

Health Related Utility Measurement in Rheumatology: An Introduction

C.H. Bakker^a, M. Rutten-van Mólken^b, E. van Doorslaer^c, K. Bennett^d,
and Sj. van der Linden^a

^aUniversity Hospital Maastricht, Division of Rheumatology, P.O. Box 5800, 6202 AZ Maastricht, ^bUniversity of Limburg, Department of Health Economics, P.O. Box 616, 6200 MD Maastricht, ^cErasmus University Rotterdam, Institute for Medical Technology Assessment, P.O. Box 1738, 3000 DR Rotterdam (Netherlands) and ^dMcMaster University, Department of Clinical Epidemiology and Biostatistics, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5 (Canada)

Abstract

Utility measures of health-related quality of life are preference values that patients attach to their overall health status. In clinical trials, utility measures summarize both positive and negative effects of an intervention into one single value between 0 (equal to death) and 1 (equal to perfect health). These measures allow for comparison of patient outcomes of different diseases and allow for comparison between various health care interventions. There are two different approaches to utility measurement. The first is to classify patients into categories based on their responses to a number of questions about their functional status, as for instance the Quality of Well-Being questionnaire. The second approach is to ask patients to assign a single rating to their overall health by means of rating scale, standard gamble, time trade-off, or willingness to pay. The Quality Adjusted Life Year (QALY) as outcome measure includes both effects in terms of quality and quantity of life. Utilities are used as weights to adjust life years for the quality of life in order to calculate QALYs. Both QALYs and utilities are useful in decision-making regarding appropriate procedures for groups of patients.

Key words: Utility measurement; Quality adjusted life years; Rheumatology.

Introduction

Quality of life may be affected by rheumatic diseases. In fact, quality of life is a broad concept of multiple viewpoints and includes all factors that impact upon an individual's life. Health-related quality of life includes only those factors that are part of an individual's health, which can be defined according to the World Health Organization as a state of complete physical, mental and social well-being (WHO, 1958).

Recognition of the impact of chronic diseases on the patient and the desirability to evaluate treatment effects has led to the development of instruments that measure quality of life. Health-related quality of life instruments commonly used in rheumatology usually assess pain, stiffness and physical mobility. However, these instruments are *specific*, as they aim at a specific disease (e.g. ankylosing spondylitis), or at a specific population of patients (e.g. rheumatic patients). The rationale for specific instruments lies in its potential for increased responsiveness, because only those aspects of quality of life are included for which a priori change can be expected. The disadvantage is that they cannot be used for comparisons between different

patient populations, which is possible with more generic instruments. *Generic* instruments are applicable in a wide variety of populations because they cover a broad spectrum of aspects relative to quality of life. Guyatt et al. (1989) distinguishes two major subcategories of generic instruments: health profiles and utility measures. Using *health profiles*, scores of separate items are obtained which can be combined in a few subdimensional scores and sometimes into one single index score. An example of a health profile is the Sickness Impact Profile (SIP) (Bergner et al., 1981) which measures physical, mental, emotional and social aspects of function.

Utility Measurement

Utility measures of health-related quality of life are single measures of the value or preference that the respondents attach to their overall health status. Respondents can be the general public, health care providers or patients. In health care decision-making it would be very advantageous to have such a single numerical measure that really reflects the value of the overall health improvement. It would be useful in making decisions about treatments for individual patients (clinical decision-analysis) and decisions regarding appropriate procedures and technology for groups of patients (technology assessment) (Torrance and Feeny, 1989).

In this article we will discuss patient utilities as a measure of effect in evaluating treatments. In our clinical trials we have chosen to measure patient utilities since they reflect the relative value of different health states to people we believe should benefit from services provided by the health care system (Mehrez and Gafni, 1989). In clinical trials, utility measures can be valuable because patients combine positive and negative effects of an intervention into one single value between 0 (equal to death) and 1 (equal to perfect health). With commonly used generic and specific instruments these positive and

negative effects are measured separately, and the investigator has little or no information on the patient's trade-offs among therapeutic improvements and treatment side effects.

Measurement of health-related utility

There are two approaches to utility measurement (Bell et al., 1990). The *first approach* is to classify patients into categories based on their responses to questions about their functional status. In the Quality of Well-Being (QWB), formerly called the Index of Well-Being, this approach is used. Patients are to complete a questionnaire on their performance within three dimensions: mobility, physical and social activity. Each dimension consists of five, four and five levels of performance, respectively. Patients are thus classified into one of the 43 possible combinations of levels. Each combination of levels describes a unique health state. Each health state and a standard list of symptoms and problems have already been valued by the general public and by patients with rheumatoid arthritis by means of a categorical rating from 0, equal to "as bad as dying", to 1, equal to "completely well" (Balaban et al., 1986). These ratings are used as values assigned to each health state into which the patient (responding to the QWB questionnaire), is classified. The values are then modified by the presence or absence of problems and symptoms of the standard list. They are added to establish an overall QWB value ranging from 0 (dead) to 1 (healthy).

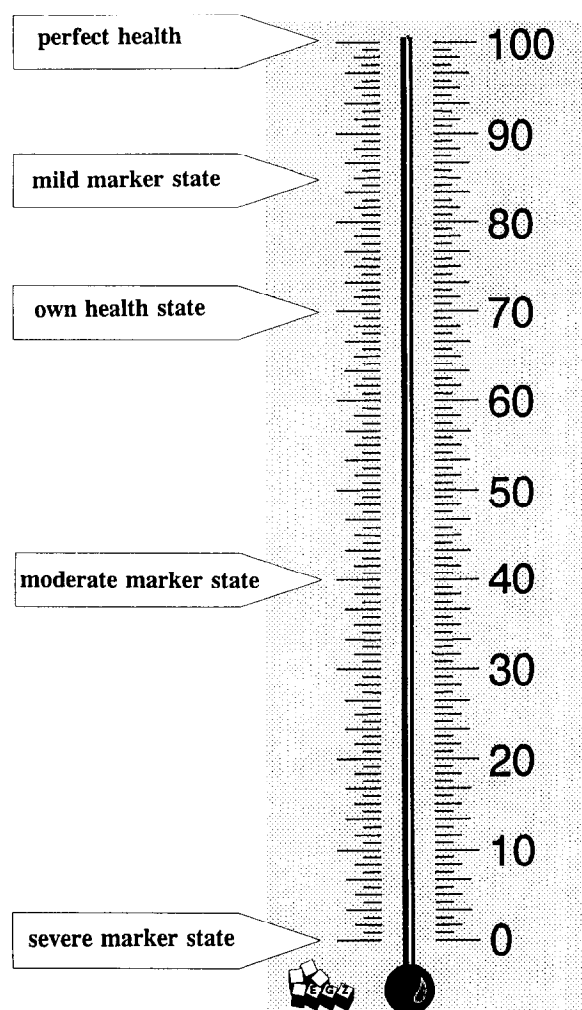
The *second approach* to utility measurement is to ask patients directly to assign one value to their overall health. The four methods which are most frequently used to elicit utility values are rating scale, standard gamble, time trade-off and willingness to pay.

A *rating scale* consists of a line on paper with clearly defined end points or anchors. It requires that the patient identifies the most and least preferred health states to use as anchor, usually labeled as "perfect health"

and as “death”. Then the patient is asked to place in order of preference his own health state and so-called marker health states on the rating scale between these anchors; such that the intervals between the placements reflect the differences the patient experiences between the health states. A useful visual aid for the rating scale is a large ‘thermometer’ with a scale from 0 to 100 (Fig. 1).

The *standard gamble* technique is based directly on the Von Neumann-Morgenstern

Fig. 1. The rating scale, a thermometer.



utility theory and is the original method of measuring utilities (Von Neumann and Morgenstern, 1944). The standard gamble method consists of paired comparison in which the patient must choose between two alternatives. Alternative 1 is a choice with two outcomes: either a good outcome, i.e. living in perfect health, with probability P ; or a bad outcome, i.e. dying, with probability $1 - P$. Alternative 2 has one outcome which is intermediate in desirability between the good and bad outcomes of Alternative 1, as for instance the patient's current health state. Probability P is varied until the patient is indifferent to the two alternatives. At this point the required utility for the patient's health state has been obtained. It is assumed that patients with a better health accept less risk in order to improve than the more severely affected patients (Torrance, 1987). The standard gamble is supplemented by the use of a probability wheel.

The *time trade-off* method is an implicit technique like the standard gamble, but it does not include risks. Both methods implicitly deduce utilities from the patients' responses to decision situations, whereas in the rating scale method, the preference values are explicitly provided by the patient (Torrance, 1987). The time trade-off is also a paired comparison in which the patient must choose between two alternatives. Alternative 1 is to maintain the patient's health state for the rest of their life (time t), while the other alternative is a shorter (time x) but healthy life. Time x is varied until the patient is indifferent to the two alternatives, at this point the required preference value for the patient's health state is x/t . It is assumed that the less desirable the patient's health state, the larger the amount of lifetime (in years or months) the patient will trade-off in order to be free from his health state (Torrance, 1987). Usually a visual aid is also used with this technique.

By means of the *willingness to pay* questionnaire, patients are asked how much money they are willing to pay for a hypothetical cure. Thompson (1986) reported that rheumatoid

arthritis patients were willing to pay 22% of their household income for a complete cure of their arthritis.

These four methods of measuring health-related utility are not interchangeable, because they are based on different assumptions and do not all include the risk component.

Patient utilities published in the literature

In the field of rheumatology, only one published randomized controlled drug trial used utility measures. Bombardier et al. (1986) reported the results of a multicenter trial in which auranofin (oral gold) was compared with placebo in the treatment of patients with rheumatoid arthritis. Outcome assessment included clinical measures (e.g. number of tender joints and erythrocyte sedimentation rate) and quality of life measures. The last set of measures included arthritis-specific instruments (e.g. Health Assessment Questionnaire (HAQ)) as well as generic instruments (e.g. Quality of Well-Being Questionnaire (QWB) and Patient Utility Measurement Set (PUMS)). Like the clinical measures, the arthritis-specific quality of life scales showed significant improvement of the auranofin group compared to the placebo group. At the same time the auranofin group reported more side effects. By means of an overall assessment as reflected in the utility score (PUMS and QWB), positive and negative effects of treatment were balanced. In the patient's opinion functional improvements were superior to injurious side effects because the PUMS as well as the QWB showed a significant improvement of auranofin in comparison to placebo.

Maastricht Utility Measurement questionnaire

As an example of an utility measurement instrument we will now describe the Maastricht Utility Measurement questionnaire. By means of this questionnaire we have elicited utility values from patients with ankylosing spondylitis or fibromyalgia (unpublished data).

The Maastricht Utility Measurement questionnaire, a Dutch translation and adaptation of the McMaster Utility Measurement questionnaire (Bennett et al., 1991), will be explained by the next three steps: (1) definition of health, (2) description of health states and (3) valuation of health states.

Definition of health. Various authors use various dimensions to define the concept of health. In the Maastricht Utility Measurement questionnaire health has been defined by six dimensions: (1) activities of daily living, (2) self-care functions, (3) emotional functions, (4) leisure activities, (5) pain, and (6) side-effects of treatment. Each dimension consists of five levels of severity: Level 1 reflects the best situation and level five the worst (Table 1 shows the backtranslation of the dimensions and its levels of the Maastricht Utility Measurement questionnaire).

Description of health states. The combination of the levels indicated by the patient in the interview, one for each dimension, was used to define 'patient's own health state'. Marker states were created by the combination of six levels, one for each dimension. Perfect health was described by combining the first levels of all six dimensions, a severe marker state was described by combining the fifth level of all six dimensions. Also a mild and moderate marker state were described by a combination of six levels (Table 2 shows the description of the mild marker state). These marker states are valuable during the measurement process, as they encourage the respondent to consider a broad range of possibilities before determining their own health state on the spectrum of possibilities (Torrance and Feeny, 1989).

Valuation of health states. The measurement of utilities is performed using rating scale and standard gamble technique. After patients have read the description of the marker states and described their health state, they are firstly asked to rank and value the health states by means of a rating scale, a thermometer with perfect health equal to 100 at the top and a se-

Table 1. Backtranslation of the six dimensions of health (Roman numeral) and its levels.

I. General Daily Activities and Mobility

Think of limitations caused by tiredness, tightness of the chest or pain while working, doing the housework, shopping, walking, climbing stairs, using public transport, driving a car, cycling, etc.

- (1) Able to perform all daily activities and duties at a normal level of mobility
- (2) Able to perform daily activities, but with some difficulties
- (3) Limited in the performance of daily activities
- (4) Limited considerably in the performance of daily activities
- (5) Unable or hardly able to perform daily activities

II. Personal Care

Think of e.g. eating, washing, taking a shower or a bath, going to the toilet, etc.

- (1) Completely capable to perform all self-care activities
- (2) Now and then having difficulty in the performance of self-care activities
- (3) Having difficulty in the performance of self-care activities
- (4) Considerable difficulty in performing self-care activities
- (5) Help needed for all self-care activities

III. Anxieties, Frustrations and Worries Related to the Course of the Disease

- (1) No anxieties, no worries, not concerned about the course of the disease
- (2) Normally no anxieties, sometimes concerned about the course of the disease
- (3) Depressed because of the inability to function normally
- (4) Often anxious, often concerned about the course of the disease
- (5) Depressive, unhappy and frustrated

IV. Leisure Activities

Think of e.g. going out, practising sports, hobbies, etc.

- (1) Able to participate in all leisure activities without difficulty
- (2) Able to participate in all leisure activities but with some difficulty
- (3) Ability to participate in leisure activities is limited
- (4) No longer able to participate in any leisure activity which requires a certain degree of physical effort or mobility
- (5) Not able to participate in any leisure activity

V. Pain

- (1) No pain
- (2) Occasionally pain
- (3) Often mild to moderate pain
- (4) Often severe pain
- (5) Continuously severe pain

VI. Side Effects of Treatment

Think of e.g. nausea, vomiting and/or diarrhoea, GI upset, skin rash, mouth ulcers.

- (1) No side effects
- (2) Occasionally mild side effects
- (3) Occasionally moderate — severe side effects
- (4) Often moderate — severe side effects
- (5) Severe side effects

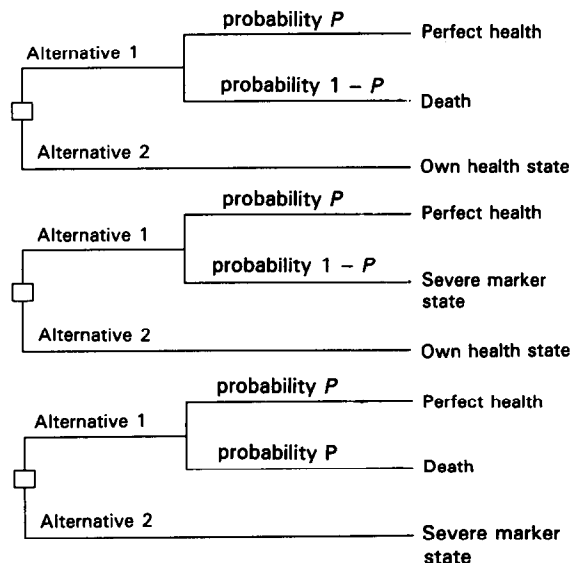
vere marker state equal to 0 at the bottom (Fig. 1). In addition, the thermometer gives the patient the opportunity to become familiar with the states and gives the investigator an indication of the ordinal rankings of the health states and information on the intensity of those preferences. Next, the standard gamble technique is performed with a probability wheel as a probe. In the standard gamble method, the health states are valued under risk, as opposed to under certainty as in the rating scale method.

In patient utility measurement, patients are usually asked how they value their own health state in comparison to perfect health and death (Fig. 2a). However, in rheumatic diseases with rather low disease-related mortality direct confrontation with the risk of dying may be inappropriate in a preference

Table 2. Description of the mild marker state.

- Able to perform all daily activities and duties at a normal level of mobility
- Completely capable to perform all self-care activities
- Normally no anxieties, sometimes concerned about the course of the disease
- Ability to participate in leisure activities is limited
- Occasionally pain
- Occasionally mild side effects

Fig. 2. (a) Standard Gamble: Value your own health state in comparison to perfect health and death. (b) Two-Step Standard Gamble: First Step, Value your own health state in comparison to perfect health and the severe marker state. (c) Two-Step Standard Gamble: Second Step, Value your own health state in comparison to perfect health and death.



assessment exercise. Therefore a two-step utility assessment is suggested. The patients are first asked to value their own health state in comparison to perfect health and the severe marker state (Fig. 2b), and then to value the severe marker state in comparison to perfect health and death (Fig. 2c), which gives a utility value for the severe marker state. *Utility* values for the patient's own health state can be calculated by using the results of these two-steps of the standard gamble (Drummond et al., 1987).

At follow-up visits it is also possible to ask the patients to compare their health state at baseline with their health state at follow-up. This additional question enables the patient to directly express the change in health-related quality of life in a utility value.

To illustrate the method of standard gamble questions that we ask patients, we now present a series of choices. Imagine you are a fibromyalgia patient and to find out how you value your health state at this moment, the first set of alternatives is: Alternative A: 100% chance of living perfectly healthy and 0% chance of living like the severe marker state. Alternative B: living in your current health state (suppose you have fibromyalgia). All health states sustain for the rest of your life. We assume you choose Alternative A. The next set is: Alternative A: 10% chance of living perfectly healthy and 90% chance of living like the severe marker state. Alternative B remains the same. We assume your fibromyalgia is not that bad that you take such a big risk and choose Alternative B. Then we continue with the next set: Alternative A: 90% chance of living perfectly healthy and 10% chance of living like the severe marker state. Alternative B remains the same. If you feel not that bad, you probably will not take the risk and therefore choose your own health state, the standard gamble stops and this last choice is reported. If you feel your fibromyalgia is that bad that you would take the risk, we continue to the next set: 20% chance of living perfectly healthy and 80% of living like the severe marker state. If you choose Alternative B, the next set of chances is 80% chance of living perfectly healthy and 20% of living like the severe marker state. If you choose A, the next set is 30% chance of living perfectly healthy and 70% of living like the severe marker state, et cetera. This is the first step of the two-step standard gamble question as presented in Fig. 3b.

Quality Adjusted Life Years

Quality of life measurements are often used in studies of interventions in rheumatic diseases, because these interventions are primarily directed towards the prevention or reduction of morbidity rather than mortality. Most interventions in rheumatology have no effect on survival, but do affect quality of life.

These effects can be positive as well as negative. Assigning utilities to intervention-induced changes in quality of life is a way to balance effects of different sizes, both positive and negative, and combine them in an overall summary value. Furthermore, using utilities as outcome measures allows not only a comparison of different interventions in rheumatic diseases, but also a comparison between these interventions and interventions in non-rheumatic diseases which also aim at improving the quality of life.

Besides interventions primarily directed at improving the quality of life, a number of health care interventions affect life expectancy. Moreover, a number of health care interventions do affect both the quality and quantity of life. To allow the comparison of the effectiveness between these various interventions, an additional outcome measure, the so-called Quality Adjusted Life Year (QALY) has been introduced.

A QALY, a concept which was first introduced by Weinstein (Weinstein and Stasson, 1977) is a single comprehensive outcome measure that includes effects in terms of both quality of life and survival. Suppose, for example, that the quality of life of an individual patient suffering from ankylosing spondylitis improves from 0.70 to 0.79 by effective drug therapy. This improvement will last for the remaining lifetime of 25 years. Suppose the survival of a patient who has had a transplant heart increases by 2 years at a quality of 0.8 and an additional half year at a quality of 0.6. This patient gains $(2 \times 0.8) + (0.5 \times 0.6) = 1.9$ QALYs, whereas the ankylosing spondylitis patient gains $25 \times (0.79 - 0.70) = 2.25$ QALYs.

In calculating QALYs, the remaining life years are weighted by using a quality-index for the patients' health state during these years. These weights can be elicited by performing utility measurement. Utilities are not the same as QALYs but are used as weights to adjust life years for the quality of life in order to

calculate QALYs (Torrance and Feeny, 1989).

Both utilities and QALYs can be related to costs, resulting in a cost-utility analysis, which is useful to planners and policy makers. QALYs have the potential advantage over utilities that the meaning of costs per unit of utility gained may not have the intuitively appealing meaning as costs per QALY gained. However, there are still a number of problems. Our aim is not to discuss them extensively, but to mention some of them. A major problem is the assumed independence between life years and quality of life which allows the direct multiplication of life years with utilities. However, it is likely that the utility that an ankylosing spondylitis patient assigns to living with moderate ankylosing spondylitis will not only be determined by this particular health state, but also by the number of years this health state is expected to last. For a possible solution to this problem see Mehrez and Gafni (1989) and Gafni (1989).

Another problem concerns the comparability of utilities and QALYs across different interventions. A comparison is not allowed when QALYs are not based on the same underlying methods of utility measurement (rating scale, standard gamble, time trade-off) which in turn have to be based on the same underlying dimensions of health. It is shown that different methods of utility measurement do produce essentially different results (Read et al., 1984). And who is to judge which dimensions are the right ones? Dimensions used in various utility measurement instruments so far may not be as sensitive to changes in chronic conditions as they are to changes achieved by acute care (Donaldson et al., 1988). This raises criticism concerning the consequences of using QALYs for the distribution of health care resources (Smith, 1987; Loomes and McKenzie, 1989).

QALYs are originally developed in the context of cost-utility analysis, thus allowing a broad comparison of interventions across

disease categories. Despite a number of problems which still have to be solved, the utility and QALY-approach will become more useful as more and more interventions are analyzed using the same underlying health dimensions and the same underlying methods for obtaining the utilities. By means of utility measurement, value judgement which otherwise implicitly guides decisions about the distribution of health care technologies is now made explicit. This has at least the potential to increase rationality in decision-making.

References

- Balaban, D.J., Sagi, P.C., Goldfarb, N.I. and Nettler, S. (1986) Weights for scoring the Quality of Well-being instrument among rheumatoid arthritics. A comparison to general population weights. *Medical Care*, 24, 973-980.
- Bell, M.J., Bombardier, C. and Tugwell, P. (1990). Measurement of functional status, quality of life and utility in rheumatoid arthritis. *Arthritis and Rheumatism*, 33, 591-601.
- Bennett, K., Torrance, G.W. and Tugwell, P. (1991). Methodologic challenges in the development of utility measures of health-related quality of life in rheumatoid arthritis. *Controlled Clinical Trials*, 12 (Suppl.), 118-128.
- Bergner, M., Bobbitt, R.A., Carter, W.B. and Gilson, B.S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care*, 14, 787-805.
- Bombardier, C., Ware, J., Russell, J., Larson, M., Chalmers, A. and Read, L. (1986). Auranofin therapy and quality of life in patients with rheumatoid arthritis. *American Journal of Medicine*, 81, 565-578.
- Donaldson, C., Atkinson, A., Bond, J. and Wright, K. (1988). Should Qualys be program-specific? *Journal of Health Economics*, 7, 239-257.
- Drummond, M.F., Stoddart, G.L. and Torrance, G.W. (1987). *Methods for the economic evaluation of health care programmes*. Oxford: Oxford Medical Publications.
- Gafni, A. (1989) The quality of QALY's (quality-adjusted-life-years): do QALY's measure what they at least intend to measure? *Health Policy*, 13, 81-83.
- Guyatt, G.H., Veldhuyzen Van Zanten, S.J.O., Feeny, D.H. and Patrick, D.L. (1989). Measuring quality of life in clinical trials: a taxonomy and review. *Canadian Medical Association Journal*, 140, 1441-1448.
- Loomes, G. and McKenzie, L. (1989). The use of QALY's in Health Care Decision Making. *Social Science and Medicine*, 28, 299-308.
- Mehrez, A. and Gafni, A. (1989). Quality-adjusted life years, utility theory and health-years equivalents. *Medical Decision Making*, 9, 142-149.
- Read, J.L., Quinn, R.J., Berwick, D.M., Fineberg, H.V. and Weinstein, M.C. (1984). Preferences for health outcomes. Comparison of assessment methods. *Medical Decision Making*, 4, 315-329.
- Smith, A., (1987) Qualms about QALYs. *Lancet*, i, 1134-1136.
- Thompson, M.S. (1986). Willingness to pay and accept risks to cure chronic disease. *American Journal of Public Health*, 76, 392-396.
- Torrance, G.W. (1987). Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases*, 40, 593-600.
- Torrance, G.W. and Feeny, D. (1989). Utilities and quality-adjusted life years. *International Journal of Technology Assessment in Health Care*, 5, 559-575.
- von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Weinstein, M.C. and Stasson, W.B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine*, 296, 716-721.
- WHO (1958) *The first ten years of the World Health Organization*. Geneva: WHO.

Correspondence to: C.H. Bakker, University Hospital Maastricht, Division of Rheumatology, P.O. Box 5800, 6202 AZ Maastricht, Netherlands.