41:791–801.
19. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
20. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095–108.
21. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, Busschbach JJV. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. *Ned Tijdschr Geneesk* 2005;149:1574–8. In Dutch.
22. Ware Jr JE, Sherbourne CD. The MOS 36-item Short-Form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
23. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A. Transition questions to assess outcomes in rheumatoid arthritis. *Br J Rheumatol* 1993;32:807–11.



24. Prevo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44–8.
25. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
26. Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
27. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ* 1997;6:327–40.
28. Van Gestel AM, Haagsma CJ, van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum* 1998;41:1845–50.
29. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
30. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
31. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. *Control Clin Trials* 1991;12(Suppl.):S142–S58.
32. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35.
33. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott D. The effectiveness of anti-TNF-therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. *Rheumatology* 2010;49:2313–21.
34. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50:869–79.
35. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity of Euroqol—a generic health status instrument— in patients with rheumatoid arthritis. *Br J Rheumatol* 1994;33:655–62.



Chapter



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

L. Buitinga
L.M.A. Braakman-Jansen
E. Taal
M.A.F.J. van de Laar

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.



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 (RA). 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 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 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 remaining 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: $1 - (\text{number of life years given up} / \text{remaining expected life years})$.

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 (Figure 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 clearness 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 1). 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.

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 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 1.

Table 1 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.

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 2). 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 2. 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 conclusion 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.

Table 2 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)
Perceived time duration		
Too long	2	1
Good	27	29
Introduction clear?		
Yes	30	30
No	0	0
Conclusion understandable?		
Yes	30	30
No	0	0
General experience**		
Unpleasant	1	3
Neutral	7	17
Pleasant	22	10

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

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 Figure 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 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 ($r^2 = 0.01$).



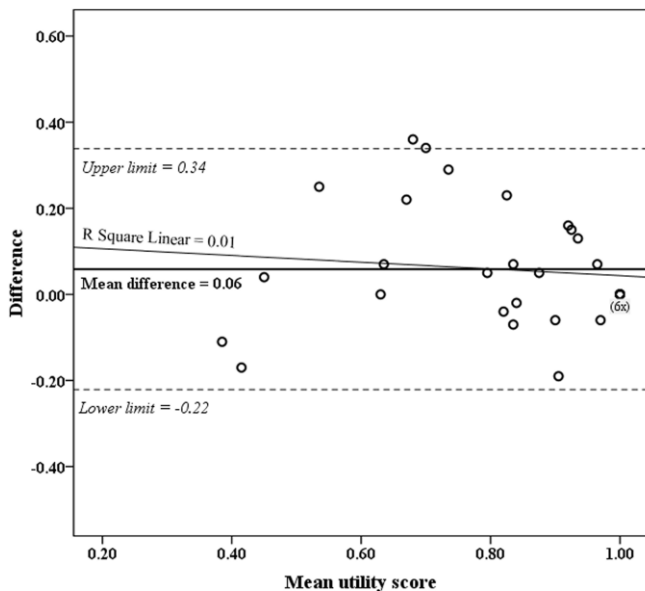


Figure 1 Bland-Altman plot: agreement between the interview and computer TTO of the 30 participants attending study 1 (interview versus computer TTO)

DISCUSSION

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



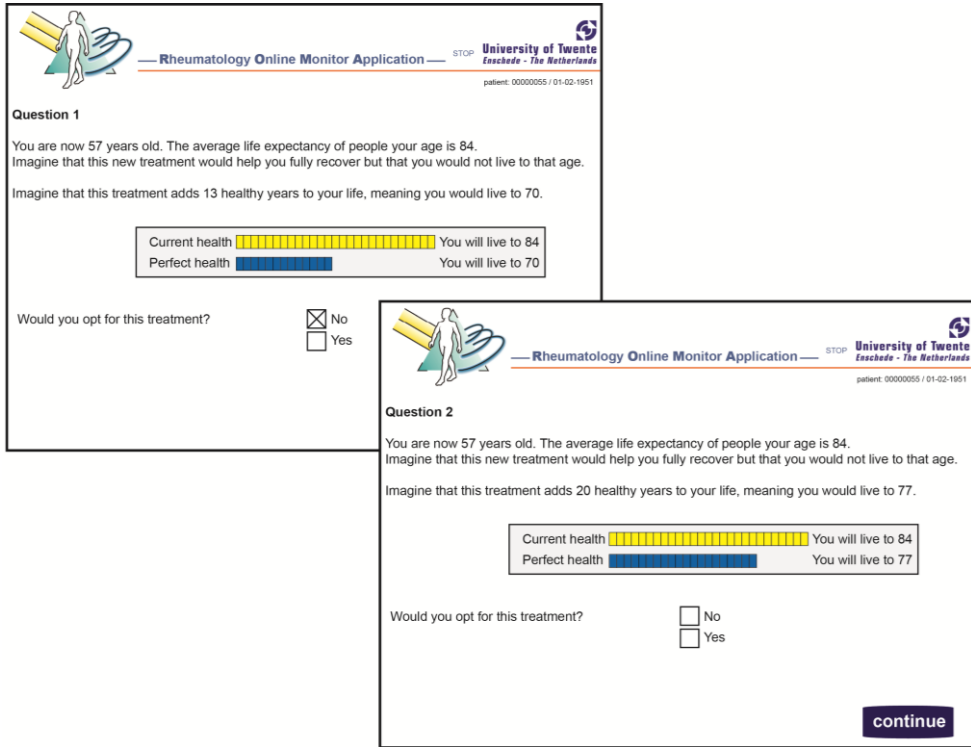


Figure 2 Two screen shots of the computer TTO*

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 7 years were added in the second question.

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

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, Hohlbauch A, Liang MH, Moreland L, et al.: The use of rheumatoid arthritis health-related quality of life patient questionnaires in clinical practice: lessons learned. *Arthritis Care 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 Busschbach JJ, Rutten FFH. Van kosten tot effecten: een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg; 2001. In Dutch.
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, Voigtlaender D, Moncur MM, Matheny TH, 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, Cull A, Gould A, Forman D, 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. U-titer: a utility assessment tool. In: Clayton PD, editor. *Proceedings of the Annual Symposium on Computer Application in Medical Care*. Washington D.C.: McGraw-Hill; 1991.701-5.
24. Lenert LA, Sturley A, Watson ME. iMPACT3: 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*. Oxford: Oxford University Press; 2003.



26. Centraal Bureau voor de Statistiek. Overlevingstafels 2006. Available from: <http://www.cbs.nl>. Accessed December 17, 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. Tjhuis 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.



Chapter



Construct validity of the interview Time Trade-Off and computer Time Trade-Off in patients with rheumatoid arthritis: a cross-sectional observational pilot study

L. Buitinga
L.M.A. Braakman-Jansen
E. Taal
M.A.F.J. van de Laar

ABSTRACT

Background The Time Trade-Off (TTO) is a widely used instrument for valuing preference-based health-related quality of life (HRQoL). The TTO reveals preferences for own current health ('utilities') on a scale anchored between death (0) and perfect health (1). Limited information on the external validity of the TTO is available. Aim of this pilot study was to examine the construct validity of both an interview TTO and a computer-based TTO in patients with rheumatoid arthritis (RA).

Methods Thirty patients visiting the outpatient rheumatology clinic participated. Construct validity was assessed by measuring convergent and discriminative validity. Convergent validity was assessed by calculating Spearman's correlations between the utilities obtained from the TTOs and pain, general health (rating scales), health-related quality of life (SF-36 and SF-6D) and functional status (HAQ-DI). Discriminative power of both TTO measures was determined by comparing median utilities between worse and better health outcomes.

Results Correlations of both TTO measures with HRQoL, general health, pain and functional status were poor (absolute values ranging from .05 to .26). Both TTOs appeared to have no discriminative value among groups of RA patients who had a worse or better health status defined by six health outcome measures. About one-third of respondents were zero-traders on each of the TTO measures. After excluding zero-traders from analysis, the correlations improved considerably.

Conclusions Both the interview TTO and computer TTO showed poor construct validity in RA patients when using measures of HRQoL, general health, pain and functional status as reference measures. Possibly, the validity of the TTO improves when using an anchor that is more realistic to RA patients than the anchor 'death'.



BACKGROUND

The Time Trade-Off (TTO) (1) is an instrument developed to assess effects of treatments in cost-utility analyses (CUAs) by measuring changes in health-related quality of life (HRQoL) directly by patients. The TTO reveals preferences for own current health ('utilities') on a scale anchored between death (0) and perfect health (1) by asking people how many life years they are willing to give up to become perfectly healthy. It is assumed that the more life years people are willing to trade off, the worse their health state is. The purpose of this measure is to capture the desirability of patients' own health state reflecting their health-related quality of life (HRQoL).

Traditionally, the TTO is administered by interview. The TTO can also be administered by questionnaire or computer. Furthermore, different methodological approaches to the TTO are used (2). This makes comparison between studies difficult. Differences in TTO procedures seem to influence utilities. For example, it has been found that utility scores are heavily influenced by the method of elicitation (ping-pong, titration) (3). Furthermore, the mode of administration (interview/computer/questionnaire) or the way the TTO question is formulated can influence utilities. Besides, the size of time frame that is used (e.g. fixed time period, life expectancy) has a great impact, since utilities are calculated as the proportion of the remaining lifetime sacrificed (2).

Few studies have examined psychometric properties of the interview TTO in rheumatoid arthritis (RA). The studies that reported on the construct validity, showed poor to moderate correlations between TTO and measures of HRQoL, functional status, disease activity and pain (4-6). It was found that the TTO was only able to discriminate between worse or better disease-specific HRQoL using the RAQoL (4,5), between worse or better outcomes on the dimensions 'symptom' and 'role' of the disease-specific AIMS-2 (5) and between worse or better mental health using the RAND-36 mental component summary scale (4). Tjshuis et al. showed that the TTO was able to discriminate between worse or better pain, worse or better disease activity and worse or better functional status (4). In contrast, Bejia et al. showed that the TTO was not able to discriminate between worse and better pain or worse and better disease activity (5).

Computer-based utility elicitation procedures to administer the TTO have been developed, for example iMPACT3 (7) and U-Titer (8). Studies in a range of conditions have used such computer-based programmes to administer a TTO using different procedures (7,9).

In this study, we report on preliminary results with respect to the construct validity of the TTO assessed in patients with RA using an interview TTO as well as a computer TTO, and using a standardised procedure for both TTOS. The first aim of this study was to examine convergent validity of the interview and computer TTO separately by correlating TTO utilities of both TTO measures with other patient-reported outcomes (PROs) in patients with RA. The second aim was to examine whether the interview and computer TTO were able to discriminate between worse and better patient-reported health outcomes.



METHODS

Patients and study design

Thirty consecutive outpatients (aged 18–85) of our rheumatology clinic who were diagnosed with RA participated. People who did not understand the Dutch language were excluded.

All participants completed the TTO twice with an interval of 14 days. Randomly the first TTO was either interview or computer-based, consequently followed by the other at the next assessment. Measures of pain, general health, health-related quality of life and functional status were administered at the first TTO assessment. Informed consent was obtained from all participants. According to legislation in the Netherlands (WMO), no approval of the ethical review board was indicated.

Measures

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?” A graphical aid was used to make the question more clearly. 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.

A life time perspective was adopted. Life expectancy calculations of the Dutch Central Bureau of Statistics (10) were used. 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 bisection method was applied to reach the point at which participants did not prefer one of the two options: staying in their health state for the rest of their lives or being perfectly healthy for a shorter life time. Therefore, 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 five years in perfect health was presented. If the person did not accept the trade, a remaining life expectancy of 15 years was presented. This process continued until the patient was indifferent between his or her own current health state according to his or her life expectancy and a shorter life in perfect health. Then, the TTO score was calculated by the formula: $1 - (\text{number of life years given up} / \text{remaining expected life years})$.

Computer TTO

Utilities were obtained by means of touch screens. Graphical presentations supported the TTO question (Figure 1). The computer TTO and the interview TTO were formatted equally, using similar formulations and graphical presentations. The bisection method was used to find the indifference point. 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. Our previous study (11) demonstrated that the test retest-reliability of this computer TTO was good (ICC = 0.88) and comparable with the test reliability of the interview TTO (4,5).

Question 1

You are now 57 years old. The average life expectancy of people your age is 84.
Imagine that this new treatment would help you fully recover but that you would not live to that age.
Imagine that this treatment adds 13 healthy years to your life, meaning you would live to 70.

Current health		You will live to 84
Perfect health		You will live to 70

Would you opt for this treatment? No Yes

Question 2

You are now 57 years old. The average life expectancy of people your age is 84.
Imagine that this new treatment would help you fully recover but that you would not live to that age.
Imagine that this treatment adds 20 healthy years to your life, meaning you would live to 77.

Current health		You will live to 84
Perfect health		You will live to 77

Would you opt for this treatment? No Yes

continue

Figure 1 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 7 years were added in the second question.

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

NRS pain and general health

Current severity of pain and current general health were both measured by a numerical rating scale (NRS), ranging from 0 (best) to 10 (worst).

SF-36

Physical and mental health were measured by calculating the physical and mental component summary scores (PCS and MCS) of the SF-36 version 2 (12), a generic descriptive instrument for measuring health-related quality life on eight dimensions (mental functioning, physical functioning, bodily pain, vitality, role limitations due to physical problems, role limitations due to emotional problems, social functioning and general health). The scores range from 0 to 100, whereby a higher score indicates a better health.

SF-6D

From the SF-36, SF-6D utility scores were derived, reflecting health state valuations of the general public (13). The utility scores range from 0 to 1, whereby a higher score indicates a better HRQoL.

HAQ-DI

The level of functional disability was assessed by the Health Assessment Questionnaire Disability Index (HAQ-DI) (14), a self-report measure consisting of eight categories (dressing and grooming, arising, eating, walking, hygiene, reach, grip and common daily activities). The HAQ score ranges from 0 to 3, whereby a higher score indicates a worse functional status.

Statistical analysis

To examine the presence of an order effect between participants who started with the interview TTO or with the computer TTO, a Mann–Whitney U-test was performed.

Construct validity was assessed by measuring convergent and discriminative validity. Convergent validity of the interview and computer version was assessed by calculating Spearman's correlations between each of the TTOs with the NRS for pain and general health, SF-36, SF-6D and HAQ-DI. Moderate correlations (0.40–0.59) are expected: all measures (except for the SF-6D) are descriptive, and most instruments only capture one or some aspects of the construct quality of life. The SF-6D yields utilities, but these are derived from a general public. A sample of 29 participants is required to demonstrate a significant moderate Spearman's correlation of 0.50 with an alpha of 0.05 (one-tailed) and a power ($1-\beta$) of 0.80. Discriminative power of the interview TTO and computer TTO was determined by comparing median utilities between worse and better pain, general health, HRQoL and functional status. Therefore, the outcome measures were dichotomised by the median score. A worse health outcome was defined by the \leq median value of the outcome measure. A better health outcome was defined by the $>$ median value of the outcome measure. Because of a difference in scaling of the NRS (Pain and General Health) and HAQ-DI, a worse health outcome on these instruments was defined by the $>$ median value of the outcome measure. A better health outcome was defined by the \leq median value of the outcome measure. The Mann–Whitney U-test was used to test significance. Data were analysed using SPSS version 16.0.

RESULTS

Subjects

All 30 patients who participated completed both TTO measures. Demographic, clinical and psychosocial characteristics and utility scores for both TTO methods of the 30 patients are shown in Table 1. Median TTO utility scores were 0.87 (interview TTO) and 0.85 (computer TTO). Ten participants did not want to trade any life year for perfect health on the interview TTO; eight participants did not want to trade any life year for perfect health on the computer TTO (zero-traders). Zero-traders did not have a significantly different self-reported health than traders (data not shown). Six zero-traders on the computer TTO were also zero-traders on the interview TTO. Two zero-traders on the computer TTO were not zero-traders on the interview TTO, and two zero-traders on the interview TTO were not zero-traders on the computer TTO.

Table 1 Demographic, clinical and psychosocial characteristics

	N (30)
Age (mean years \pm SD)	58 \pm 13
Gender (%)	
Men	27
Women	73
Disease duration (median years (IQR))	10 (19-3)
Marital status (%)	
Single	17
Married/Living together	83
Educational level (%)	
Low	50
Moderate	30
High	20
Work status (%)	
Paid work	40
Housekeeping	10
retired/unemployed/disabled	50
Utilities (Median (IQR))	
Interview TTO	0.87 (0.80-0.87)
Computer TTO	0.85 (0.58-0.85)
Pain (numerical rating scale) (median (IQR))	4 (2-6)
General Health (numerical rating scale) (median (IQR))	5 (3-6)
Descriptive health-related quality of life (SF-36) (median (IQR))	
Physical	40.05 (34.22-46.73)
Mental	40.99 (35.68-43.95)
Preference-based health related quality of life (SF-6D) (median (IQR))	0.63 (0.59-0.75)
Functional status (HAQ-DI) (median (IQR))	0.88 (0.34-1.38)

Test order

For neither interview TTO utility scores nor computer TTO utility scores an effect of test order was found ($P = 0.37$ and $P = 0.73$ respectively). So, no significant differences in utility scores existed between patients who started with the interview TTO or computer TTO.



Construct validity: Convergent and discriminative validity

Correlations between utilities and scores on rating scales for pain and general health, SF-36, SF-6D and HAQ-DI are shown in Table 2. In the entire group of RA patients, poor correlations were found between either both TTOs and the NRS measuring pain and general health, physical and mental component summaries of the SF-36, the SF-6D and the HAQ-DI. After exclusion of zero-traders, the correlations were moderate and significant between either the interview TTO and the NRS measuring pain ($r = -0.38$) and general health ($r = -0.42$), the physical component summary of the SF-36 ($r = 0.45$), and the SF-6D ($r = 0.45$). In addition, moderate and significant correlations were found between either the computer TTO and the NRS pain ($r = -0.47$) and the HAQ-DI ($r = -0.38$) (Table 2). All other correlations remained non-significant after exclusion of zero-traders. Discriminative properties of both TTOs are shown in Table 3. Both TTOs proved to have no discriminative value between worse and better health outcomes for all six outcome measures. Performing these analyses without the zero-traders resulted in larger and significant median differences in interview TTO scores between worse and better outcomes for the physical component summary of the SF-36 (difference in mean rank = 5.50), the NRS measuring pain (difference in mean rank = 5.83) and the SF-6D (difference in mean rank = 4.85). In addition, larger and significant median differences in computer TTO scores between worse and better outcomes were found for the NRS measuring pain (difference in mean rank = 6.30). No significant median differences in TTO scores between worse and better health outcomes were found for the other outcome measures.



Table 2 Spearman's correlations (95% confidence intervals) of interview TTO or computer TTO utilities with health outcome measures for the total sample and for the traders only

	Interview TTO (N=30)	Interview TTO traders only (N=20)	Computer TTO (N=30)	Computer TTO traders only (N=22)
Pain (NRS)[†]	-0.10 (-0.45, 0.27)	-0.38 (-0.70, 0.08)*	-0.26 (-0.57, 0.11)	-0.47 (-0.74, -0.05)*
General Health (NRS)[†]	-0.08 (-0.43, 0.29)	-0.42 (-0.73, 0.03)*	-0.05 (-0.40, 0.32)	-0.13 (-0.53, 0.30)
Descriptive HRQoL (SF-36)				
PCS	0.16 (-0.22, 0.49)	0.45 (0.01, 0.74)*	0.22 (-0.15, 0.54)	0.35 (-0.08, 0.67)
MCS	0.24 (-0.14, 0.55)	0.32 (-0.14, 0.67)	0.22 (-0.15, 0.54)	0.24 (-0.20, 0.60)
Preference-based HRQoL (SF-6D)	0.18 (-0.21, 0.50)	0.45 (-0.01, 0.73)*	0.11 (-0.27, 0.45)	0.20 (-0.24, 0.57)
Functional status (HAQ-DI)[†]	-0.07 (-0.42, 0.29)	-0.20 (-0.59, 0.27)	-0.21 (-0.53, 0.17)	-0.38 (-0.69, 0.05)*

NRS = numerical rating scale; HRQoL = health-related quality of life; PCS = Physical Component Summary; MCS = Mental Component Summary.

[†]Negative correlations were caused by a difference in scaling between the TTO and three health outcomes: a higher TTO score means a better HRQoL, whereas a higher score on these health outcomes means a worse health outcome.

* Significant correlations ($P < 0.05$).

Table 3 Discrimination of the interview and computer TTO between worse and better health outcomes^{a,d}

	Interview TTO utilities (N=30)			Computer TTO utilities (N=30)		
	Worse health outcomes ^{b,c}	Better health outcomes	P	Worse health outcomes ^{b,c}	Better health outcomes	P
Pain (NRS)^e	0.87 (0.81-0.97)	0.87 (0.57-1.00)	0.48	0.60 (0.50-1.00)	0.85 (0.74-0.97)	0.15
General Health (NRS)	0.87 (0.79-0.99)	0.93 (0.77-1.00)	0.31	0.87 (0.58-1.00)	0.82 (0.57-0.98)	0.23
SF-36						
PCS	0.86 (0.67-1.00)	0.88 (0.80 -1.00)	0.24	0.77 (0.53-1.00)	0.85 (0.71-1.00)	0.26
MCS	0.87 (0.63-1.00)	0.87 (0.80 -1.00)	0.45	0.84 (0.50-1.00)	0.85 (0.60-1.00)	0.28
SF-6D	0.87 (0.79-1.00)	0.88 (0.78-0.96)	0.36	0.86 (0.58-1.00)	0.81 (0.58-0.95)	0.29
HAQ-DI	0.91 (0.82-1.00)	0.85 (0.70-0.99)	0.48	0.86 (0.52-1.00)	0.85 (0.59-0.92)	0.37

^a Utilities are expressed in median scores (interquartile range); ^b For NRS (Pain), NRS (General Health) and HAQ-DI: a worse health outcome was defined by the > median value of each outcome measure; ^c For the SF-36 and SF-6D: a worse health outcome was defined by the ≤ median value of each outcome measure; ^d Median values for the NRS (Pain), 4; for the NRS (General health), 5; for the SF-36 PCS, 40.05, MCS, 40.99; for the SF-6D, 0.63; for the HAQ-DI, 0.88; ^e NRS = numerical rating scale

DISCUSSION

This pilot study showed that the construct validity of both the interview TTO and computer TTO was poor in patients with RA when using measures of HRQoL, general health, pain and functional status as reference measures. After exclusion of zero-traders from analysis, the results improved. This finding was expected, because zero-traders did not have a significantly different health status compared with traders. Indications of the poor convergent validity of the TTO were also found in other studies in RA and studies in other diseases (4-6,9,15-17). In most of these studies it was unclear how many participants were zero-traders and whether they were in- or excluded. One study reported similar results when in- or excluding zero-traders from analysis (15). In our study, we did not find the TTO to be discriminative for any of the health outcome measures used. Other studies found evidence for and against its discriminative ability (4,5,9,16). Contradicting findings were found for pain and disease activity scores in patients with RA (4,5) and for functional status scores in patients with cardiovascular disease (9,16).

All these studies were found to have differences in the TTO procedure applied. This might explain the contradicting results regarding the discriminative ability of the TTO. Beside the mode of administration, studies differed in the time frame used (remaining life expectancy (4,5,16-18), time frame dependent on age group (6) or not mentioned (15)). Furthermore, some studies described the way in which people had to think about current health (16,17) and/or about the anchors perfect health (4,5,16) and death (16), whereas other studies did not (6,9,18). One study used a symptom-free anchor ('no angina') instead of 'perfect health' (9). In many studies it was stated that a visual aid was used, although no further information was given about its representation (4-6). Besides, many studies did not report the precise method of elicitation (e.g. ping-pong) (4-6,9,18).

In our study, the TTO procedure applied was precisely described, facilitating the comparison with other studies. Strengths of this study were the fact that we used two different TTO assessments and that we used a broad set of PROs in a homogeneous population consisting of RA patients. A limitation of this study was the use of a small convenience sample.

There are several explanations possible for the results of our study, irrespective of the TTO procedure used. First, the low correlation with the SF-6D, another preference-based instrument, can be partly explained by the difference in perspective used to obtain utilities. SF-6D utilities are derived from the general public, so these scores represent a societal perspective. TTO scores were directly calculated from the patients' preferences, representing a patient perspective. Secondly, except for the SF-36 and SF-6D, the comparators used in this study only measure one aspect (e.g. functional status) of the construct quality of life. Furthermore, except for the SF-6D, the comparators are descriptive which implies that valuations of health states are not assessed. With these measures patients are asked about their levels of impaired health or pain, whereas personal preferences toward their health state remain unrevealed. It is possible that people with the same health state report different utilities if they have different 'aspirations' (18). Nease et al. illustrate this by the example that inability to walk 'more than a city block' does not have to be a limitation if someone does not desire to be active (18). Therefore, it would be worthwhile to examine in future studies whether it is better to validate the TTO against individualized measures of personal preferences, such as the SEIQOL (19,20) or MACTAR (21). Thirdly, it has been found that preferences are prone to biases inherently to the nature of the TTO, such as loss aversion. Loss aversion can be observed when a choice has to be made between 'remaining the status quo' (remaining in the current health state) and 'accepting an alternative to it' (trading off life years for perfect health). In that case people will evaluate the advantages and disadvantages of the alternative in terms of losses and gains (22). The TTO asks people about their willingness to trade off life years (a loss)



for optimal health (a gain) (23). Because 'losses loom larger than gains' (22), people become reluctant to give up life years. This will result in higher utilities, as supported by findings of Van Osch et al. (24). Furthermore, TTO utilities might be influenced by other factors that are unrelated to current health (15), such as family-related aspects, for example having children (2) or seeing grandchildren grow up (17). Finally, the nature of the disease can influence utilities. Asking patients to trade off life years may feel unrealistic, because patients with RA do not perceive their disease as life-threatening (6). Therefore, people may be less willing or not willing at all to trade off life years. Our results are indicative of this: irrespective of health, a relatively large number of participants were not willing to trade any life year for perfect health. For chronic illnesses such as RA there may be more realistic health-related anchors, for example 'becoming dependent on others' and 'having increased physical limitations', which were reported by RA patients to worry them (25,26). It could be examined whether the validity of the TTO improves when changing the trade-off about dying earlier in other more realistic (health-related) trade-offs. The use of a 'chained' TTO procedure could also improve the validity of the TTO. In a chained procedure, the health state of interest is not directly compared with death but indirectly with the aid of an intermediate anchor health state (27-29). A limitation is that a chained procedure is more complex, because it adds an additional step to the valuation process, possibly leading to extra noise (28). Limited research has been performed on the chained TTO and has been mainly applied in temporary health states (28-30). For chronic health states it has been shown that chained TTOs are systematically biased upwards (when the worst endpoint was varied) or downwards (when the best endpoint was varied), but that it is possible to correct for these biases (31). However, the respondents were not patients, but healthy people and women at high risk for breast cancer. Research in chronically ill patients examining the validity of the chained TTO for chronic states is lacking.

CONCLUSIONS

In conclusion, both the standardised interview TTO and standardised computer TTO showed similar poor results regarding construct validity when using measures of HRQoL, general health, pain and functional status as reference measures. Possibly, the validity of the TTO can be improved by replacing the anchor 'death' by an anchor that is more realistic to RA patients. Future studies in which direct patient reported utilities are derived, could start with the development of a TTO instrument using realistic anchors for RA patients. This instrument could be validated against individualized measures of personal preferences, such as the SEIQOL or MACTAR instrument.

ACKNOWLEDGEMENTS

We would like to thank the participants and the rheumatology department for their contribution to this study. We are also very grateful to André Brands, who programmed the computer TTO. This study was financially supported by a grant from the foundation: 'Stichting Onderzoek en Zorgontwikkeling Reumacentrum Twente'.

REFERENCES

1. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7:118–33.
2. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Econ* 2005;14:39–53.
3. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitation. *Med Decis Making* 1998;18:76–83.
4. Tjihuis 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.
5. 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.
6. Witney AG, Treharne GJ, Tavakoli M, Lyons AC, Vincent K, Scott DL, et al. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology* 2006;45:975–81.
7. Lenert LA, Sturley A, Watson ME. iIMPACT3: internet-based development and administration of utility elicitation protocols. *Med Decis Making* 2002;22:464–74.
8. Sumner W, Nease R, Littenberg B. U-titer: a utility assessment tool. In: Clayton PD, editor. *Proceedings of the Annual Symposium on Computer Application in Medical Care*. Washington D.C.: McGraw-Hill; 1991:701–5.
9. Nease Jr RF, Kneeland T, O'Connor GT, Sumner W, Lumpkins C, Shaw L, et al. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. *J Am Med Assoc* 1995;273:1185–90.
10. Centraal Bureau voor de Statistiek. Overlevingstafels 2006. Available from: <http://www.cbs.nl>. Accessed December 17, 2007.
11. Buitinga L, Braakman-Jansen LMA, Taal E, van de Laar MAFJ. A computer Time Trade-Off: a feasible and reliable alternative for the interview Time Trade-Off in rheumatoid arthritis. *Clin Exp Rheumatol* 2011;29:783–9.
12. Ware Jr JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
13. 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.
14. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
15. Maor Y, King M, Olmer L, Mozes B. A comparison of three measures: the time trade-off technique, global health-related quality of life and the SF-36 in dialysis patients. *J Clin Epidemiol* 2001;54:565–70.
16. Lalonde L, Clarke AE, Joseph L, Mackenzie T, Grover SA. Comparing the psychometric properties of preference-based and nonpreference-based health-related quality of life in coronary heart disease. *Qual Life Res* 1999;8:399–409.
17. Khanna D, Ahmed M, Furst DE, Ginsburg SS, Park GS, Hornung R, et al. Health values of patients with systemic sclerosis. *Arthritis Rheum* 2007;57:86–93.
18. Nease Jr RF, Tsai R, Hynes LM, Littenberg B. Automated utility assessment of global health. *Qual Life Res* 1996;5:175–82.
19. McGee HM, O'Boyle CA, Hickey A, O'Malley K, Joyce CRB. Assessing the quality of life of the individual: the SEIQoL with a healthy and a gastroenterology unit population. *Psychol Med* 1991;21:749–59.
20. O'Boyle CA, McGee H, Hickey A, O'Malley K, Joyce CRB. Individual quality of life in patients undergoing hip replacement. *Lancet* 1992;339:1088–91.
21. Tugwell P, Bombardier C, Buchanan WW. The MACTAR patient preference disability questionnaire - an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol* 1987;14:446–51.
22. Kahneman D, Tversky A. Choices, values, and frames. *Am Psychol* 1984;39:341–50.
23. Van Osch SMC, Wakker PP, Van Den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making* 2004; 24:511–7.



24. Van Osch SMC. The construction of health state utilities; 2007. PhD thesis. Leiden University Medical Center (LUMC), Leiden University, Department of Medical Decision Making.
25. Lempp H, Scott D, Kingsley G. The personal impact of rheumatoid arthritis on patients' identity: a qualitative study. *Chronic Illn* 2006;2:109–20.
26. McPherson KM, Brander P, Taylor WJ, McNaughton HK. Living with arthritis – what is important? *Disabil Rehabil* 2001;23:706–21.
27. Spencer A. The implications of linking questions within the SG and TTO methods. *Health Econ* 2004;13:807–18.
28. Locadia MA, Stalmeier PFM, Oort FJ, Prins MH, Spranger MAG, Bossuyt PMM. A comparison of 3 valuation methods for temporary health states in patients treated with oral anticoagulants. *Med Decis Making* 2004;24:625–33.
29. Jansen SJT, Stiggelbout AM, Wakker PP, Vliet-Vlieland TPM, Leer JH, Nooy MA, et al. Patients' utilities for cancer treatments: a study of the chained procedure for the Standard Gamble and Time Tradeoff. *Med Decis Making* 1998;18:391–9.
30. Johnston K, Brown J, Gerard K, O'Hanlon M, Morton A. Valuing temporary and chronic health states associated with breast screening. *Soc Sci Med* 1998;47:213–22.
31. Stalmeier PFM. Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. *Med Decis Making* 2002;22:53–64.



Chapter



Future expectations and worst-case future scenarios of patients with rheumatoid arthritis: a focus group study

L. Buitinga
L.M.A. Braakman-Jansen
E. Taal
M.A.F.J. van de Laar

ABSTRACT

Objective Over the past 15 years, developments in the treatment of rheumatoid arthritis (RA) have resulted in better clinical outcomes. The aim of the present study was to explore how patients think their RA will influence their lives in the future, and which of these future expectations would be the worst for them to experience.

Methods A focus group study was performed in 16 RA patients. Three groups were heterogeneously composed, based on age group (18–40, 40–65, 65–80), gender and having a paid job or not. Patients were asked about the expected future impact of RA and worst-case future scenarios. Transcripts were coded by three researchers under the main components of the International Classification of Functioning, Disability and Health. The codes were discussed until agreement was reached about all codes.

Results Dependency on others, increasing dependency on medication, inability to walk, activity limitations and worsening fatigue were mentioned as worst-case future scenarios. Further concerns were raised about the acceptance of RA and possible disappearance of physicians' expertise. Nevertheless, hope and positive feelings were expressed toward continuous medication improvements.

Conclusion The present study provided insight into RA patients' future expectations and worst-case future scenarios. The results may be of help in the development of support interventions to put concerns and worst-case future scenarios into a realistic perspective. Furthermore, insight into patients' worst-case future scenarios could be used to improve the validity and responsiveness of the Time Trade-Off, an instrument to measure preference-based health-related quality of life.



INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease characterized by symmetrical, erosive polyarthritis. As a result of advances in treatment and treatment strategies that control disease activity and delay radiological damage (1,2), patients experience a reduction in symptoms such as fatigue and an increase in well-being, which are more important outcomes to them than traditional outcomes such as joint swelling and stiffness (3).

In light of the continuous treatment innovations that provide considerable benefits to patients' health, it is interesting to examine how RA patients think their disease will influence their future lives (i.e. future expectations related to RA), and which of these future expectations would be the worst for them to experience (i.e. worst-case future scenarios). This insight can reveal whether patients' expectations and worst-case future scenarios related to their RA are realistic and can help in the development of support interventions to put these into a realistic perspective and to help patients in coping with fears and concerns, in order to improve their health-related quality of life (HRQoL).

Knowing patients' worst-case future scenarios can also be valuable for the measurement of preference-based HRQoL. The Time Trade-Off (TTO) instrument (4) measures patients' valuations of their own health state (patients' utilities) by asking the number of life years that patients are willing to trade for perfect health (i.e. to avoid being in their current health state). Thus, patients have to compare their current health state with the anchor state 'death'. In this way, an estimation of the severity of the current health state is obtained. It is assumed that the more life years patients are willing to trade off, the worse their health state is. However, 'dying earlier' may not be perceived by RA patients as a realistic consequence of their disease. It is possible that replacing the anchor 'death' by a worst-case anchor that is perceived by RA patients as a realistic consequence of RA improves the validity and responsiveness of this instrument.

The first aim of the present study was to examine RA patients' future expectations. The second aim was to reveal worst-case future scenarios.

METHODS

Patients and study design

RA patients were recruited consecutively from the outpatient rheumatology clinic of a Dutch hospital in February 2009. Three heterogeneous focus groups of six patients each were composed, based on age group (18–40, 40–65, 65–80), gender and having a paid job or not, in order to enhance discussion (5). A semi-structured question format was used. The discussion was conducted by a moderator (L.B.). A note-taker (A.B.) was present. Every session took about an hour and a half. Data saturation was checked afterwards to determine the desirability of organizing additional groups. The conversations were recorded and transcribed verbatim.

Measures

The focus group sessions comprised two subjects: expectations about the future influence of RA (questions: *How do you envisage yourself living with rheumatism in the future and in what way do you think that your rheumatism will influence your life in the future?*) and worst-case future scenarios derived from these future expectations (questions: *Of all the expectations that have been listed so far, which one would you consider to be the worst? What are you most afraid of?*). In case of non- or



limited response, or in order to clarify responses, sub-questions were asked. The focus group schedule and the questions were constructed by three researchers (L.B., A.B., E.T.). All questions are shown in Appendix 1. To ensure that we did not miss out any worst-case future scenario mentioned, the note-taker wrote down all scenarios on cards. Subsequently, participants were asked to indicate if any scenario was missing or if scenarios belonged together.

Analysis

The conversations were individually coded and interpreted by three researchers (L.B., A.B., E.T.). The World Health Organization International Classification of Functioning, Disability and Health (ICF) (6) was used to categorize the coded expressions. To clarify the content of the expressions in a logical manner, the expressions were categorized under the main components of the ICF (body functions and structures, activities and participation, environmental factors and personal factors). The first, second, third and fourth level categories were not described. We categorized the objects of the expressions. To clarify this: when a participant expressed concerns about *the ability to walk* in the future, we did not categorize this worry as a personal factor, but we linked the object of this concern (italicized) to the ICF main component 'activities and participation'. The codes were discussed (by L.B., A.B. and E.T.) until agreement was reached about all codes.

Ethical approval

According to local regulations in the Netherlands (WMO), no approval of the ethical review board was needed. Only (non-intervention) studies with a high burden for patients have to be reviewed. Patients underwent no intervention or treatment, and the burden of participation was relatively low.

RESULTS

The demographic and clinical characteristics of the 16 participating RA patients are shown in Table 1. Data saturation was achieved sufficiently. An overview of participants' future expectations linked to the ICF domains is shown in Table 2.



Table 1 Demographic and clinical characteristics of the 16 participants in the three focus groups

Focus group	Participant	Age (years)	Gender	Marital status	Educational level	Work status	Disease duration (years)
1	1	55	Female	Single	Moderate	Paid work	5
	2	56	Female	Single	High	Paid work	19
	3	70	Female	Married	Low	Retired	3
	4	49	Male	Married	Moderate	Paid work	5
	5	62	Male	Married	Moderate	Paid work	7
2	1	60	Male	Married	High	Unemployed	19
	2	59	Female	Married	Low	Housekeeping	1
	3	68	Female	Single	Low	Retired	24
	4	60	Male	Married	Low	Paid work	4
	5	52	Female	Married	Moderate	Unemployed	1
	6	72	Male	Single	Low	Retired	37
3	1	72	Male	Single	High	Retired	24
	2	59	Male	Married	Low	Paid work	6
	3	68	Female	Married	High	Retired	39
	4	41	Female	Living together	High	Paid work	10
	5	72	Female	Living together	High	Retired	2



Body functions and structures

Fatigue was a prevailing, uncontrollable problem. Only two participants reported future concerns about *worsening fatigue*. For example:

‘. . . And then I think to myself. . . if this should get any worse, all I am doing is going to work, coming home, eating and throwing in the towel, that’s it really. . .’ (woman, 56 years, RA for 19 years).

Many participants experienced *side effects of medication* or had experienced these in the past, which they hoped not to face again. Uncertainties and fears were expressed concerning side effects:

‘I think I’m on quite a lot of different medicine at the moment, so I really think, couldn’t these be less? How long can my stomach and the rest of me put up with all of this?’ (woman, 70 years, RA for three years).

‘In fact, the very thought terrifies me. All those stomach troubles and the internal bleeding. It’s no picnic if you are sitting on the toilet at night and you suddenly see a whole mass of blood’ (man, 73 years, RA for 37 years).

One participant expected to have a *reduced life expectancy because of the medication* she had to take. Nevertheless, the reduced life expectancy did not outweigh the benefits of the medication:

‘It’s common knowledge that when you’re given medication for your rheumatism your life expectancy will probably decrease, and be a bit shorter. At least, that’s what I read all over the place. I do take it on board, also with regard to the medicines, but I put it into context by thinking: what is the average life expectancy anyway? How old am I now? And what are the risks of this actually happening to me?’ (woman, 72 years, RA for two years).

Participants did not mention any future concerns regarding *pain*. Participants experienced pain, but most people compared their current pain level with the level at diagnosis, and indicated that the pain had decreased tremendously. One participant explicitly stated that it would not be a problem if the pain worsened. Participants indicated that they got used to the pain; that they had to cope with it. For example:

‘With the pain being like it is now. . . well . . . even if it becomes worse. . . no, that is not such a problem, no. If you compare it to what it was like when I first got it. . .’ (man, 62 years, RA for seven years).

‘Now, almost none [pain], occasionally. You just get used to it, of course.’ (woman, 68 years, RA for 24 years).

‘You have to live with that. . . or you go under to it. . .’ (woman, 70 years, RA for three years).



Activities and participation

Many participants expressed concerns about *becoming dependent on others*. They felt that being unable to be independent and needing help from others regarding all kind of activities would be hard to take. For example:

‘. . .that you still have a little bit of independence and that you can still do things more or less . . . such as small errands to the shops’ (woman, 70 years, RA for three years).

In one focus group, concerns were particularly raised about self-care because of the intimacy of activities related to it. For example:

‘. . .that I’ll be dependent on my husband or children. . . I hope that it never gets to that stage, that would be terrible for me. . . And this is what terrifies me. . . I always had to help my mother to wash herself and to have a shower and I don’t know how many other things: no, it is not pleasant’ (woman, 59 years, RA for one year).

Many participants expressed uncertainties and fears about *activity limitations*, such as limitations in hobbies, driving or shopping. For example:

‘If I start to focus on the future, then I sometimes feel anxious.’

Interviewer: *‘And what are you afraid of exactly?’*

Participant:

‘That you can no longer do the things that you do now. . . Playing the guitar, for one. I’m really anxious about this. It is not work-related, it’s just a hobby, but I really love it. If I am not able to play anymore. . .’ (man, 49 years, RA for five years).

One participant anticipated the possibility of reduced capacities:

‘. . .I always fancied being an artist. . . I cancelled all my managerial positions a few years ago and that’s when I began to paint . . . I think to myself, if it gets to the point where I cannot do it anymore, then at least I will have a couple of my own paintings hanging on the wall’ (man, 59, RA for six years).

The process of *ageing* was seen as an important factor for influencing future capabilities:

‘I think that it might deteriorate a bit further. . . but that probably has a lot to do with one’s age as well. So in that sense you’re never be as fit as you used to be years ago anyway’ (woman, 55 years, RA for five years).



Losing independence and having limitations in the ability to perform activities are intertwined: the inability to perform activities leads inevitably to the need for assistance from others:

‘Yes, that you can still drive a car and that you’re still a bit independent and that you can manage to do everything to a certain point. . . like errands and shopping and you know pretty well that you can’t fill a shopping trolley and push it around, but still manage to do a bit’ (woman, 70 years, RA for three years).

One person emphasized her concerns about the *ability to walk*: although dependency would be terrible, the inability to walk was regarded as a tremendous, unimaginable, limitation:

‘I think that I would have this too, that I would find it terrible if I couldn’t walk anymore. Then of course you would be dependent on others, but I find it hard to imagine how it would be to be unable to walk . . . there are support aids available, but that’s not the point, the very thought that you cannot get into a car by yourself or jump on a bike, then you’re completely stuck. That is a major hindrance of course’ (woman, 68 years, RA for 39 years).

Environmental factors

One participant especially expressed concerns about becoming *increasingly dependent on medication*:

‘Being dependent on medication: that’s what I find terrible. The idea that you cannot live without taking the medicines. And that the amount will steadily increase. It begins with a cure for rheumatism, then something for the stomach complaints is added, and something for brittle bones, and before I knew it I was up to nine a day. Simply to counteract the side effects of the medicines a little. There’s no end to it. . . That’s what worries me the most’ (man, 72 years, RA for 24 years).

Positive feelings and hopes about *future improvements in medication* were expressed, also because of previous experiences. For example:

‘Then if you are told, this is the last available treatment . . . but in the meantime, new medicines appear . . . and it will continue. . . and then I’ll improve again. . . and this has all happened in the last few years . . . of course there are all sorts of side effects, but you know that something else will be made available after that’ (woman, 56 years, RA for 19 years).

Nevertheless, concerns were reflected in many expressions of hope. For example:

‘I was 45 when I got rheumatism and my children were still young. At the time, I sometimes found myself thinking: goodness, how am I going to manage? But just keep taking the medicine and hope that better medicines will become available, because – as I say – you have to stay hopeful’ (woman, 68 years, RA for 24 years).



One participant expressed the fear of *disappearing expertise* because of better career perspectives for physicians elsewhere:

‘The thing that worries me is. . . how long will it be before the doctors are dragged away to fill some academic position elsewhere? I have the feeling that we have brilliant doctors and specialists working here who will sooner or later say goodbye’ (woman, 72 years, RA for two years).

Personal factors

One participant hoped to keep the ability to accept the disease and to enjoy life:

‘I hope that I can continue to accept the fact that it’s there and focus on enjoying what we still can do. Because if you lose sight of that, then you’re truly lost’ (woman, 70 years, RA for three years).

Worst-case future scenarios

Of the future expectations listed, *being dependent on others* would be the worst ‘fear’ for the future for most participants. Not being able to function independently, but needing the assistance of others would be a great burden to them. In one focus group, dependency was particularly related to activities of *self-care*, such as showering and going to the toilet. Some participants expected to get used to it eventually. Uncertainty about the actual reaction when facing dependency was also raised. For example:

‘I haven’t given it much thought yet. But I can imagine that also for myself; that for me would be the worst thing too. Being washed by strangers. Maybe one gets used to it. . . I’d find it difficult the first three times and will become really nervous about it half an hour beforehand, but after two or three times one would get used to it, at least I hope so. . . It is true, it’s hard to say in advance how you would actually deal with everything that will cross your path. You don’t know yourself yet, you just don’t know how you would cope or react in that particular situation’ (man, 60 years, RA for 19 years).

For one participant an *increasing dependency on medication* was worrying and would be the worst-case future scenario. The inability to live without medication and the inevitability of accumulating medication worried him. For two participants, *losing the ability to walk* would be the worst thing they could experience in the future, for another it was the *inability to perform hobbies*, and for another it was *worsening fatigue*:

‘I’m more bothered by the fact that the feeling of tiredness doesn’t go away. Sometimes I think to myself, good grief: there I go – I have got to sit down again. That’s indeed my biggest bug-bear. I hope that doesn’t get worse’ (woman, 55 years, RA for five years).



Table 2 Future expectations and worst-case future scenarios categorized under the ICF

ICF category	Future expectations	Worst-case future scenarios
Body Functions and Structures	Expected decrease in life expectancy owing to medication	Worsening fatigue
	Concerns about worsening fatigue	
	Concerns about side effects of medication	
Activities and Participation	Expected decrease in activity performance as a result of ageing	Dependency on others
	Concerns about dependency on others	Inability to perform hobbies
	Concerns about activity limitations	Inability to walk
	Concerns about ability to walk	
Environmental Factors	Expected disappearance of expertise as a result of better career perspectives for physicians elsewhere	Increasing dependency on medication
	Concerns about increasing dependency on medication	
	Positive feelings and hope of continuously improved medication	
Personal Factors	Concerns about maintaining the ability to accept the disease	

ICF = International Classification of Functioning, Disability and Health.

DISCUSSION

The present focus group study in RA patients revealed diverse future expectations related to all ICF categories. Regarding worst-case future scenarios, dependency on others, worsening fatigue, inability to perform hobbies, the inability to walk and increasing dependency on medication were mentioned.

Remarkably, patients did not report any concern about the pain they experienced. Pain was not much of a problem for them. Either they compared their current pain level with the level at diagnosis, or they indicated that they got used to the pain. By contrast, a semi-structured interview study in RA patients (7) showed that pain was considered to be a main problem of the disease and was seen as problematic by the majority of participants (86%). Furthermore, a semi-structured interview study about RA patients' concerns revealed worries about increasing pain in the future (8). It is likely that this

difference can be explained by the fact that treatment strategies for RA have improved tremendously over the past decade.

Previous studies (3,7,9-10) have already demonstrated that dependency on others is a major problem for RA patients. The results of the present study regarding patients' concerns were consistent with those of other studies (8,11-12) with respect to concerns about losing independence and becoming a burden to others. Moreover, consistent with the present study, a previous study reported concerns about the inability to perform daily activities (8), and a more recent study about fatigue (13).

Not all expectations and worst-case future scenarios are likely to be realistic from a clinical perspective because of continuous advances in the treatment and management of RA (e.g. the introduction of tumour necrosis factor blockers (14,15) and treat-to-target strategies (1)). Nevertheless, from the patients' perspective, these expectations and worst-case scenarios are perceived as realistic.

The present results support the development of interventions that teach patients to cope with their concerns and to put expectations and worst-case future scenarios into a realistic perspective. Furthermore, the results could be used to improve the TTO. Studies applying the TTO have shown that a number of people are reluctant to trade any life years for perfect health. These people are called 'zero-traders' (16). Because of their unwillingness to trade any life years for perfect health, zero-traders give the impression that their health is perfect, leading to an overestimation of HRQoL. Consequently, the validity and responsiveness of the TTO are limited. An explanation for the existence of zero-traders is that giving up life years (i.e. dying earlier) is not perceived by RA patients as a realistic consequence of RA. This was confirmed in the current study, as dying earlier was not mentioned by most participants as a future expectation. Replacing the current anchor 'death' by an anchor that is perceived by RA patients as a realistic consequence of RA (e.g. 'dependency on others' or 'inability to walk') could improve the validity and responsiveness of the TTO. Obviously, a larger, quantitative study of RA patients is needed to validate the results.

A limitation of the present study was that participants might have felt inhibited to talk freely because of the group format. Furthermore, social pressure could have affected responses. Nevertheless, the possibility for participants to identify themselves with others sharing similar experiences could have facilitated disclosure (17). Besides, the process of group interaction has the advantage of facilitating participants to clarify and revise their opinions based on the contribution of others (18).

CONCLUSION

The present study provided insight into RA patients' future expectations and worst-case future scenarios. Regarding worst-case future scenarios, dependency on others, worsening fatigue, inability to perform hobbies, the inability to walk and increasing dependency on medication were mentioned. The results may be of help in the development of support interventions. Furthermore, insight into patients' worst-case future scenarios could improve the validity and responsiveness of the TTO, an instrument for measuring preference-based HRQoL.

ACKNOWLEDGEMENTS

The authors would like to thank the staff members from the outpatient rheumatology clinic and the participants for their contribution to this study. This study was financially supported by a grant from the foundation 'Stichting Onderzoek en Zorgontwikkeling Reumacentrum Twente'.



REFERENCES

1. Vermeer M, Kuper HH, Hoekstra M, Haagsma CJ, Posthumus MD, Brus HLM, et al. Implementation of a treat to target strategy in very early rheumatoid arthritis: results of the DREAM remission induction cohort. *Arthritis Rheum* 2011;63:2865–72.
2. Villeneuve E, Emery P. Rheumatoid arthritis: what has changed? *Skeletal Radiol* 2009;38:109–12.
3. Hewlett S, Carr M, Ryan S, Kirwan J, Richards P, Carr A, et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care* 2005;3:131–42.
4. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7:118–33.
5. Barbour RS. Making sense of focus groups. *Med Educ* 2005;39:742–50.
6. World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization; 2001.
7. Taal E, Rasker JJ, Seydel ER, Wiegman O. Health status, adherence with health recommendations, self-efficacy and social support in patients with rheumatoid arthritis. *Patient Educ Couns* 1993;20:63–76.
8. Bath J, Hooper J, Giles M, Steel D, Reed E, Woodland J. Patient perceptions of rheumatoid arthritis. *Nurs Stand* 1999;14:35–8.
9. Melanson PM, Downe-Wamboldt B. Confronting life with rheumatoid arthritis. *J Adv Nurs* 2003;42:125–33.
10. van Lankveld W, Naring G, van der Staak C, van't Pad Bosch P, van de Putte L. (1993). Stress caused by rheumatoid arthritis: relation among subjective stressors of the disease, disease status, and well-being. *J Behav Med* 1993;16:309–21.
11. Lempp H, Scott D, Kingsley G. The personal impact of rheumatoid arthritis on patients' identity: a qualitative study. *Chronic Illn* 2006;2:109–20.
12. Lütze U, Archenholtz B. The impact of arthritis on daily life with the patient perspective in focus. *Scand J Caring Sci* 2007;21:64–70.
13. Repping-Wuts H, Repping T, van Riel P, van Achterberg T. Fatigue communication at the out-patient clinic of rheumatology. *Patient Educ Couns* 2009;76:57–62.
14. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 2004;50:1400–11.
15. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;363:675–81.
16. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Econ* 2005;14:39–53.
17. Wong LP. Focus group discussion: a tool for health and medical research. *Singapore Med J* 2008;49:256–61.
18. Powell RA, Single HM. Focus groups. *Int J Qual Health Care* 1996;8:499–504.



Focus group questions**A Appendix 1****Main question 1.**

How do you envisage yourself living with rheumatism in the future and in what way do you think that your rheumatism will influence your life in the future?

Sub questions. (in case of non- or limited response)

- *Do you envisage your rheumatism getting better or worse in the future? Can you describe this in more detail (in what way will it get better/become worse)?*
- *(If everyone expects the rheumatic disease to improve): Just suppose for a moment that it will become worse than you thought; how might your life be influenced by your rheumatism then? How do you think you will handle this?*
- *What kind of influence does rheumatism have on your life now?*
- *Could you name one or two concrete situations in which you noticed this type of influence?*
- *Do you expect that this will continue to be the case in the future? In what way will your rheumatism have an influence on your life in the future?*

Main question 2.

Of all the expectations that have been listed so far, which one would you consider to be the worst? What are you most afraid of?

Sub questions. (to clarify responses)

- *Have these already been mentioned, or not?*
 - *Can you indicate what it is that makes you so afraid? Why are you so afraid of this?*
 - *Are there expectations listed on the board that are named differently, but have the same meaning for you?*
- Is there anything that you feel is missing in this discussion about expectations for the future?*



Chapter



Worst-case future scenarios of patients with rheumatoid arthritis: a cross-sectional study

L. Buitinga
L.M.A. Braakman-Jansen
E. Taal
M.A.F.J. van de Laar

ABSTRACT

Objective The time trade-off is a health-related quality of life instrument that measures valuations for health states (utilities) by asking patients to value their health state anchored on a scale between death (0) and perfect health (1). Dying earlier is not perceived as a realistic worst-case consequence of the disease by RA patients. Of the previous focus groups study on RA patients, five worst-case future scenarios emerged. The aim of this study was to examine which potential worst-case scenario was the most appropriate for RA patients to use in utility calculation.

Methods In a cross-sectional study of 74 consecutive RA patients visiting the rheumatology outpatient clinic, participants were presented with descriptions of the five worst-case future scenarios. In pairwise comparisons, patients had to choose the scenario that would be the worst to experience. The worst-case future scenario was defined by the scenario that was chosen by a significantly greater proportion of participants than could be expected based on chance (20%). Therefore, analysis based on a single fraction (\hat{p}) was used and 95% CI was calculated.

Results The scenario being dependent on others was chosen most often as the worst to experience [by 35% of participants (95% CI 24%, 46%)] and significantly more often than could be expected based on chance ($\hat{p} = 0.35$, $z = 6.45$, $P = 0.00$).

Conclusion The scenario being dependent on others is likely to be the most appropriate worst-case future scenario for RA patients. Using an alternative anchor could improve the validity and responsiveness of the time trade-off in RA patients.



INTRODUCTION

RA is a chronic, autoimmune disease characterized by chronic symmetric polyarthritis. Most patients will have joint destruction, functional impairments and increased mortality [1]. RA can have a great impact on patients' health-related quality of life (HRQoL) (2,3). HRQoL can be defined as the extent to which physical, emotional and social well-being are impacted by the medical condition and/or its treatment (3). Fortunately, in the last century, developments in treatment have improved patients' clinical status dramatically (4,5) and also patients' HRQoL (6-8).

Various methods have been developed to measure HRQoL; either descriptive or valuation methods can be used. The first type of method is designed to give a detailed description of patients' health states (e.g. SF-36). The second type of method is designed to obtain valuations for patients' health states (utilities). Health states are valued on a scale anchored between death (0) and perfect health (1). Utilities can be obtained directly from patients or indirectly by attaching utility weights of the general public to health states of patients. Direct utility elicitation methods are instruments, such as the standard gamble (9,10) or time trade-off (TTO) (11).

The TTO is a frequently used method in case of direct utility elicitation. The TTO asks patients about the number of life years they are willing to trade for perfect health, i.e. to avoid being in their current health state. Thus, patients have to compare their current health state with the anchor state death. In this way, an estimation of the severity of the current health state is obtained. It is assumed that the more life years patients are willing to trade off, the worse their health state is. However, studies applying the TTO have shown that people do not always trade life years for perfect health (zero-traders) (12). A range in the proportions of zero-traders from 1.7% to 58% of participants has been found (12). Zero-traders can lead to ceiling effects, as they score the highest possible TTO score (1.00). Because of their unwillingness to trade any life years for perfect health, zero-traders give the impression that their health is perfect, leading to an overestimation of their HRQoL. As a consequence, the validity and responsiveness of the TTO is limited. A possible explanation is that giving up life years, i.e. dying earlier, is not perceived as a realistic consequence of their disease by patients with RA (13). Furthermore, patients have often adapted to their life with RA and have cognitively integrated the illness into their lives by redefining their internal standards of what comprises health, changing their values and priorities and/or redefining the concept of quality of life. Therefore, perfect health is not seen as an important or realistic goal. Replacing the current anchor death by an anchor that is perceived as a realistic consequence of RA from the perspective of patients may reduce the number of zero-traders, which could improve the validity and responsiveness of the TTO in patients with RA. The TTO utility scales are constructed for use in cost-utility analyses and for assistance in decisions about resource allocation. Changes in the validity, reliability and responsiveness of the utility scales might lead to changes in decisions about resource allocation.

To find alternative anchors for the TTO, we examined RA patients' future expectations and worst-case future scenarios in a previous focus group study (13). Concerns and worst-case future scenarios were related to dependence on others, worsening fatigue, inability to perform hobbies, inability to walk and increasing dependence on medication. All five worst-case future scenarios that were mentioned in the focus group study were used in the current study.

Before an alternative TTO can be developed, a quantitative study should be performed to validate the results of the focus group study. Therefore, the aim of this study was to examine which of the five worst-case future scenarios that emerged from the focus group study was the most appropriate worst-case future scenario for RA patients to be used as an anchor in the TTO to obtain utilities.



METHODS

Patients and study design

Consecutive patients diagnosed with RA (aged 18-75) visiting the rheumatology outpatient clinic of Medisch Spectrum Twente, Enschede, The Netherlands, participated in the study. Patients who could speak and read the Dutch language were handed an information letter and asked by the attending rheumatologist to participate. Patients who agreed to participate were introduced to the researcher (L.B.) who informed them about the study in more detail. The study was done within 2-4 weeks thereafter.

After informed consent was obtained, patients had to choose in pairwise comparisons the scenario that would be the worst to experience. Subsequently, they had to complete questionnaires about their current experience with the five worst-case future scenarios, about sociodemographic information, pain (numerical rating scale), functional status (HAQ consensus) and quality of life (SF-36: mental and physical component scales). The researcher was present (L.B.) to introduce the questionnaires. The study was approved by the Central Committee on Research Involving Human Subjects.

Measures

Pilot study

Previous to this study, a pilot study was conducted to assess whether the scenarios were conceivable, clear and comprehensive. One adaptation was made. The scenario that was labelled as no longer able to do any hobbies was renamed as no longer able to do any leisure activities in your free time. Participants remarked that the term hobbies inclined them to only think about specific activities like handicraft or collecting items, rather than about any activity that they might want to do in their free time.

Study

Worst-case future scenarios

Participants were presented with descriptions of the five worst-case future scenarios. In pairwise comparisons, they were asked to choose the scenario that would be the worst to experience. Each scenario was compared with all the other scenarios, resulting in 10 comparisons. An opt-out option allowed participants to indicate that they would not choose any scenario from the pair. Four random versions of the 10 comparisons were designed to rule out possible sequence effects. By means of a computer, four random sequences of the numbers 1-10 were obtained. In advance, participants were assigned to one of the four versions by using blockwise randomization. The five scenarios were being dependent on others, being extremely fatigued, being unable to perform leisure activities anymore, being unable to walk anymore and being dependent on medication.



Current experience of worst-case future scenarios

In five questions, participants had to indicate on a numerical scale from 0 (not experienced) to 10 (fully experienced) the extent to which they currently experienced being in the state as described in the five worst-case future scenarios.

Pain

Current severity of the pain was assessed by a numerical rating scale (NRS), ranging from 0 (best) to 10 (worst).

Functional status

The level of functional disability was assessed by the Dutch consensus HAQ (HAQ consensus) (14) based on the HAQ Disability Index (HAQ-DI) (15), a self-report measure consisting of eight categories (dressing and grooming, rising, eating, walking, hygiene, reach, grip and common daily activities). The HAQ-DI score ranges from 0 to 3, with higher scores indicating lower levels of functioning. The HAQ-DI has been proved valid (16,17).

HRQoL (physical and mental health)

Physical and mental health were assessed by calculating the physical component summary (PCS) and mental component summary (MCS) scores of the SF-36 version 2 (18,19), a generic descriptive instrument for measuring HRQoL on eight dimensions (mental functioning, physical functioning, bodily pain, vitality, role limitations because of physical problems, role limitations because of emotional problems, social functioning and general health). Evidence for the validity of the PCS and MCS of the SF-36 in RA patients has been found (20). The PCS and MCS scores range from 0 to 100, whereby a higher score indicates a better health.

Statistical analysis

Although our study has an explorative character, we performed a power analysis based on the main objective of the study. Using an alpha of 0.05 (two-tailed) and a power (1- β) of 80%, a sample of 72 participants is required to detect a difference of 15% between the proportion of participants who chose a certain scenario most often as the worst to experience and the proportion based on chance (expected proportion based on chance is 20%).

To calculate which of the five scenarios (A, B, C, D and E) was chosen most often as the worst to experience, a ranking of scenarios was calculated based on the 10 comparisons (AB, AC, AD, AE, BC, BD, BE, CD, CE and DE). Each scenario was compared with the other four scenarios. In each comparison, the scenario that was chosen as the worst to experience was given 1 point, the other scenario was given 0 points. Adding up the points for each scenario resulted in a score for each scenario ranging from 0 (never chosen as the worst to experience) to 4 (in all comparisons chosen as the worst to experience). The scenario with the highest score was considered to be the worst-case future scenario for the participant. If participants turned out to have two or more scenarios equally ranked as the worst to experience, we considered their answer on the comparisons of these equally ranked scenarios as their worst-case future scenario. The worst-case future scenario was defined by the scenario that was chosen by a significantly greater proportion of participants than could be expected based on chance (20%).



Therefore, analysis based on a single fraction (\hat{p}) was used. Furthermore, 95% CI for proportions were calculated.

In addition, it was examined whether the results of the single fraction analysis changed when participants who already experienced one or more of these states were left out of the analysis. Participants who scored > 8 on the question whether they currently experienced a certain state as described in the scenarios were considered to experience that state.

Furthermore, it was examined for the scenario that was chosen most often as the worst to experience, whether there were also participants who did not choose this scenario to be worse than any of the other scenarios. Data were analysed using PASW Statistics (version 18; SPSS, Inc., Chicago, IL, USA).

RESULTS

Subjects

Of the 84 participants who agreed to participate, 74 participants eventually participated. No significant differences in gender, disease activity and general health (NRS general health) were found between the included participants ($n=74$) and the non-included participants ($n=10$). A significant difference in age was found ($U = 185.00$, $z = -2.56$, $P < 0.01$) between the included participants {median age [interquartile range (IQR)] = 58 (52.42-66.08)} and non-included participants [median age (IQR) = 68 (64.12-72.46)]. Demographic and clinical characteristics of the 74 participants are reported in Table 1.

Worst-case future scenario

The worst-case future scenarios are shown in Figure 1. The total number of comparisons that participants completed was 740. In 53 of these comparisons (7.2%), no choice was made.

The scenario being dependent on others was chosen most often as the worst to experience by 35% of participants (95% CI 24%, 46%). This percentage was significantly higher than could be expected based on chance ($\hat{p}= 0.35$, $z = 3.26$, $P = 0.00$). The percentages of the latter four scenarios did not significantly deviate from chance (Table 2).

Median (IQR) scores regarding the experience of the scenarios are shown in Table 1. Interestingly, being dependent on others is less frequently experienced than the other four scenarios. Twelve participants (16%) experienced one or more of the worse states (experience score > 8) (Table 1). For the remaining 62 participants who did not experience any of the states described in the scenarios (experience score ≤ 8), the scenario being dependent on others was still chosen most often as the worst to experience (Table 2). The percentages of the other four scenarios did not significantly deviate from chance. Four participants (5.4% of the 74 participants) did not choose the scenario being dependent on others as the worst to experience in any of the 10 comparisons.

Table 1 Personal and disease-related characteristics of the 74 participants

		Total group (n=74)
Mean age (SD)		58 (8)
Gender, %	Men	38
	Women	62
Marital Status, %	Single	16
	Married/living together	84
Educational Level, %	Low	59
	Moderate	23
	High	18
Work status, %	Paid work	44
	Housekeeping	14
	Retired/unemployed/disabled	42
Medication, %	DMARD	70
	Biologic	30
Median disease duration (IQR), years		7.00 (3.75-13.00)
Mean disease activity (SD) (DAS28 1-10)		2.47 (1.02)
Median pain (IQR) (NRS)		3.00 (2.00-5.00)
Median HAQ-DI (IQR)		0.50 (0.14-1.00)
Mean physical health (SD) (PCS)		42.39 (8.59)
Mean mental health (SD) (MCS)		50.97 (10.85)
Median experience of being dependent on others (IQR) (0-10)^a		2.00 (0.00-4.00)
Median experience of being dependent on medication (IQR)^a		5.00 (3.00-8.00)
Median experience of being extremely fatigued (IQR)^a		4.00 (2.00-6.00)
Median experience of no longer being able to do any leisure activities in the free time (IQR)^a		3.00 (1.00-5.00)
Median experience of no longer being able to walk (IQR)^a		3.00 (1.00-5.00)
Participants (n) being in one or more of the worst health states (current experience > 8), %		12 (16)

^a A score of 0: the scenario has not (yet) been experienced; a score of 10: the scenario has been fully experienced as it has been described. DAS28: disease activity score in 28 joints.



Figure 1 Descriptions of the five worst-case future scenarios

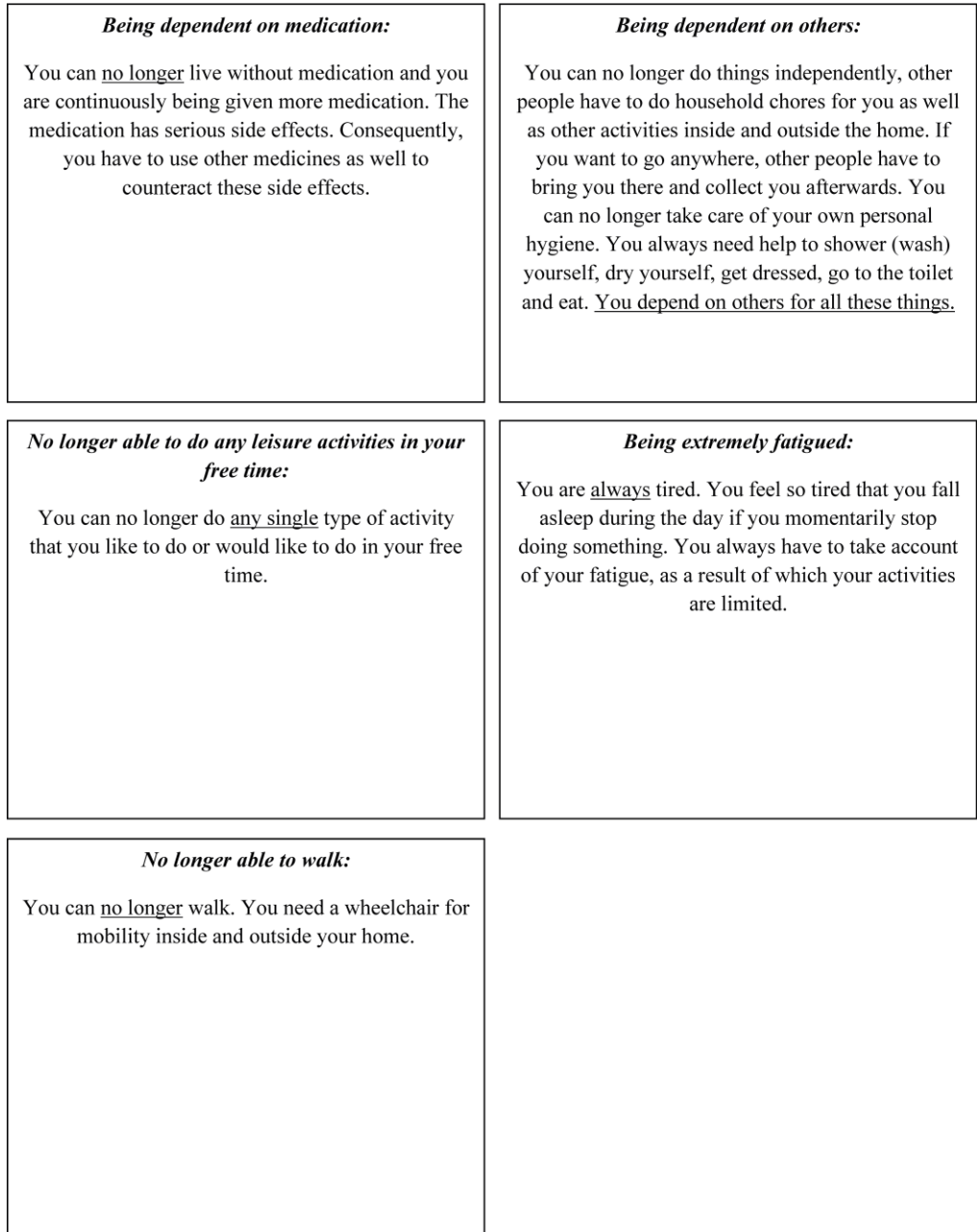


Table 2 Percentage of participants divided over every worst-case future scenario and significance tests for all participants ($n=74$) and for the participants ($n=62$) who were not considered to be in the worst state (experience score ≤ 8)

	Percentage of participants (95%CI) ($n=74$)	P [significantly different from chance (20%)]	Percentage of participants (95%CI) (experience score ≤ 8) ($n=62$)	P [significantly different from chance (20%)] (experience score ≤ 8)
Being dependent on others	35 (24, 46)	0.00	37 (25, 49)	0.01
No longer being able to walk	18 (9, 27)	0.61	15 (6, 24)	0.28
Being dependent on medication	12 (5, 19)	0.09	11 (3, 19)	0.09
No longer being able to do any leisure activities in your free time	11 (4, 18)	0.05	11 (3, 19)	0.09
Being extremely fatigued	12 (5, 19)	0.09	13 (5, 21)	0.16
Being indifferent	12 (5, 19)		13 (5, 21)	

DISCUSSION

This study of RA patients revealed that the scenario being dependent on others was most often indicated as the worst to experience. Moreover, it was the only scenario that was indicated significantly more often as the worst to experience than could be expected based on chance; even after repeating the analysis without patients who already experienced one or more of the worst states.

Although this study is the first to reveal RA patients' worst-case future scenarios, previous studies (21-24) have already demonstrated that dependency on others is one of the most important problems for RA patients, except for the study of Gossec et al. (25) that did not show that this was one of the most important health outcomes for RA patients. Functional limitations, mobility and fatigue were also mentioned within the top three of most important problems or outcomes, although not all these studies mentioned each of them. All these studies did have pain among the top three of most important problems or outcomes. Remarkably, in our previous focus group study, pain was not considered to be a worst-case future scenario, and therefore, we did not include pain in this study. A possible explanation is that we asked for future scenarios, whereas the previous studies asked for current problems or outcomes. Furthermore, in the focus groups some patients stated that they experienced pain, but compared their pain levels with pain levels at the time of diagnosis and indicated that the pain had



decreased tremendously. Also, participants indicated that they got used to the pain; that they had to cope with it. This may also explain why they did not mention pain as a worst-case future scenario.

When applying the scenario being dependent on others as a new anchor in the TTO, patients would be asked how many years in dependency they are willing to make a trade-off to become perfectly healthy. Therefore, people have to choose between a life in their current health state and a life in a better health state for a shorter number of years (being perfectly healthy) followed by a number of years in the worse health state (being dependent on others).

This study showed that none of the participants experienced this worst state of being dependent on others. Being in a state that is used in the trade of the TTO will bias the utility score downward, leading to floor effects that produce an underestimation in HRQoL; it is expected that these participants are, irrespective of their current health, willing to accept no years in perfect health, and thus, all years in a state they already experience. This will result in a utility score of 0, indicating a very bad current health state. However, as no participants have indicated to be in the state of being dependent on others, downward biases are not likely when being dependent on others is used as an anchor in the TTO.

Some participants in our study (5.4%) did not choose the scenario being dependent on others as the worst to experience in any of the 10 comparisons. For these participants, the sensitivity of a TTO with being dependent as an anchor state may be decreased; the utility score may be biased downwards.

Interestingly, the scenario being dependent on others was experienced least, although it was most often chosen as the worst to experience. It seems that patients who do not have much experience with a state find this state the worst to experience. This explanation is supported by research that has examined differences in values of health states between patients and members of the general public. Patients have been found to assign higher values to their health states than people of the general public who generally do not have any experience with these health states (26-28). Furthermore, research has shown that patients assign lower utility scores to states they do not experience than to states they do experience (29). Therefore, states that have not been experienced (hypothetical states) are worse for people than states that have been experienced. An explanation is that people undervalue their own ability to adapt (30,31).

When an adapted TTO with being dependent on others as the anchor state turns out to enhance the validity and responsiveness of the TTO, it could be used, for instance, to measure the effects of treatments at the group level from a patient perspective to determine the optimal treatment for groups of RA patients.

It has to be remarked that, although the scenario being dependent on others was chosen most often as the worst to experience by RA patients, 65% of participants had not chosen this scenario most often as the worst to experience. The other four scenarios were also chosen by 12-18% of the participants as worst to experience. In addition, 12% of participants were indifferent. Other studies also showed that, although dependency on others was one of the most important problems for RA patients, it was not selected by the majority of patients in the top three of most important outcomes (21,22). Hewlett et al. (21) concluded from this finding that no outcome is universally important. For patients who do not consider being dependent on others as their worst-case future scenario, the use of it as an anchor in the TTO can be problematic. Therefore, the scenario should be compared with the current scenario death to examine how much of a problem this is. Nevertheless, it should also be noted that even when dying earlier turns out to be the worst-case future scenario for RA patients, it is probably not perceived as a realistic consequence of RA, as it was not mentioned in the focus group study to be a concern and/or worst-case future scenario. The sample of RA patients used in this study seems to be a representative sample of the Dutch RA population in usual care. The short disease duration, low disease activity scores (DAS28) and low pain scores found in our sample were comparable with the clinical characteristics of a Dutch sample of consecutive patients seen in standard rheumatology care that was part of the



Quantitative Patient Questionnaires in Standard Monitoring of Patients with Rheumatoid Arthritis (QUEST-RA) study (32).

This study is the first to reveal that being dependent on others was the worst-case future scenario for RA patients. A further step is to examine the psychometric properties of the TTO with being dependent on others as an alternative to the current anchor, death.

ACKNOWLEDGEMENTS

The authors would like to thank the participants and the Rheumatology Department of Medisch Spectrum Twente, Enschede, The Netherlands, for their contribution to this study.



REFERENCES

- 1 Combe B. Progression in early rheumatoid arthritis. *Best Pract Res Clin Rheumatol* 2009;23:59-69.
- 2 Khanna D, Maranian P, Palta M, Kaplan R, Hays R, Cherepanov D, et al. Health-related quality of life in adults reporting arthritis: analysis from the National Health Measurement Study. *Qual Life Res* 2011;20:1131-40.
- 3 Strand V, Khanna D. The impact of rheumatoid arthritis and treatment on patients' lives. *Clin Exp Rheumatol* 2010;28(Suppl. 59):S32-S40.
- 4 Vermeer M, Kuper HH, Hoekstra M, Haagsma CJ, Posthumus MD, Brus HLM, et al. Implementation of a treat-to-target strategy in very early rheumatoid arthritis: results of the Dutch Rheumatoid Arthritis Monitoring remission induction cohort study. *Arthritis Rheum* 2011;63:2865-72.
- 5 Villeneuve E, Emery P. Rheumatoid arthritis: what has changed? *Skeletal Radiol* 2009;38:109-12.
- 6 Strand V, Mease P, Burmester G, Nikai E, Coteur G, van Vollenhoven R, et al. Rapid and sustained improvements in health-related quality of life, fatigue, and other patient-reported outcomes in rheumatoid arthritis patients treated with certolizumab pegol plus methotrexate over 1 year: results from the RAPID 1 randomized controlled trial. *Arthritis Res Ther* 2009;11:R170.
- 7 Klarenbeek NB, Güler-Yüksel M, van der Kooij SM, Han KH, Runday HK, Kerstens PJSM, et al. The impact of four dynamic, goal-steered treatment strategies on the 5-year outcomes of rheumatoid arthritis patients in the BeSt study. *Ann Rheum Dis* 2011;70: 1039-46.
- 8 Curtis JR, Singh JA. Use of biologics in rheumatoid arthritis: current and emerging paradigms of care. *Clin Ther* 2011;33:679-707.
- 9 Von Neumann J, Morgenstern O. *Theory of games and economic behavior*. New York: Wiley; 1953.
- 10 Torrance GW. Toward a utility theory foundation for health status index models. *Health Serv Res* 1976;11:349-69.
- 11 Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7:118-33.
- 12 Arnesen T, Trommald M. Are QALYs based on time tradeoff comparable? - A systematic review of TTO methodologies. *Health Econ* 2005;14:39-53.
- 13 Buitinga L, Braakman-Jansen LM, Taal E, van de Laar MA. Future expectations and worst-case future scenarios of patients with rheumatoid arthritis: a focus group study. *Musculoskeletal Care* 2012;20 June [Epub ahead of print].
- 14 Boers M, Jacobs JW, van Vliet Vlieland TP, van Riel PL. Consensus Dutch Health Assessment Questionnaire. *Ann Rheum Dis* 2007;66:132-3.
- 15 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
- 16 Bruce B, Fries J. The Stanford health assessment questionnaire: dimensions and practical applications. *Health Qual Life Outcomes* 2003;1:20.
- 17 Ten Klooster PM, Taal E, Van De Laar MA. Rasch analysis of the Dutch Health Assessment Questionnaire Disability Index and the Health Assessment Questionnaire II in patients with rheumatoid arthritis. *Arthritis Care Res* 2008;59:1721-8.
- 18 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 19 Ware JE Jr. SF-36 health survey update. *Spine* 2000;25:3130-9.
- 20 Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). *Br J Rheumatol* 1998;37:425-36.
- 21 Hewlett S, Carr M, Ryan S, Kirwan J, Richards P, Carr A, et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care* 2005;3:131-42.
- 22 Taal E, Rasker JJ, Seydel ER, Wiegman O. Health status, adherence with health recommendations, self-efficacy and social support in patients with rheumatoid arthritis. *Patient Educ Couns* 1993;20:63-76.
- 23 Van Lankveld W, Naring G, Van der Staak C, Van 't Pad Bosch P, Van de Putte L. Stress caused by rheumatoid arthritis: relation among subjective stressors of the disease, disease status, and well-being. *J Behav Med* 1993;16:309-21.

- 24 Melanson PM, Downe-Wamboldt B. Confronting life with rheumatoid arthritis. *J Adv Nurs* 2003;42:125-33.
- 25 Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadelo S, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680-5.
- 26 Ratcliffe J, Brazier J, Palfreyman S, Michaels J. A comparison of patient and population values for health states in varicose veins patients. *Health Econ* 2007;16:395-405.
- 27 Ariza-Ariza R, Hernández-Cruz B, Carmona L, Ruiz-Montesinos MD, Ballina J, Navarro-Sarabia F. Assessing utility values in rheumatoid arthritis: a comparison between Time Trade-Off and the EuroQol. *Arthritis Care Res* 2006;55:751-6.
- 28 Happich M, Von Lengerke T. Valuing the health state 'tinnitus': differences between patients and the general public. *Hearing Res* 2005;207:50-8.
- 29 De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* 2000;9:109-26.
- 30 Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychol Sci* 1998;9:340-6.
- 31 Dolan P. Developing methods that really do value the 'Q' in the QALY. *Health Econ Policy Law* 2008;3: 69-77.
- 32 Sokka T, Kautiainen H, Toloza S, Mäkinen H, Verstappen SMM, Lund Hetland M, et al. QUEST-RA: quantitative clinical assessment of patients with rheumatoid arthritis seen in standard rheumatology care in 15 countries. *Ann Rheum Dis* 2007;66:1491-6.



Chapter

Chapter summary & discussion

Chapter summary

Advances in the treatment and management of RA have led to significant improvements in patients' symptoms and outcomes. Currently, the clinical status of a significant proportion of patients can be elevated to a state of remission in which there is no apparent disease activity and progression of joint damage (1,2). Since the recognition that not only objective clinical measures are important but also measures that incorporate the patient's perspective, the number of PROs has expanded including measures of HRQoL. The increased use of PROs in different disciplines necessitates the thorough examination of their psychometric properties. The expanding use of computers to conduct PROs warrants the comparison of outcomes obtained by computer with those obtained by the traditionally administered interview or questionnaire. PROs such as the HAQ and SF-36, also referred to as self-reported health outcome measures, have been validated extensively in RA. With respect to HRQoL instruments used in economic analysis, the EQ-5D and SF-6D are commonly used because they represent a societal perspective. Extensive research has addressed the psychometric performance of the EQ-5D and, to a lesser extent, that of the SF-6D. Nevertheless, further research is required into the responsiveness of these instruments to improvement. As research into treatment of RA has evolved rapidly, leading to effective therapies, HRQoL instruments should be able to reflect changes in patients' health status. Studies of different diseases have shown that these instruments have limitations in their discriminative ability because of, for example, floor or ceiling effects in utility scores. Less research has been performed in RA regarding instruments that directly obtain valuations of health states from patients, i.e. the TTO and the SG. Because patients have a unique and full knowledge of, and experience with, their own health condition and are directly affected by interventions, we believe that it is important to obtain patients' valuations. More research is needed to examine the psychometric properties of these direct utility instruments in RA.

In **chapter 3**, the responsiveness of the EQ-5D was compared with the responsiveness of the SF-6D in RA patients with active disease treated with effective TNF-blockers. Both instruments were only moderately responsive to improvement over a one-year period, and the SF-6D was generally more responsive than the EQ-5D. Neither systematic floor nor ceiling effects were found in the utility scores that could have influenced the potential of the instruments to detect change. Nevertheless, the already high floor of the SF-6D (utility score of 0.30) limits the use of the instrument in patients with more severe RA. It is important to note that, although the SF-6D was more responsive than the EQ-5D, its use in cost-utility analysis will result in a less favorable cost-utility ratio because mean changes in utility scores are used to calculate the number of QALYs gained by a novel health technology. The mean change in the EQ-5D utility score is larger than the mean change in the SF-6D utility score. This underlines the larger variance around the EQ-5D, providing a lower responsiveness statistic. Using the SF-6D, however, would result in less uncertainty regarding the relative cost-effectiveness of two treatments.

In **chapter 4**, a computer TTO was developed and compared with an interview TTO regarding feasibility and level of agreement. The results showed that the computer TTO was a feasible and reliable alternative for the interview TTO. Both TTO instruments, however, did not give similar results. A clinically relevant median difference between both TTOs was found, implying that utilities obtained with both TTOs are not interchangeable. Fortunately, no systematic biases in the differences were found over the range of scores. Some adaptations should be made to ensure the computer TTO can be administered without assistance. Further research in a larger sample of RA patients is needed to verify the results.



In **chapter 5**, the construct validity of the TTO was examined in a small sample of RA patients. The preliminary results showed that the computer and interview TTO were poorly correlated with, and poorly able to discriminate between, self-reported health outcome measures. After the exclusion of people who refused to trade any life years ('zero-traders'), the results improved with respect to convergent validity. Possible explanations are the fact that most self-reported outcome measures only measure one aspect of the construct quality of life and are only of a descriptive nature. In addition, the TTO has been found to be prone to biases. For example, because of the fact that people are loss averse, they are less willing to trade off life years, resulting in upward biases in the utility scores. Another important explanation could be the fact that RA patients are less willing to trade any of their life years in return for perfect health because they do not perceive their disease as life-threatening. Possibly, there are more realistic health-related anchors for RA patients. To improve the current TTO, the use of anchors perceived as realistic by RA patients could be evaluated or a procedure called 'chaining' could be used, although a chained TTO is more complex and has not been validated widely in chronic states.

In **chapter 6**, RA patients' future expectations and worst-case future scenarios were examined by means of focus groups. Insight into patients' expectations and worst-case future scenarios related to their RA could be used in interventions to place them in a realistic perspective and to help patients to cope with their fears and concerns in order to improve HRQoL. Worst-case future scenarios that are perceived as realistic by RA patients could be used as an alternative for the anchor 'death' in the TTO, possibly improving the validity and responsiveness of the instrument. The results showed that expectations and worst-case future scenarios were mentioned in the following areas: dependency on others, worsening fatigue, inability to perform hobbies, the inability to walk, and increasing dependency on medication. Although not all expectations and worst-case scenarios are realistic from a clinical point of view, for patients they are realistic. Remarkably, pain was not mentioned as either a concern or problem, in contrast to some other studies. Differences between studies can be explained by the rapid advances in the treatment of RA and by the differences in the questions being asked. In addition, although patients did experience pain, they compared their pain levels with pain levels during the time of diagnosis and indicated that the pain had decreased tremendously. Participants also indicated that they got used to the pain; that they had to cope with it. This may also explain why they did not mention pain as a worst-case future scenario. A larger quantitative study of RA patients is needed to validate the results.

In **chapter 7**, a quantitative study was performed to examine which of the five worst-case future scenarios that emerged from the focus group study was the most appropriate worst-case future scenario for RA patients to be used as an anchor in the TTO to obtain utilities. The results showed that the scenario 'being dependent on others' was the most appropriate scenario, because it was most often indicated as the worst to experience. None of the participants experienced this worst state scenario already. Being in a state used in the TTO will bias utility scores downwards, leading to floor effects that produce an underestimation in HRQoL; people who already experience a state used in the TTO will automatically receive a utility score of 0. For these people it is not possible to do the trade. Interestingly, the scenario 'being dependent on others' was experienced least, although it was most often chosen as the worst to experience. It seems that patients who do not have much experience with a state find this state the worst to experience. It has to be remarked that, although the scenario 'being dependent on others' was chosen most often as the worst to experience by RA patients, 65% of participants had not chosen this scenario most often as the worst to experience. For these 65% of patients, the use of dependency on others as an anchor in the TTO can be problematic. Therefore, the scenario should be compared with the current scenario 'death' to examine how much of a problem this is. Nevertheless, it should also be noted



that even when ‘dying earlier’ turns out to be the worst-case future scenario for RA patients, this anchor is probably not perceived as a realistic consequence of RA, since it was not mentioned in the focus group study to be a concern and/or worst-case future scenario.

General discussion

HRQoL is a complex construct that misses a clear conceptualization. Although at a conceptual level the term is often considered to include valuations or judgments regarding domains of physical, psychological and social functioning (3), in practice the term is mainly used to denote descriptive instruments that give an indication of patients’ level of functioning in various domains.

The measurement of HRQoL is important for assessing the impact of diseases and the effects of treatments. HRQoL is a critical outcome in CUAs where the costs and effects of different treatments and in different diseases are compared. A problem for CUAs is that results from different studies are often hard to compare because various instruments are used to measure utilities that do not generate similar results. In addition, different methodological variants of instruments are applied, which further limits comparability. It is important that utility instruments capture the full range of disease severity and that they are psychometrically sound. This thesis has shown that direct and indirect utility instruments have limitations regarding their psychometric properties in patients with RA, and that the data collection strategy may influence results. In addition, alternative anchors have been examined that could improve the measurement of HRQoL that represents the patient perspective.

The societal versus patient perspective

An important issue for debate in the field of health economics concerns the perspective for generating valuations of health states; should we ask those patients who are personally affected with that particular health condition in question or members of society in general? The use of each of the perspectives is supported by plausible arguments. Both perspectives, however, entail problems that restrict the accurate measurement of the impact of the disease on HRQoL and the impact of interventions on changes in HRQoL. Patients seem to underestimate the losses associated with their disease whereas members of the general public seem to overestimate these losses (4). The use of the societal perspective is currently the norm in CUAs.

Improving HRQoL measurement that represents the societal perspective with item response theory

In this thesis, we showed with respect to instruments that reveal a societal perspective, namely the EQ-5D and SF-6D, that they are only moderately responsive to improvement in RA patients with active disease receiving very effective treatment. Other studies (5-8) have shown that, although evidence was found for their reliability and validity, the EQ-5D and SF-6D have limitations with providing similar results and with adequately reflecting disease severity. The study of Brazier et al. (6) showed, for example, that the majority of participants with a utility score of 1.00 on the EQ-5D did not have this perfect health score on the SF-6D. As a matter of fact, a wide range in SF-6D utility scores was observed. The finding of floor effects in the SF-6D (many participants scored the lowest possible score on the SF-6D) and ceiling effects in the EQ-5D (many participants scored the highest possible score on the EQ-5D) suggests that these instruments are not good at discriminating among people who are in poor or better health. These floor and ceiling effects were not observed in studies of RA and in our study reported in chapter 3. Nevertheless, the SF-6D has a high floor, whereby 0.30 is the worst possible utility score.



An explanation for all these findings might be that important dimensions are missing in these instruments and/or that the number of the response levels is too limited. If instruments lack dimensions that have a significant impact on patients' health state, utility scores will be distorted. Some studies have examined the expansion of dimensions of the EQ-5D, such as adding 'cognitive functioning' (EQ-5D+C) (9) or a set of five extra dimensions (sleep, memory and concentration, energy and fatigue, sight and hearing and contact with others) (10). Krabbe et al. (9) showed a better content validity of the EQ-5D+C and a similar reliability as the original EQ-5D, although an increase in dimensions is expected to increase the variability in responses and decrease the reliability. According to Krabbe et al., it is not likely that the EQ-5D+C will benefit from a further extension of dimensions, but they note that it should be explored whether improvement can be obtained within dimensions, for instance by increasing the number of response levels. Perneger and Courvoisier (10) showed that the extended ten-item version of the EQ-5D explained a larger proportion of variance in self-perceived health (VAS) than the original EQ-5D, and a decreased ceiling effect. However, an increase in the number of dimensions for the EQ-5D exponentially increases the possible number of health states that have to be valued (10).

With respect to including more response levels, a study of Pickard et al. (11) showed that, when comparing a five-level EQ-5D with the three-level EQ-5D using Rasch analysis, the EQ-5D-5L has a greater descriptive richness whereby less ceiling effect and a better discriminative ability compared with the original EQ-5D were found. Other studies in cancer patients (12) and in the population at large (13) also showed a greater discriminatory power in terms of absolute informativity of the EQ-5D-5L and less ceiling effect than the original EQ-5D (EQ-5D-3L). Another general population study (14) gave evidence for the increased descriptive power and a broadened measurement continuum of the EQ-5D-5L compared with the EQ-5D-3L. Thus, expanding the response levels improves the measurement properties of the instrument. A disadvantage of this increase in the number of response levels is that a much larger number of health states needs to be valued for generating value sets. More research is needed to examine whether the adapted EQ-5D has better psychometric properties than the original EQ-5D.

A promising approach for improving instruments so that they capture the full range of disease severity and are psychometrically sound without suffering from floor and ceiling effects is the use of modern psychometric techniques, such as item response theory (IRT). With IRT it is possible to construct a calibrated item bank. From such an item bank, short forms or computerized adaptive testing protocols (CAT) can be developed. The latter methods of assessment help to ensure that patients only respond to questions that are relevant to their specific quality of life level, and that only minimal questions need to be answered by patients, while retaining or surpassing the measurement precision of classical instruments. In addition, IRT can be used to examine whether the number and content of items are sufficient to comprise the underlying domain continuum, and whether each scale is unidimensional, which is an important assumption of IRT (15). An important advantage of IRT over classical test theory (CTT) is the fact that item properties (discrimination and difficulty) are invariant with respect to the sample used, i.e. they remain stable when administering items to different samples of the same population (16). Pickard et al. (11) who examined the expansion of response levels of the EQ-5D domains from three to five, used Rasch analysis, and they were able to identify the item levels on the five-level and three-level version that were statistically equivalent. This linking of item levels facilitates the estimation of societal preference weights for the five-level EQ-5D from the three level version. In addition, they were able to confirm the expected broader coverage of the underlying health continuum of each domain. More research is desirable using IRT modeling, including research that uses IRT to improve the SF-6D. With IRT it is possible to examine the information value of each item of the SF-6D and each response level to the underlying latent continuum (for the SF-6D: pain, mental health, physical functioning, social functioning, role limitations and vitality). Item information curves can show how

broad the coverage of the continuum by an item is, and how much precision the item adds to the measurement of the underlying latent continuum, i.e. how reliable an item is for measuring the continuum. These item information functions of the items of the SF-6D can be summed up to represent the information value of each SF-6D dimension. Higher information indicates more precision (or reliability). Thus, with IRT it can be examined whether the appropriate (most informative) items were selected (of the SF-36) to form the SF-6D. In this way, the SF-6D could be improved whereby the high floor of the SF-6D could be lowered.

Improving HRQoL measurement that represents the societal perspective with discrete choice experiments

Because of different concerns related to the current TTO procedure, e.g. regarding the complexity of the task for some respondents resulting in response inconsistencies and subsequent data exclusions which limit the representativeness of utility values (17), research has been carried out into the use of discrete choice experiments (DCEs). The DCE has been argued to be a simpler elicitation method because people only have to choose out of two or more options (descriptions of health profiles based on different attributes) which option they prefer. A problem of DCE is the fact that the data cannot be incorporated directly into QALY calculations, because the data are not anchored on the conventional perfect health-death (0-1) scale. Research has been carried out that anchors DCE values on health state utilities (18,19), but the method they used to do this has received criticism, as length of life is excluded from the valuation task (20). Bansback et al. (17) have developed a novel DCE method (DCE_{TTO}) that resembles the time trade-off task to estimate values on the health utility scale for the EQ-5D. The DCE_{TTO} was able to reduce the number of data exclusions, leading to more representative values. More research should be carried out to examine the value of this novel DCE method as an alternative way of obtaining preferences.

Measuring HRQoL from the patient perspective: influence of the data collection strategy

Different methodologies can be used to administer utility instruments. With respect to direct utility instruments, all sorts of variants are possible. Studies have shown that there were differences within the TTO methodology (21) and SG methodology (22-24) and that these differences yielded diverse results. A review of Arnesen et al. (21) showed that median utility scores varied substantially between studies with similar characteristics. Arnesen et al. (21) note that it is unclear whether differences between instruments are caused by differences in the TTO methodology or by differences in the characteristics of the population in which the TTO is applied. To be able to compare TTO utility scores between studies, it is important that only one version of the TTO is used rather than the different possible variants that are currently applied. The preliminary results of chapter 4 show that a computer TTO is a reliable and feasible alternative for the interview TTO in RA patients, although they cannot be used interchangeably. Because all RA patients completed both the interview and the computer TTO, the sample characteristics were similar for both TTO versions. Therefore, differences can be addressed to variations in the data collection strategy. Nevertheless, more research is needed that compares different variants of the TTO in the same patient group.



Improving HRQoL measurement that represents the patient perspective: an alternative anchor for the TTO

In chapter 5, it was shown that both the interview and the computer TTO had poor construct validity. Nevertheless, it has to be noted that the reference measures used were not preference-based. Validating the TTO against individualized measures of personal preferences, such as the SEIQoL or MACTAR instrument, might reveal better results. In this thesis we examined possible alternative anchors that could be used to replace the anchor 'death' by an anchor that might be perceived as more realistic by RA patients. Future research should be conducted to examine whether the validity and responsiveness of the TTO improves if such an alternative anchor is used. It should be noted, however, that not all problems with the TTO will be resolved. The alternative TTO will still not satisfy all QALY assumptions, such as a constant proportional time trade-off (CPTTO), which means that people do not trade off the same proportional amount of time independently of the absolute duration in the TTO task (25). When applying CPTTO to the TTO with an alternative anchor, then, under CPTTO it is assumed that someone who is willing to live five years out of 10 years in, for example, a state of dependency on others, is also willing to live 10 years out of 20 years in a state of dependency on others. Although some authors showed for the traditional TTO that people satisfy a constant proportional time trade-off, at least at the aggregate level (25,26), others showed that people do not satisfy a constant proportional time trade-off (27,28). In addition, the alternative TTO will still violate EUT under which people are expected to behave rationally, because of e.g. loss aversion (see theoretical background in chapter 2). Although the TTO does not imply the loss of life years when using an RA-related anchor, such as the anchor 'being dependent on others' revealed in chapters 6 and 7, people are still confronted with a loss, thus there will remain some level of loss aversion present during the TTO completion. Besides, as TTO utilities can be influenced by a lot of different factors, it might not be possible to take all these factors into account (21). The question is, therefore, how far these problems will influence the validity and responsiveness of the adapted TTO. In addition, because of the problems with measuring utilities, the question may also be whether we should measure valuations of health states using methods such as the TTO or SG. Irrespective of whose perspective is used to obtain valuations, utilities are biased estimates (4,29). During the valuation task, people have to be capable of predicting how they will feel in the future (so-called 'affective forecasting') (29). Evidence exists that people fail to predict their future well-being (4,29,30). Although Dolan and Kahneman (4) developed a method to capture actual experiences rather than future preferences, it is a time-consuming task for participants.

Improving HRQoL measurement that represents the patient perspective with item response theory

IRT could also be used to improve the HRQoL measurement that represents the patient perspective. Recent work of Krabbe (31) has concentrated on the combination of discrete choice (DC) models with basic features of the Rasch model. By using a multi-attribute response model based on a dichotomous question that asks patients whether their own health state is worse or better than other health states, health states can be calibrated on a scale from worst to best and subsequently anchored between 0 and 1. If certain health states are valued differently between groups of patients or between patients and healthy people, IRT can be applied to analyze this differential item functioning (DIF). If needed, group-specific weights can be attached to the values of health states. As a result, the valuation of health states will become comparable between groups of patients with different diseases and between patients and healthy persons. Eventually this may lead to an integration of the patient and societal perspective on the valuation of health states.



Strengths and limitations of this thesis

This thesis has paid attention to HRQoL valuated in the form of utilities as used for the economic analysis of health technologies. Evidence from this thesis indicates that utilities obtained using the current measurement techniques are limited from both a scientific and psychometric perspective as we have demonstrated in RA patients. Since these instruments provide data that have an important role in economic evaluations of health technologies to support reimbursement decisions, it is essential that these instruments are responsive and generate valid data. Therefore, improved versions of the instruments are warranted. This thesis is the first to compare an interview TTO with a similarly formatted computer TTO with respect to feasibility and agreement. Because of the fact that both TTOs were formatted in a similar way, differences could be addressed to the data collection strategy. The use of a cross-over design had the advantage of minimizing the confounding factor of inter-individual variation. Both the interview and computer TTO were validated whereby a range of important self-reported outcome measures was used. In addition, no previous study has asked RA patients explicitly about their future expectations and worst-case future scenarios, and alternative anchors for the TTO in RA have not been examined so far.

Several limitations must be mentioned here. First, the data of both TTO studies came from a convenience sample which limited the generalizability of the results. It has to be noted, however, that other studies in RA samples (32-34) have also shown the poor construct validity of an interview TTO. Second, the TTO was mainly validated against reference measures of a descriptive nature rather than against preference-based measures, which could have explained the poor results. Validating the TTO against individualized measures of personal preferences, such as the SEIQoL or MACTAR instrument, might reveal better results. Third, the studies in chapters 6 and 7 that examined worst-case future scenarios related to RA have a disease-specific character. Results cannot be generalized and projected onto other disease populations. One implication of this is that this alternative TTO cannot be directly applied in CUAs that compare different diseases, because instruments need to be generic. Anyhow, this TTO might be suitable for measuring treatment effects at the group level from a patient perspective in order to determine the optimal treatment for groups of RA patients. Fourth, in the last study, chapter 7, the current TTO anchor 'death' was not included in the comparisons of scenarios. Therefore, how much of a problem the current anchor 'death' really is for RA patients remains unknown. As 65% of participants would not choose 'being dependent on others' as their worst-case future scenario, it would be unwise to exchange the current anchor 'death' for this anchor in case less than 65% would not choose 'dying earlier' as their worst-case future scenario. Nevertheless, even if the current anchor 'death' turns out to be the worst-case future scenario for RA patients, it is probably not perceived as a realistic consequence of RA, as it was not mentioned in the focus group study to be a concern and/or worst-case future scenario.

The way forward

Based on this thesis, the following steps are recommended as the way forward. More research should be carried out into the validity of the TTO in RA patients using different methodological variants. It is important to include individualized measures of personal preferences as reference measures rather than using only descriptive measures that assess patients' level of functioning in various domains. Future research should examine how much of a problem the current TTO anchor 'death' really is for RA patients to determine the relevance of using an alternative anchor. However, even if 'dying earlier' is the worst-case future scenario for patients it might not be perceived as a realistic consequence of RA, so this should be examined to be able to verify this expectation. Subsequently, future research in representative



and sufficiently sized RA samples should examine whether the adapted TTO with the anchor 'being dependent on others' induces fewer zero-traders and has a better validity and responsiveness than the traditional TTO. Therefore, a cross-over design could be used to administer both TTOs in the same patients. Research could also examine alternative anchors for the TTO in other chronic diseases. Furthermore, the possibilities with DCEs and IRT should be explored. Promising results were found concerning the decrease in ceiling effect and improvement in discriminative ability of the extended five-level EQ-5D (EQ-5D-5L) when comparing this EQ-5D-5L with the traditional three-level EQ-5D using Rasch analysis (11). In addition, DC models and IRT can also improve the measurement of preferences of patients, but research is required to examine this.

Conclusion

This thesis has shown that important HRQoL instruments have problems with their psychometric properties. To improve the psychometric properties of the TTO in RA patients, this thesis has suggested replacing the current worst-case anchor 'death' by an anchor that is perceived as more realistic by RA patients. The results indicate that 'being dependent on others' is the most appropriate anchor for RA patients to use in the TTO. Further research should reveal whether this alternative TTO has indeed better psychometric properties than the traditional TTO.



REFERENCES

1. Villeneuve E, Emery P. Rheumatoid arthritis: What has changed? *Skeletal Radiol* 2009;38:109-12.
2. Vermeer M, Kuper HH, Hoekstra M, Haagsma CJ, Posthumus MD, Brus HLM, et al. Implementation of a treat to target strategy in very early rheumatoid arthritis: results of the DREAM remission induction cohort. *Arthritis Rheum* 2011;63:2865-72.
3. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000;9:887-900.
4. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008;118:215-34.
5. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* 2003;56:317-25.
6. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873-84.
7. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DPM. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *J Rheumatol* 2008;35:592-602.
8. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Verstappen SMM, Watson K, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Qual Life Res* 2009;18:1195-1205.
9. Krabbe PFM, Stouthard MEA, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol* 1999;52:293-301.
10. Perneger TV, Courvoisier DS. Exploration of health dimensions to be included in multi-attribute health-utility assessment. *Int J Qual Health Care* 2011;23:52-9.
11. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Med Care* 2007;45:259-63.
12. Kim SH, Kim HJ, Lee SI, Jo MW. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. *Qual Life Res* 2012;21:1065-73.
13. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health* 2008;11:275-84.
14. Janssen MF, Birnie E, Bonsel GJ. Quantification of the level descriptors for the standard EQ-5D three-level system and a five-level version according to two methods. *Qual Life Res* 2008;17:463-73.
15. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value Health* 2009;12(Suppl. 1):S18-S26.
16. Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, editors. *Assessing Quality of Life in Clinical Trials: Methods of Practice*. Oxford: Oxford University Press;2005:55-73.
17. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31:306-18.
18. Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health* 2010;13:1005-13.
19. Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *Eur J Health Econ* 2011. Article first published online: 30 September 2011. Doi: 10.1007/s10198-011-0352-9.
20. Flynn TN, Louviere JJ, Marley AAJ, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Popul Health Metr* 2008;6:6.
21. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Econ* 2005;14:39-53.
22. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;15:209-31.
23. Hammerschmidt T, Zeitler HP, Gulich M, Leidl R. A comparison of different strategies to collect standard gamble utilities. *Med Decis Making* 2004;24:493.



24. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitations. *Med Decis Making* 1998;18:76-83.
25. Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decis Making* 2005;25:460-7.
26. Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J Health Econ* 2003;22:445-58.
27. Buckingham JK, Birdsall J, Douglas JG. Comparing three versions of the time tradeoff: time for a change? *Med Decis Making* 1996;16:335-47.
28. Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Habbema JDF, De Haes JCJM. The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *J Clin Epidemiol* 1995;48:1207-14.
29. Kahneman D, Sugden R. Experienced utility as a standard of policy evaluation. *Environ Resour Econ* 2005;32:161-81.
30. Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychol Sci* 1998;9:340-6.
31. Krabbe, PFM. Towards a generalized HRQoL measurement model. *Presented at the University of Twente, department of Health Technology and Services Research*. Enschede, The Netherlands, 2012.
32. Tjihuis 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.
33. 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.
34. Witney AG, Treharne GJ, Tavakoli M, Lyons AC, Vincent K, Scott DL, et al. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology* 2006;45:975-81.



Samenvatting
(Dutch summary)

De snelle vooruitgang in behandeling van reumatoïde artritis (RA) heeft geleid tot een aanzienlijke verbetering in de symptomen van patiënten en uitkomsten. Tegenwoordig is het zelfs zo dat een groot deel van de patiënten in een staat van remissie kan worden gebracht, wat betekent dat er bij deze patiënten geen aanwijsbare ziekteactiviteit en achteruitgang in gewrichtsbeschadiging worden waargenomen. Voor het meten van de klinische toestand van patiënten worden zowel objectieve als subjectieve maten gebruikt. Met name in de afgelopen jaren is het belang van subjectieve, patiëntgerapporteerde uitkomstmaten erkend. Hierdoor is het aantal patiëntgerapporteerde uitkomstmaten aanzienlijk toegenomen, waaronder instrumenten die gezondheidsgelateerde kwaliteit van leven (KvL) meten. Vanwege het toegenomen gebruik van deze maten in verschillende disciplines is het van belang om de psychometrische eigenschappen van deze instrumenten grondig te evalueren. Bij de afname van instrumenten worden steeds vaker computers gebruikt. Het is daarom ook belangrijk om resultaten die zijn verkregen met een computer te vergelijken met de resultaten die zijn verkregen met een interview of vragenlijst. Voorbeelden van patiëntgerapporteerde uitkomstmaten zijn de zogenoemde 'Health Assessment Questionnaire' (HAQ), een maat voor functionele status, en de 'Medical Outcomes Study Short Form-36' (SF-36), een maat voor KvL, bestaande uit fysiek, psychologisch en sociaal functioneren. Deze instrumenten zijn uitgebreid gevalideerd bij RA-patiënten. Gezondheidsgelateerde KvL-maten worden ook gebruikt als uitkomstmaten voor economische analyses ter ondersteuning van vergoedingsbeslissingen. De voorkeur hierbij gaat uit naar instrumenten die waarderingen voor gezondheidstoestanden ('utiliteiten') meten vanuit het maatschappelijk perspectief, zoals de EuroQol-5D (EQ-5D) en de Short Form-6D (SF-6D). Er is uitgebreid onderzoek gedaan naar de psychometrische prestaties van de EQ-5D en in mindere mate ook naar die van de SF-6D. Toch is aanvullend onderzoek nodig naar de responsiviteit van deze instrumenten voor verbeteringen in de gezondheid van patiënten. Aangezien het onderzoek naar effectieve behandelingen voor RA zich snel heeft ontwikkeld, is het belangrijk dat gezondheidsgelateerde KvL-instrumenten veranderingen in de gezondheidstoestand van patiënten als gevolg van deze behandelingen kunnen weergeven.

Studies in verschillende ziekten hebben laten zien dat deze instrumenten beperkingen hebben wat betreft hun discriminerend vermogen, omdat er onder andere vloer- en plafondeffecten aanwezig zijn in de utiliteitsscores van deze instrumenten. Tot nu toe is er vooral bij patiënten met RA weinig onderzoek gedaan naar de psychometrische eigenschappen van instrumenten die waarderingen vanuit een patiëntenperspectief meten, namelijk de Time Trade-Off (TTO) en Standard Gamble (SG). Vanwege het feit dat patiënten unieke en uitgebreide kennis van en ervaring met hun ziekte hebben en vanwege het feit dat zij direct worden beïnvloed door interventies, zijn wij van mening dat het belangrijk is om waarderingen van patiënten te verkrijgen. Meer onderzoek is nodig naar de psychometrische eigenschappen van deze zogenoemde 'directe utiliteitsmaten' bij RA.

Dit proefschrift heeft onderzoek gedaan naar de psychometrische eigenschappen van enkele gezondheidsgelateerde KvL-instrumenten bij RA-patiënten. Daarnaast is van één van de veelgebruikte directe instrumenten (de TTO) onderzocht of een computerversie als alternatief kan dienen voor de interviewversie. Tot slot is onderzoek gedaan naar alternatieve ankers voor de TTO bij RA-patiënten, wat het meten van gezondheidsgelateerde KvL gericht op het patiëntenperspectief zou kunnen verbeteren.

In **hoofdstuk 3** werd de responsiviteit van de EQ-5D vergeleken met die van de SF-6D bij patiënten met actieve RA die werden behandeld met effectieve TNF-blokkers.

Beide instrumenten waren slechts gematigd responsief voor verbetering over een periode van een jaar. De SF-6D was in het algemeen responsiever dan de EQ-5D. Noch systematische vloer- noch systematische plafondeffecten werden gevonden in de utiliteitsscores die de mogelijkheid van de instrumenten om verandering te detecteren zouden kunnen hebben beïnvloed. Toch is de al aanwezige

hoge vloer in de SF-6D (utiliteitscore van 0.30) een beperking voor het gebruik van het instrument bij patiënten met een ernstigere RA. Het is belangrijk op te merken dat, hoewel de SF-6D responsiever was dan de EQ-5D, het gebruik ervan in kosten-utiliteitsanalyses (KUAs) zal leiden tot een minder gunstige kosten-utiliteitsratio vanwege het feit dat gemiddelde veranderingen in utiliteitscores worden gebruikt om het aantal gewonnen Quality-Adjusted Life-Years (QALYs) met een nieuwe gezondheidstechnologie te berekenen. De gemiddelde verandering in de utiliteitscore van de EQ-5D is groter dan die van de SF-6D. Dit onderstreept de grotere variantie van de EQ-5D, aangezien voor dit instrument een lagere waarde voor de responsiviteit werd gevonden. Echter, het gebruik van de SF-6D zou leiden tot minder onzekerheid wat betreft de relatieve kosten-utiliteit van twee behandelingen.

In **hoofdstuk 4** werd een computer TTO ontwikkeld en vergeleken met een interview TTO wat betreft praktische bruikbaarheid en mate van overeenstemming. De resultaten lieten zien dat de computer TTO een praktisch bruikbaar en betrouwbaar alternatief was voor de interview TTO. Echter, beide TTO instrumenten gaven niet dezelfde resultaten. Het verschil in de mediaanscore tussen beide TTO's was een klinisch relevant verschil, wat impliceert dat utiliteiten verkregen met beide TTO's niet verwisselbaar zijn. Gelukkig werden er geen systematische verschillen gevonden over het bereik van scores. Om er zeker van te zijn dat de computer TTO zonder hulp kan worden voltooid, zouden enige aanpassingen moeten worden gedaan. Aanvullend onderzoek in een grotere steekproef van RA-patiënten is nodig om de gevonden resultaten te bevestigen.

In **hoofdstuk 5** werd de constructvaliditeit van de TTO onderzocht in een kleine steekproef van RA-patiënten. De voorlopige resultaten lieten zien dat zowel de computer als de interview TTO slecht waren gecorreleerd met, én slecht in staat waren om onderscheid te maken tussen, andere zelfgerapporteerde uitkomstmaten. Na het uitsluiten van mensen die weigerden om jaren in te leveren ('zero-traders') verbeterden de resultaten wat betreft convergente validiteit. Mogelijke verklaringen zijn het feit dat de meeste zelfgerapporteerde uitkomstmaten slechts één aspect van het construct KvL meten en dat zij slechts beschrijvend van aard zijn. Bovendien blijkt de TTO vatbaar te zijn voor meetfouten ('biases'). Een voorbeeld is dat mensen vanwege het feit dat ze verliesangst hebben minder bereid zijn om levensjaren in te leveren, waardoor opwaartse 'biases' in hun utiliteitscores worden gevonden. Een andere belangrijke verklaring zou kunnen zijn dat patiënten hun ziekte niet als levensbedreigend zien waardoor zij minder bereid zijn om levensjaren in te ruilen voor perfecte gezondheid. Mogelijk zijn er realistischere ankers voor RA-patiënten. Om de huidige TTO te verbeteren zou het gebruik kunnen worden geëvalueerd van ankers die patiënten als realistisch zien. Ook zou een procedure als 'chaining' kunnen worden gebruikt, hoewel een 'chained TTO' ingewikkelder is en niet uitgebreid is gevalideerd bij chronische ziekten.

In **hoofdstuk 6** werden toekomstverwachtingen en worst-case toekomstscenario's van RA-patiënten onderzocht door middel van focusgroepen. Inzicht in verwachtingen en worst-case toekomstscenario's van patiënten gerelateerd aan hun RA zouden kunnen worden gebruikt in interventies die verwachtingen en toekomstscenario's in een realistisch perspectief zetten en die patiënten helpen om met hun angsten en zorgen om te kunnen gaan zodat hun KvL verbetert. Worst-case toekomstscenario's die als realistisch worden gezien door RA-patiënten zouden kunnen worden gebruikt als een alternatief anker voor het anker 'dood' in de TTO, wat mogelijk de validiteit en responsiviteit van het instrument verbetert. Verwachtingen en worst-case toekomstscenario's werden genoemd over afhankelijkheid van anderen, toenemende vermoeidheid, onbekwaamheid hobby's uit te voeren, onbekwaamheid te lopen en een toenemende afhankelijkheid van medicatie. Vanuit een klinisch oogpunt zijn niet alle verwachtingen en worst-case toekomstscenario's realistisch. Echter, voor patiënten waren ze wel realistisch.

Opmerkelijk was dat in onze studie pijn niet als zorg of probleem werd genoemd in tegenstelling tot in sommige andere studies. Verschillen tussen studies kunnen worden verklaard door de snelle vooruitgang in de behandeling van RA en door verschillen in de vragen die worden gesteld. Daarbij komt dat patiënten, hoewel ze zeker wel pijn ervaarden, hun pijnniveaus vergeleken met de niveaus ten tijde van de diagnose. Sindsdien was de pijn enorm afgenomen. Ook gaven deelnemers aan dat ze gewend waren geraakt aan de pijn; dat ze ermee moesten leren omgaan. Dit kan ook verklaren waarom ze pijn niet als een worst-case toekomstscenario noemden. Een grotere kwantitatieve studie bij RA-patiënten is vereist om de gevonden resultaten te valideren.

In **hoofdstuk 7** werd een kwantitatieve studie uitgevoerd om te onderzoeken welk van de vijf worst-case toekomstscenario's uit de focusgroepstudie het meest geschikt was voor patiënten met RA om te gebruiken als anker in de TTO om utiliteiten te verkrijgen. De resultaten lieten zien dat het scenario 'afhankelijk zijn van anderen' het meest geschikte scenario was, omdat dit scenario het vaakst werd gekozen als ergste om mee te maken. Geen enkele deelnemer ervaarde deze slechtste toestand op het moment al. In een toestand zijn die in de TTO wordt gebruikt zal utiliteitsscores benedenwaarts vertekenen, wat leidt tot vloereffecten die zorgen voor een onderwaardering in gezondheidsgelateerde KvL; mensen die deze toestand in de TTO al ervaren zullen automatisch een score van 0 krijgen. Voor deze mensen is het niet mogelijk om een trade-off te doen. Opmerkelijk was dat het scenario 'afhankelijk zijn van anderen' het minst werd ervaren, hoewel deze het vaakst werd gekozen als ergste om mee te maken. Het lijkt erop dat patiënten die niet veel ervaring hebben met een toestand, deze toestand het ergste vinden om mee te maken. Echter, er moet opgemerkt worden dat, hoewel het scenario 'afhankelijk zijn van anderen' het vaakst door RA-patiënten werd gekozen als ergste om mee te maken, 65% van de deelnemers dit scenario niet het vaakst had gekozen. Voor deze 65% kan het gebruik van afhankelijkheid van anderen als anker in de TTO problematisch zijn. Vandaar dat dit scenario zou moeten worden vergeleken met het huidige scenario 'dood' om te onderzoeken hoe groot het probleem is. Toch moet worden opgemerkt dat zelfs als het anker 'eerder overlijden' het worst-case toekomstscenario voor RA-patiënten blijkt te zijn, dit anker waarschijnlijk niet als een realistisch gevolg van RA wordt gezien, aangezien het niet in de focusgroepstudie werd genoemd als zorg of worst-case toekomstscenario.

In **hoofdstuk 8** werden de resultaten van het proefschrift samengevat en bediscussieerd en werden sterke kanten en beperkingen van het proefschrift besproken. Verder werden mogelijkheden genoemd voor verbetering van instrumenten en werden aanbevelingen gedaan voor toekomstig onderzoek. Het proefschrift heeft laten zien dat utiliteiten die worden verkregen met huidige meetinstrumenten beperkt zijn vanuit zowel een wetenschappelijk als psychometrisch perspectief, zoals we bij RA-patiënten hebben onderzocht. Aangezien deze instrumenten data verschaffen die een belangrijke rol spelen in economische analyses van gezondheidstechnologieën ter ondersteuning van vergoedingsbeslissingen, is het essentieel dat deze instrumenten responsief zijn en valide data genereren. Vandaar dat het nodig is verbeterde versies van instrumenten te ontwikkelen. Om het meten van gezondheidsgelateerde KvL dat het patiëntenperspectief weergeeft te verbeteren, is onderzoek gedaan naar alternatieve ankers voor de TTO die realistischer zijn voor RA-patiënten dan het huidige anker 'dood'. Een andere veelbelovende benadering om het meten van gezondheidsgelateerde KvL te verbeteren is item respons theorie (IRT). Eerder onderzoek heeft laten zien dat met behulp van Rasch-analyse kan worden aangetoond dat de EQ-5D met vijf niveaus (EQ-5D-5L) inderdaad een breder gezondheidscontinuüm meet dan de originele EQ-5D met drie niveaus. Ook zou met behulp van IRT gekeken kunnen worden of voor de huidige SF-6D de meest informatieve items (uit de SF-36) zijn gekozen. Op deze manier zou de SF-6D kunnen worden verbeterd en zou de hoge vloer kunnen worden

verlaagd. Er is ook onderzoek gedaan naar het combineren van Discrete Keuzemodellen met IRT. Dit zou een goede vervolgstap zijn om het meten van gezondheidsgerelateerde KvL gericht op het patiëntenperspectief te verbeteren.

In hoofdstuk 8 werden sterke punten van het onderzoek genoemd, zoals het feit dat dit proefschrift als eerste een interview TTO met een equivalente computer TTO heeft vergeleken. Bovendien heeft geen enkele eerdere studie RA-patiënten expliciet gevraagd naar hun toekomstverwachtingen en worst-case toekomstscenario's. Ook is er niet eerder gekeken naar alternatieve ankers voor de TTO. Echter, er zijn ook een aantal beperkingen te noemen. Zo waren data van beide TTO-studies afkomstig uit een 'convenience sample', wat de generaliseerbaarheid van de resultaten beperkt. Echter, andere studies bij RA-patiënten hebben ook een slechte constructvaliditeit van een interview TTO gevonden. Verder zijn uitsluitend instrumenten van beschrijvende aard gebruikt om de TTO te valideren. Een andere beperking heeft betrekking op het ziektespecifieke karakter van het gevonden anker. Aangezien uitsluitend RA-patiënten zijn gevraagd naar hun worst-case toekomstscenario als gevolg van hun reuma, is het gevonden worst-case anker niet zomaar toepasbaar bij andere ziekten. Een implicatie hiervan is dat deze alternatieve TTO niet direct kan worden gebruikt in KUAs die ziekten vergelijken, omdat instrumenten daarvoor generiek moeten zijn. Tot slot is in de laatste studie het huidige anker 'dood' niet meegenomen in de vergelijkingen van de scenario's. Het is dus niet bekend hoe groot het probleem van het gebruik van het anker 'dood' nu eigenlijk is. Echter, ook al zou het anker 'dood' het vaakst door RA-patiënten worden gekozen als ergste om mee te maken, eerder overlijden wordt waarschijnlijk niet gezien als een realistisch gevolg van RA. Ten slotte werden er in hoofdstuk 8 aanbevelingen gedaan voor vervolgonderzoek. De belangrijkste aanbeveling is dat toekomstig onderzoek de psychometrische eigenschappen van de TTO met een alternatief anker moet vaststellen in grote representatieve steekproeven van RA-patiënten. Ook kan IRT een belangrijke rol spelen bij het verbeteren van het meten van gezondheidsgerelateerde KvL.

Dankwoord
(Acknowledgements)

Toen ik begon met schrijven van dit proefschrift dacht ik: als ik maar bij dat dankwoord kom, dan heb ik de *belangrijkste uitdaging* gehad. Nu het zover is, kan ik zeggen dat ik de *uitdaging*, op de verdediging na, waarschijnlijk wel heb gehad. Echter, het *belangrijkste* nog niet, namelijk het uiten van mijn dankbaarheid. Dank aan allen die mij de afgelopen jaren hebben begeleid, aangemoedigd, gesteund en hun vertrouwen hebben geuit. Mede dankzij hen heb ik dit proefschrift kunnen schrijven en mede dankzij hen kan ik het afronden met een gevoel van grote trots.

In de eerste plaats wil ik mijn dankbaarheid uiten aan mijn promotor, prof. dr. Mart van de Laar, en dagelijkse begeleiders, dr. Erik Taal en dr. Annemarie Braakman-Jansen, met wie het allemaal begon.

Beste Mart, jouw oprechtheid, professionaliteit en betrokkenheid van begin tot eind, heb ik zeer gewaardeerd. Ik bewonder je altijd positieve kijk op de dingen en vooral de manier waarop je gebeurtenissen kunt ombuigen tot iets positiefs. Tijdens één van onze besprekingen, volgens mij toen een artikel een aantal keren was afgewezen, zei je tegen mij: Laurien, je proefschrift is een grote koektrommel waar steeds meer koekjes in komen; we moeten alleen de formules nog vinden om er chocoladekoekjes van te maken. Nu zijn de artikelen allemaal geaccepteerd en of het nu chocoladekoekjes zijn geworden of gewoon biscuitjes, ze vinden hun weg in de schappen. Mart, jij hebt mij de belangrijkste ingrediënten geleerd: geduldig zijn en doorzetten. Het is ons gelukt! Dank!

Beste Erik, jouw onuitputtelijke kennis en vaardigheden waarmee je verbanden kunt leggen en jouw eindeloze geduld voor al mijn vragen, bewonder ik zeer. Mede door je kritische blik en continue betrokkenheid bij het project van begin tot eind, heb ik het proefschrift kunnen voltooien. Vooral op statistische vragen, maar eigenlijk op elke vraag die ik had, had je een antwoord waardoor ik snel weer verder kon. Je nauwkeurige en snelle feedback op artikelen en andere stukken heb ik erg gewaardeerd. Dank!

Beste Annemarie, jouw vrolijkheid en energie waren erg belangrijk op het moment dat een artikel was afgewezen. Op zo'n moment kon je me opbeuren en het gevoel geven dat ik al heel wat had bereikt. Je blijdschap en enthousiasme wanneer een artikel was geaccepteerd of wanneer ik een presentatie op een congres mocht houden, gaven me een extra trots gevoel. Ik ben je dankbaar dat je me af en toe hebt aangemoedigd om even vrij te nemen en lekker te ontspannen. Elke keer besepte ik hoe belangrijk dat was. Dank!

Een speciaal woord van dank aan alle deelnemers van de studies in dit proefschrift. Zonder hen was het ook zeker niet gelukt. Hun bereidwilligheid om vragenlijsten in te vullen en om hun verhaal met mij te delen, heb ik zeer gewaardeerd. Dank!

Zeker ook een woord van dank aan de reumatologen, verpleegkundigen en secretaresses van de poli reumatologie van het MST. Zij hebben mij geholpen bij het vragen van patiënten voor mijn onderzoek en input gegeven tijdens de ochtendbespreking en wetenschapsbesprekingen. Dank voor jullie interesse en betrokkenheid!

Verder wil ik graag de co-auteurs, dr. Wietske Kievit, dr. Henk Visser en prof. dr. Piet van Riel, bedanken voor hun bijdrage aan dit proefschrift. In het bijzonder wil ik Wietske Kievit bedanken voor haar intensieve hulp bij het uitzoeken van data en haar bereidwilligheid mij wegwijs te maken in de grote hoeveelheid data. Wietske, ik heb veel van jou geleerd over het mergen van bestanden. Dat komt nu elke keer weer goed van pas. Dank!

Graag ook een woord van dank aan de leden van mijn promotiecommissie. Zij hebben mijn proefschrift bestudeerd en zijn bereid zitting te nemen in mijn promotiecommissie. Dr. Gerben Westerhof,

prof. dr. Karlein Schreurs, prof. dr. Maarten IJzerman, dr. Annelies Boonen, dr. Paul Krabbe en prof. dr. Hans Severens, hartelijk dank daarvoor!

Ook wil ik graag alle collega's van PGT bedanken voor hun steun en gezelligheid. In het bijzonder wil ik dr. Stans Drossaert bedanken voor haar inzet tijdens het promovendi-overleg en haar betrokkenheid bij de promovendi. Verder wil ik graag de secretaresses Marieke Smellink-Kleisman en Ria Stegehuis-de Vegte bedanken voor hun praktische ondersteuning en ontspannende kletsjes tussendoor. Dank!

Graag wil ik alle promovendi van de afdeling PGT, waarvan er zich inmiddels al een heleboel doctor mogen noemen, heel hartelijk danken voor hun gezelligheid en betrokkenheid. Een bijzonder woord van dank aan Stephy Nikolaus, Martine Fledderus, Marloes Vermeer, Roos van der Vaart, Pia Hunger, Ingrid Nota, Jojanneke Korte, Saskia Kelders, Sanne Lamers, Rilana Prenger en Maria Haagsma. Ik koester alle mooie, grappige en gezellige momenten die we hebben beleefd en ik hoop dat er nog veel van die momenten bij zullen komen. Het is bijzonder dat we zoveel met elkaar hebben gedeeld. Dank voor jullie gezelligheid, spontaniteit, openheid en steun!

Stephy en Martine, mijn lieve paranimfen en kamergenootjes, wat hebben we veel tijd met elkaar doorgebracht; daar kunnen onze partners jaloers op zijn ☺. Jullie stonden altijd voor mij klaar om de successen te delen, maar ook om me te steunen als het wat minder ging. Dat waardeer ik enorm. Ik ben er trots op dat jullie aan mijn zijde zullen staan tijdens de verdediging. Dank!

Lieve Marloes, ik ben jou zoveel chocolademuffins verschuldigd voor je hulp met spss dat het enige gezondheidsrisico's met zich meebrengt als je ze in één keer van me krijgt. Ik hoop daarom (maar vooral omdat je zo gezellig bent) dat we elkaar nog vaak blijven zien, zodat ik mijn schuld in termijnen kan aflossen ☺. Dank voor je hulp en betrokkenheid!

Lieve Harry, Diane, Robin, Bernadette, Jurjen en Mirthe, jullie staan altijd voor me klaar. Dat waardeer ik zeer. Mede door de vele kopjes koffie, de uitgebreide barbecues en het gezellige kamperen samen, heb ik voldoende ontspanning gekregen ☺. Dank daarvoor!

Lieve Ineke en Cor, dank voor jullie medeleven en betrokkenheid tijdens de afgelopen jaren. Ik denk graag terug aan de leuke vakanties en gezellige momenten samen en ik weet zeker dat er nog heel veel meer van die momenten bij zullen komen ☺. Dank!

Lieve Ans en Dick, zonder jullie onvoorwaardelijke steun had ik dit nooit kunnen bereiken. Ik denk met een warm gevoel terug aan de gezellige tijd met zijn vieren thuis waarbij humor niet mocht ontbreken. Jullie hebben me geleerd dat het belangrijk is om te genieten van de kleine dingen en om te waarderen wat je hebt. Ik vind dat jullie daar ook een titel voor mogen krijgen! Dank voor alles!

Lieve Mijke, ongetwijfeld kennen wij elkaar het langst. Vanaf het eerste moment dat ik je zag, dacht ik: wij horen bij elkaar ☺. We hebben aan één blik of aan een half woord genoeg; dat vind ik heel bijzonder. Ik bewonder je doorzettingsvermogen, oneindige energie en spontaniteit. Dank zus, dat je er altijd voor me bent!

Lieve Martijn, je bent voor mij van onschatbare waarde en hebt een speciaal plekje in mijn hart gekregen. Van jouw wijsheid en jouw kijk op het leven kan ik nog veel leren. Je weet denk ik niet half hoe trots ik ben op jou. Dank voor je luisterend oor, je tomeloze geduld en onvoorwaardelijke steun!

Laurien

Curriculum Vitae

Laurien Buitinga werd op 9 maart 1985 geboren te Hengelo (OV). In 2003 behaalde zij het eindexamen Gymnasium aan de Bataafse Kamp te Hengelo (OV). In datzelfde jaar begon zij met de opleiding Psychologie aan de Universiteit Twente in Enschede en in 2004 behaalde zij haar propedeuse. In 2007 behaalde zij haar mastertitel in de richting Psychogerontologie aan de Radboud Universiteit Nijmegen. In haar afstudeeronderzoek vergeleek zij werkenden en gepensioneerden wat betreft positief welbevinden en angst en depressie. In november 2007 keerde zij terug naar de Universiteit Twente en begon met haar promotie-onderzoek naar waarderingen van gezondheidsgerelateerde kwaliteit van leven bij patiënten met reumatoïde artritis.

Laurien Buitinga was born in Hengelo (OV), the Netherlands, on March 9, 1985. In 2003, she completed her secondary education at de Bataafse Kamp in Hengelo (OV). In the same year, she started with the study Psychology at the University of Twente in Enschede, and in 2004 she finished her propaedeutics. In 2007, she obtained her master's degree in Psychogerontology at the Radboud University of Nijmegen. During her graduation research she compared working people with retired people regarding positive well-being and anxiety and depression. In November 2007, she returned to the University of Twente and started her PhD study on preference-based health-related quality of life of patients with rheumatoid arthritis.

List of publications

- Buitinga L, Braakman-Jansen LMA, Taal E, van de Laar MAFJ. A computer Time Trade-Off: a feasible and reliable alternative for the interview Time Trade-Off in rheumatoid arthritis. *Clin Exp Rheumatol* 2011;29:783-9.
- Buitinga L, Braakman-Jansen LMA, Taal E, Kievit W, Visser H, van Riel PLCM, van de Laar MAFJ. Comparative Responsiveness of the EuroQol-5D and Short Form 6D to Improvement in Patients With Rheumatoid Arthritis Treated With Tumor Necrosis Factor Blockers: Results of the Dutch Rheumatoid Arthritis Monitoring Registry. *Arthritis Care Res* 2012;64:826-32.
- Buitinga L, Braakman-Jansen LMA, Taal E, van de Laar MAFJ. Future Expectations and Worst-Case Future Scenarios of Patients with Rheumatoid Arthritis: A Focus Group Study. *Musculoskeletal Care* 2012. Article first published online: 20 June 2012. DOI: 10.1002/msc.1026.
- Buitinga L, Braakman-Jansen LMA, Taal E, van de Laar MAFJ. Worst-case future scenarios of patients with rheumatoid arthritis: a cross-sectional study. *Rheumatol* 2012. Article first published online: 30 July 2012. DOI: 10.1093/rheumatology/kes176.
- Buitinga L, Braakman-Jansen LMA, Taal E, van de Laar MAFJ. Construct validity of the interview Time Trade-Off and computer Time Trade-Off in patients with rheumatoid arthritis: A cross-sectional observational pilot study. *BMC Musculoskelet Disord* 2012;13:112.

