DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS

129

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
129

PATIENT-REPORTED OUTCOME MEASUREMENT IN RHEUMATOID ARTHRITIS

MARIKA TAMMARU

TARTU UNIVERSITY
PRESS

Department of Internal Medicine, University of Tartu, Estonia

Dissertation is accepted for commencement of the degree of Doctor of Medical Sciences on November 15, 2006 by the Council of the Faculty of Medicine, University of Tartu, Estonia

Supervisor: Professor Margus Lember MD, PhD, Department of Internal Medicine, University of Tartu, Estonia

Reviewers: Associate professor Anu Aluoja PhD, Department of Psychiatry, University of Tartu, Estonia; Associate professor Ruth Kalda MD, PhD, Department of Polyclinic and Family Medicine, University of Tartu, Estonia

Opponent: Professor Désirée van der Heijde MD, PhD, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, The Netherlands

Commencement: December 12, 2006

Publication of this dissertation is granted by University of Tartu

ISSN 1024–395x
ISBN 9949–11–485–3 (trükis)
ISBN 9949–11–486–1 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 630

To Bethany

CONTENT

1. ABBREVIATIONS

ACR	American College of Rheumatology
AIMS, AIMS2	Arthritis Impact Measurement Scales
AIMS2-SF	Arthritis Impact Measurement Scales Short Form
ARA	American Rheumatism Association
EQ-5D	The EuroQol-5 Dimensions
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire Disability Index
MACTAR	McMaster-Toronto Arthritis Patient Preference Disability Questionnaire
MHAQ	Modified Health Assessment Questionnaire
NHP	Nottingham Health Profile
NHPD	Nottingham Health Profile Distress Index
PSI	Person Separation Index
Qol	Quality of life
RA	Rheumatoid arthritis
RAQoL	Rheumatoid Arthritis Specific Quality of Life Scale
RUMM	Rasch Unidimensional Measurement Model
SF-36	Short-Form 36
SIP	Sickness Impact Profile
SPSS	Statistical Package for the Social Sciences
UK	United Kingdom
US	United States (of America)
VAS	Visual analog scale

2. LIST OF ORIGINAL PUBLICATIONS

I. Tammaru M, Strompl J, Maimets K, Hanson E. The value of the qualitative method for adaptation of a disease-specific quality of life assessment instrument: the case of the Rheumatoid Arthritis Quality of Life Scale (RAQoL) in Estonia. Health and Quality of Life Outcomes 2004;2:69.

II. Tammaru M, Singh G, Hanson E, Maimets K. The adaptation, reliability and validity testing of the Estonian version of the Health Assessment Questionnaire's Disability Index. Rheumatology International, accepted for publication.

III. Tammaru M, McKenna SP, Meads DM, Maimets K., Hanson, E. Adaptation of the rheumatoid arthritis quality of life scale for Estonia. Rheumatology International 2006;267:655–62.

IV. Tammaru M, Lember M, Polluste K, McKenna SP. Cultural adaptation of quality of life measures. Scandinavian Journal of Rheumatology, accepted for publication.

3. INTRODUCTION

By linking the received care to the outcomes, health outcome research is concerned with evaluating the quality of healthcare. It specifies improvement of care and achieving the optimum benefit from available resources as its overall goals, hence being closely interrelated with economical assessment in the health sector (Rychlik, 2002). Health outcome research is intended to provide scientific evidence relating to decisions made by all who participate in the healthcare process. Although it can be argued that patients may know the effects of ill-health and treatment better than anyone, little formal attention has traditionally been given to outcomes as patients perceived them. Clinicians have been concerned primarily with physiological outcomes; healthcare administrators are interested in the indicators of the process of system proceeding.

However, during the last few decades there has been mounting change. The need to measure patients' own views regarding the experience of illness and treatment has been recognized (Rychlik, 2002; O'Connor, 2004). Assessment of patients' attitudes, experience and perceptions has formed the foundations of patient-centered approach in outcome research: patient-reported outcome research. From merely describing existing health states, patient-reported outcome research is currently expanding into new areas of investigation, including the way health states are perceived and experienced, new techniques for developing measures, and tests that may prevent future undesirable health states (O'Connor, 2004).

The prerequisite for successful evaluation is reasoned choice of methodology. After defining the purpose and conceptualizing the construct — attribute or ability — to be measured, a researcher in the field of patient-reported outcome research is confronted with successive questions concerning the selection of the measurement method to be applied in the particular population. Should the measure be developed in the context of research or can an international instrument be adapted? If it is decided upon adaptation, which one from the host of existing patient-reported outcome measures would be the best to select? What adaptation strategy should be applied and how to ensure quality of the newly adapted instrument (O'Connor, 2004)?

The way to answer these questions is significantly influenced by characteristics of the population under research. Compared to most European countries, Estonia possesses some distinguishing traits that should be taken into account when research using patient-reported outcome measures is considered. Due to the small population (1.3 million), studies involving only Estonian patients may not reach the required statistical power and restrictions related to funding also apply. Often the most practical approach for Estonian clinical and health sciences is to join in cooperative research projects, thereby the adaptation of internationally established assessment tools should usually be preferred to the development of original instruments. Considering the development of Estonian

society during the last half of the 20th century and the dramatic changes it has undergone in the last two decades, particular attention should be paid to verifying suitability and validity of those measures that assess constructs influenced by social aspects (e.g. quality of life). The relatively short history of the usage of questionnaires in health care, as well as in other areas of life in Estonia, should also be taken into account. A lack of experience being tested, which is especially the case for the older generation, provides comprehensibility as vital characteristics of assessment instruments.

To acquire patient-reported outcome measures for use in Estonia that allow participation in international research projects and are acceptable for patients, strategies should be established for the instruments' selection and adaptation. This study presents a procedure of assessing suitability, adaptation and testing of the psychometric properties of two instruments (the Health Assessment Questionnaire Disability Index, HAQ-DI and the Rheumatoid Arthritis Specific Quality of Life Scale, RAQoL) for assessment of patient-reported outcome in rheumatoid arthritis. Using these examples, it is aimed to evaluate the applicability of different strategies for patient-reported outcome instrument adaptation and relying upon the results of the research to make recommendations on the adaptation of patient-reported outcome measures in Estonia.

The example of rheumatoid arthritis was chosen for the following reasons. Firstly, the burden of rheumatoid arthritis, as a chronic and progressive condition, has a significant impact on individual sufferers and caregivers and also on society as a whole. Secondly, in the absence of cure and the improvement of patients' quality of life being a primary goal of therapy, measurement of patient-reported outcomes has an exceptional role in the assessment of therapeutic outcomes in rheumatoid arthritis. Thirdly, economic evaluations of rheumatoid arthritis, which rely on patient-related outcomes as a vital part of assessment, have become critically important for the evolvement of costly but highly effective new therapies.

Based on figures of worldwide prevalence (Kvien, 2004; Scott, 2004), 6500 to 13000 sufferers of this disabling condition are expected to live in Estonia. To be able to offer the best possible treatment and care to this group, heterogeneous in disease severity, requirements and opportunities, the general knowledge about the patients' current status should be accumulated and outcome research launched. To reach this goal, appropriate patient-reported outcome measures have to be put into use in Estonian rheumatoid arthritis research.

4. REVIEW OF LITERATURE

4.1. Assessment of patient-reported outcomes

Outcome research is a method that evaluates the effectiveness of health technologies for everyday clinical practice. The overall goal of outcome research is to improve health care and to achieve the optimum benefit from the available resources, on account of which the outcomes of different medical treatments should be assessed. The outcomes to be considered should be both those that are of clinical relevance and those that are of relevance to the patient (Rychlik, 2002; O'Connor, 2004). Traditionally, clinical endpoints assessed by physical, laboratory or radiographic measures have been utilized for outcome evaluation (Perrin, 1997). Clinical outcomes tend to be those considered most important by investigators, though they may not capture all of the outcomes that are regarded important by the patient. The perspective of the patient (patient-reported outcome) is increasingly considered within outcome research, including the influence of treatment on the patient's quality of life (Qol), the functional and health status of the patient, and the patient's satisfaction with the treatment (Rychlik, 2002; Adèr and Mellenbergh, 1999).

Patient-reported outcomes have become one of the main measures for evaluation (Rychlik, 2002). To assess the perspective of the patient, the quantitative approach, which prevails in outcome research, employs scores which are obtained by using standardized instruments, using either a patient self-completion format or interviewer-administered questionnaires. Although the interviewer-administered questionnaires produce data with the fewest amount of ambiguities and missing data, the self-complete instruments are regularly applied. Patient self-report scales are cheaper to administer and analyze, they can be used in postal surveys and they are usually less taxing for respondents (Bowling, 1995).

4.2. Generic and disease-specific instruments

The selection of the appropriate patient-completed measures for any clinical research should follow careful consideration. The choice of instrument should be driven by well-advised research goals; it is essential to be clear about the construct (e.g. health status, Qol) that needs to be assessed (Doward and McKenna, 2004).

An important question regarding the design of clinical researches is whether a generic or a specific patient-completed questionnaire should be used (McKenna, 1993). Generic measures apply across a wide range of disease and population groups, which allows inter-group comparisons of great importance in economic decision making. Disease-specific instruments apply to a specific

disease or condition grouping only. Specific measures can investigate the relevant areas in sufficient depth without enquiring into a multitude of possible problems and hence being valued for intra-group comparisons and follow-ups (Perrin, 1997; McKenna, 1993).

4.3. Psychometric properties of instruments and their testing

Indicators of psychometric properties, including reliability, validity, unidimensionality, responsiveness and acceptability, are crucial characteristics that have to be taken into account when choosing a research instrument (Streiner and Norman, 1995; Bowling, 1995; Fitzpatrick *et al*., 1998; Tennant *et al*., 2004a). However, these indicators should not be interpreted as being fixed properties of a measure, but as also depending on the specific setting — purpose of research, study design, population being studied — of the instrument's application (Fitzpatrick *et al*., 1998; O'Connor, 2004).

4.3.1. Reliability

Reliability is the first requirement for any good measurement (Murphy and Davidshofer, 2005). It refers to the ability of a measure to produce consistent results on different occasions, when there is no evidence of change (Bowling, 1995). In practice, measures are never perfectly consistent; theory of reliability estimates the effect of inconsistency on the accuracy of measurement.

The observed measurement score is a sum of a true score and an error of measurement. The true score represents a combination of the factors that lead to consistency in the measurement, which are stable characteristics of an individual or the attribute being measured. Errors of measurement represent discrepancies between obtained scores and corresponding true scores; they are formed by features of an individual or the situation, affecting scores but having nothing to do with the attribute being measured. The basic assumption of reliability theory is that measurement errors are random, i.e. they are uncorrelated with true scores and with errors of other measurements and are equally likely to be positive or negative.

It has been shown that the variance of the observed scores (σ_X^2) is the sum of the variance of true scores (σ_T^2) and the variance of measurement errors (σ_e^2). The reliability coefficient (r_{xx}) is defined as the ratio of true score variance to the total (obtained) variance of scores ($r_{xx} = \frac{\sigma_T^2}{\sigma_X^2}$ or, equivalently:

$r_{xx} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}$), and it indicates the proportion of variance in observed scores that is due to or accounted for variability in true scores (Murphy and Davidshofer, 2005).

Different strategies have been developed for practical reliability estimation. For patient-completed measures the most commonly applied are the test-retest reliability and internal consistency methods (Doward *et al*., 2004; Bowling, 1995).

4.3.1.1. Test-retest reliability

Test-retest reliability is generally assessed by correlations of the measure administered in different occasions on subjects and domains not expected to change in their scores over the time period between administrations (Murphy and Davidshofer, 2005). The rationale behind this method is that if the same instrument is administered twice, the absence of changes in the level of attribute being measured will determine that differences between scores of the test and the scores of the retest should be due solely to measurement error. The main problem inherent in the test-retest reliability assessment is that the stability of true scores over the period between measurements cannot be assured. However, the test-retest method is most useful when the long-term stability of the measurement is important, as it is the case for patient-reported outcomes measures which are intended to be included in longitudinal surveys (Murphy and Davidshofer, 2005; Doward *et al*., 2004). Correlation coefficients 0.85 or above between test and retest scores are required to indicate the acceptable level of measurement error (Weiner and Stewart, 1984).

4.3.1.2. Internal consistency

The internal consistency method of estimating measurement reliability uses estimation based solely on the number of items in the questionnaire (k) and the average inter-correlation among these items ($\bar{r}_{ij}$). These two factors are combined in the formula to estimate the reliability of the instrument ($r_{xx} = \frac{k\left(\bar{r}_{ij}\right)}{1 + \left(k-1\right)\bar{r}_{ij}}$), which yields correlation coefficient known also as Cronbach's Alpha coefficient. By means of informing on the interrelatedness of items Cronbach's Alpha tests homogeneity of the scale that can be seen as a prerequisite for consistency in the measurement. Representing a function of a number of items and the average inter-correlation among them, the internal

consistency method suggests that the test will be reliable if every item on the test measures essentially the same construct as all other items and if the number of items is large. Practically, the assessment of reliability using the internal consistency method requires one-time administration, which makes it less time-consuming and less expensive in comparison with the test-retest reliability testing (Murphy and Davidshofer, 2005). Cronbach's Alpha coefficient's value 0.7 or above is required to indicate acceptable relatedness of items. At the same time, scales with very high Alpha coefficients can be considered to indicate item redundancy (Streiner and Norman, 1995).

4.3.2. Validity

As the reliability theory provides a set of procedures for determining whether an instrument is consistent in assigning scores to an individual, then to understand fully which attribute is being measured, the validity of the instrument must be examined. Validity refers to the extent of which the scale measures the underlying concept of interest. Four essentially different ways of defining validity are recognized, including content, construct, predictive and concurrent validity (Murphy and Davidshofer, 2005; Fitzpatrick *et al*., 1998; Bowling, 1995).

Content and construct validity determines validity in terms of measurement, stating that an instrument is valid if it measures what it is supposed to measure. The assessment of content validity examines the instrument itself. Content validity is established by showing that the instrument representatively samples the attribute being measured. In applying this strategy, not only should the instrument's items be examined, but the response process should also be considered. There is no exact statistical measure of content validity; it rather represents a set of expert judgments (Murphy and Davidshofer, 2005). The content of patient-reported outcome measures should not be checked with clinical experts only. Patients' participation is required to justify suitability and completeness of the questionnaire content and whether it is easy to understand and complete (Hunt and McKenna, 1992; McKenna *et al*., 2003; Fitzpatrick *et al*., 1998). Often the distinct validation category, face validity, is used to refer to assessments of an extent for which an instrument appears reasonable to those who take it (Murphy and Davidshofer, 2005).

The goal of construct validation is to determine whether test scores provide a good measure of a specific construct (a patient-reported outcome). Construct validity is established by examining the relationships between instrument scores and other indicators. A detailed description of the construct provides a basis for describing the relationships to be expected between a good measure of that construct and a variety of measures. Construct validity is usually divided into convergent and discriminant validity. Testing of convergent validity looks for evidence that the measure of a particular construct correlates with measures of

other close constructs as hypothesized on a theoretical basis. Discriminant validity is defined as a lack of correlations with theoretically unrelated indicators (Murphy and Davidshofer, 2005). Different correlation coefficients are used to describe construct validity of a measure (Murphy and Davidshofer, 2005; Bowling, 1995). The complexness of this validation method is in exact explication of the construct, which would be the most challenging if Qol is investigated as a patient-reported outcome (Gill and Feinstein, 1994).

Predictive and concurrent validity, bound together as validity for decisions or criterion-related validity, determines an instrument as being valid if it can be used to make correct or accurate decisions. The method is to correlate the scores with measures of success or of the outcomes of decisions. These measures are referred to as criteria; the key to choosing criterion measures is an exact explication of the decision maker's goal. The idea of a predictive validity is to determine the correlation between instrument's scores, which are obtained before making decisions, and criterion scores, which are obtained after making decisions. This approach has its practical limitations. Concurrent validation strategy, the practical alternative to predictive validity, is to obtain both instrument scores and criterion scores at roughly the same time in a preselected group and compute the correlation between the two (Murphy and Davidshofer, 2005; Weiner and Stewart, 1984). In practice, criterion validity is often assessed by correlating the instrument of interest with another instrument or measure that is generally accepted as a more accurate or criterion variable. However, in the field of application of patient-report outcome measures, it is rarely, if ever, that a perfect gold standard measure exists against which to test the validity of new health status or Qol measure, and other approaches are recommended to judge instruments' validity (Fitzpatrick *et al*., 1998).

4.3.3. Unidimensionality

Unidimensionality of an instrument refers to a single underlying construct measured by the items. As a prerequisite to the summation of any set of items it is a desirable characteristic of measurement (Streiner and Norman, 1995; Tennant *et al*., 2004a). Until recently, internal consistency coefficients have been relied on to indicate unidimensionality, but this statistic merely indicates the degree of interrelation between the items in a scale. Factor-analytic methods have also been used to assess the dimensional structure of scales; however, their use may not always be valid (Cortina, 1993; Doward *et al*., 2004).

Rasch analysis is now considered to be the most efficient means of establishing unidimensionality of an instrument (Rasch, 1980; Andrich, 1988). It enables to evaluate how a set of items confirms to a unidimensional structure, but also how reliably the persons measured are separated. Rasch model assumes the probability of a test item to be 'passed' by a person being a function of the difference between the person's ability and the difficulty of the item and

function of only that difference. Formally expressed, the function takes the form $P_i(\theta) = \frac{e^{(0-b_i)}}{1+e^{(0-b_i)}}$, where e is the constant 2.718, $P_i(\theta)$ is the probability that respondents with ability θ will answer the item i correctly or be able to complete the task specified by that item, and b is the item i difficulty parameter.

The expected pattern of responses to an item or a task set is determined given the estimated θ and b. When the observed response pattern coincides with or does not deviate greatly from the expected response pattern, the items fit the measurement model and constitute a true Rasch scale. Generally, statistics which determine whether or not the data fit the model work on the principal of looking at the deviation of observed data from the model expectations. The absence of residual associations in the data after the Rasch trait removal, taken together with fit to the model supports the contention of the unidimensionality of the scale (Smith, 2000; van Alphen *et al*., 1994; Tennant *et al*., 2004a; Tennant *et al*., 2004b).

Fit of the data to the model can be assessed by a number of specific statistics based on the Chi^2 goodness-of-fit test with a significant Chi^2 value indicating misfit. Individual item fit shows the extent to which each item performance matches the Rasch-modelled expectations. The overall fit of the scale is given by the item-trait interaction statistic. This statistic gives an indication of any significant deviation of the data from the Rasch model and gives an indication of how well the items fit together to form a hierarchical and unidimensional scale. By analogy to the item-fit statistics in relation to items, the person-fit statistics are used to indicate the degree to which a person's response pattern conforms to the model. When the data fits the model, the overall distribution for the item and person statistics, should have a mean of approximately 0 and a standard deviation of approximately 1 (Andrich, 1988).

Further information on the consistency of the pattern of responses to individual items is provided by standardized residuals [(observed score - expected score)/standard error]. Residuals within the range + 2 to -2 are considered to be indicative of adequate consistency. Highly negative residuals indicate over-fit to the model (i.e. that the item is not adding anything extra to the scale and is duplicating other items) while highly positive residuals are indicative of under-fit to the model (i.e. that the item is not adding anything to the scale and is perhaps violating unidimensionality). The same approach can be applied for person's response pattern assessment. Residuals are used for item calibration, which defines the location and hierarchical order of the difficulty of the items along the continuum of the measured variable. Item calibration is expressed in logits, positioned along a hierarchical scale. A logit is defined as the natural log of an odds ratio. Logits of greater magnitude represent increasing item difficulty. Logit coverage, the distance between locations of most difficult

and least difficult items can be calculated in order to assess the capacity of scale to cover persons of different levels of ability (Prieto *et al*., 2003).

To evaluate how reliably the persons measured are separated, the Person Separation Index (PSI) should be estimated. It is a latent trait equivalent of the Cronbach's Alpha internal consistency reliability test. However, instead of raw scores, the ability estimates on a logit scale of each person are used to calculate reliability. A minimum PSI of 0.80 is required to ensure reliability of the model (Wright and Masters, 1982).

The application of Rasch analysis has an advantage for further cross-cultural validation of the instrument. Differential Items Functioning, the approach that is set within the unidimensional Rasch measurement model, is used as a mechanism to evaluate the cross-cultural validity of outcome measures. Given that the probability of an item to be affirmed is only a function of the difference between the person's ability and the difficulty of the item, individuals with the same level of ability will have the same probability to affirm an item irrespectively of any group membership. This assumption allows examination of the cross-cultural validity by contrasting the item response functions across cultures (Tennant *et al*., 2004b).

4.3.4. Responsiveness

Responsiveness, which can also be referred to as sensitivity to change, refers to the ability to detect real changes in the measured construct and is therefore relevant when instruments are to be used in an evaluative context for the measurement of health outcomes (Doward *et al*., 2004). Common methods of evaluating responsiveness include comparing scale scores before and after an intervention that is expected to affect the construct, and comparing changes in scale scores with changes in other related measures that are assumed to move jointly with the target measure. Claims for an instrument's responsiveness should be derived from longitudinal data, preferably comparing a group that is expected to change with a group that is expected to remain stable. There is no single agreed method of assessing responsiveness and a number of statistical techniques are used for quantifying it (Medical Outcomes Trust, 2002; Streiner and Norman, 1995).

4.3.5. Acceptability

Acceptability is the extent to which an instrument is acceptable to those who take it. It is an essential characteristic of a measure in order to obtain high response rates to make results easier to interpret, more generalisable and less prone to bias from non-response (Fitzpatrick *et al*., 1998; Medical Outcomes Trust, 2002). There are a number of factors that can influence acceptability,

including the mode of administration, questionnaire design and the health status of respondents. One basic way in which a questionnaire may fail to be acceptable is if it is expressed in a language unfamiliar to respondents. This issue is of particular importance if cross cultural adaptation of a measure is undertaken. Complying with the available guidelines is intended to ensure a high standard of translation (Fitzpatrick *et al*., 1998; Guillemin *et al*., 1993; Guillemin, 1995; Beaton *et al*., 2000). Although it is disputable whether rigorous translation can establish the appropriateness of an instrument in a new cultural context. Failure to identify differences in concerns and experiences in terms of salience and ways of expression from one culture to another, can lead to usage of wording that respondents feel irrelevant or incomprehensible in context of their condition. This results in diminished face validity but also acceptability of a translation (Hunt, 1998).

The acceptability of patient-based outcome measures has far less frequently been examined than issues such as reliability and validity, and there is less consensus as to what constitutes assessment of acceptability. Indicators of acceptability include administration time, response rates and levels of missing data (Fitzpatrick *et al*., 1998; Medical Outcomes Trust, 2002). However, it is argued that acceptability of a measure is best determined by pretesting (by means of additional questions or semi-structured interview) with patients in terms of wording, response options, and general layout (Fitzpatrick *et al*., 1998).

4.4. Development or adaptation of an instrument

Ethnicity and culture can have a significant effect on the assessment of subjective experiences including the assessment of patient-reported outcomes. Illness, health and Qol have different meanings in different societies, and even between cultures living within the same society (O'Connor, 2004; Swaine-Verdier *et al*., 2004). Therefore, before starting a research process, a decision has to be made about whether to develop new instruments primarily designed for the local language and cultural environment or to use widely accepted international measures. Although different language versions of well-known scales are in use across the world and within multicultural societies, there is a need for critical appraisal of the applicability of the patient-reported outcome measures which have been developed in a different cultural environment (Cheung and Thumboo, 2006). Questionnaires reflect the cultural norms of the society within which they are created. Some questionnaire items may not translate well, or at all; items that were of importance to the study population for whom the questionnaire was designed, may appear trivial in a different culture (Bowling, 1995). However, there is an increase in the number of multinational and multicultural research projects, a prerequisite to participate in

which is possessing of accepted international outcome measures. The opportunity to join international research is exceedingly substantial for small populations, where the data collected may not exhibit the statistical power needed to draw firm conclusions that are statistically significant. Also, development and validation of a new instrument requires substantial monetary investment and for small countries with a unique language, such a course is unlikely to be economically feasible. The adaptation of existing international instruments is usually seen as the optimum in terms of cost-effectiveness (Lovas *et al*., 2003; Guillemin, 1995).

4.5. Methodology of instruments adaptation

A majority of patient-completed instruments are developed in UK or US English. If such instruments are needed in non-English speaking countries, methods are required to produce translations of high quality. Cultural adaptation must be conducted to reach the closest possible equivalence between the original source and target versions of the questionnaire. It is generally acknowledged, that conceptual equivalence rather than strict closeness of form and content is necessary to assure comparability of language versions (Swaine-Verdier *et al*., 2004). Conceptual equivalence refers to the meaning of the constructs that are assessed. It is achieved when answers to the same questions reflect the same concept which can be meaningfully discussed in each of the cultures concerned (Marsella *et al*., 2000.). To guarantee conceptual equivalence of versions of an instrument, there is a clear need to follow standard procedures of translation and validation. Currently there are two different approaches to translation in the area of patient-reported outcome instrument adaptation.

The first, forward-backward translation approach constitutes a base for widely accepted cross-cultural adaptation and validation guidelines (Guillemin *et al*., 1993; Guillemin, 1995; Beaton *et al*., 2000). These guidelines, based on a review of cross-cultural adaptation in the medical, sociological and psychological literature, recommend adaptation procedure incorporating stages of translation, back translation, committee review and patients' pre-testing. Bilingual translators produce the independent translations from the source language to the target language. The translators and the recording observer synthesize the results of the translations to form the pooled target language version of the instrument. The pooled version is then translated back to the source language by translators totally blind to the original version. This stage is seen as a process of validity checking to make sure that the translated version includes the same items as the original version. The expert committee's role is to consolidate all the versions of the questionnaire and develop what would be considered the prefinal version of the questionnaire. The recommended

composition of this committee comprises methodologists, health professionals, language professionals, and the translators involved in the process up to this point. The pre-testing of the new questionnaire seeks to use the prefinal version on subjects or patients from the target settings. Each subject completes the questionnaire and is questioned to probe about what he or she thought was meant by each questionnaire item and the chosen response. The stage is intended to ensure that the adapted version is retaining its equivalence in an applied situation. Additional testing for the retention of the psychometric properties is recommended by the guidelines (Beaton *et al*., 2000).

Although widely implemented, the back-translation approach has attracted some criticism (Hunt *et al*., 1991; Leplège and Verdier, 1995; da Mota Falcao *et al*., 2003). The scientific basis of the back-translation method has been questioned. It is argued that little information of any value will be obtained from back-translation. Attention is directed to the possibility that the translation taking into account all the requirements and attaining conceptual equivalence may be very different in form and apparent content from the source item. Hence a good back-translation is quite likely to differ considerably from the original. The possible misleading impressions that may occur in a back-translation process are demonstrated on theoretical examples of combinations of forward and back-translations of different quality. The tendency to include professionals for the translation procedure is also criticized. This is believed to lead to translation that may be inappropriate for subsequent respondents because of its over-academic style (Swaine-Verdier *et al*., 2004).

Supporters of the second, dual-panel approach, see a solution in recruiting translators with varied profiles to work as teams at group meetings. The dual-panel method of translation is recommended by the European Group for Health Measurement and Quality of Life Assessment (Hunt *et al*., 1991); it has been used in the development of 20 disease-specific Qol measures in up to 30 languages (McKenna *et al*., 2004a). The translation procedure is divided into two stages. The first stage (‘bilingual’) panel utilizes people that are bilingual in the source and target languages. Preference in the selection of panel members is given to individuals without clinical background; professional translators should also be excluded. The instrument is expected to express lay rather than medical specialists’ views on the impact of the condition of interest; professional translators may find it difficult to target their translations to an appropriate level. The goal of teamwork is to produce the initial translation of the instrument. The second stage (‘lay’) panel involves people of average or below average educational level who will be presented with the initial version of the instrument produced at the first stage. The objective of the second panel is to evaluate the appropriateness of wording and content clarity of the translation. Thereafter the draft version of the measure is field-tested with patients or subjects from the target group via one-to-one semi structured interviews. The interviews are aimed to determine whether respondents are able to complete the instrument and if the content is acceptable to them. A survey for psychometric

properties testing is expected to follow (Hunt *et al*., 1991; Swaine-Verdier *et al*., 2004; Thorsen *et al*., 2001).

By our best knowledge, currently there are no comparative studies on the applicability of different approaches for patient-reported outcome instruments adaptation. In one paper (Lovas *et al*., 2003) the researchers described an attempt to establish a standard for adapting outcome instruments into the local language. However, only the applicability of the dual-panel approach was assessed in that case.

4.6. Qualitative research methods and their contribution for assessment of patient-reported outcomes

Qualitative research methods that have long been used in the social sciences are increasingly being used in health research during the last decades. For outcome research, qualitative methods present a unique opportunity for understanding the impact of a condition from the perspective of those affected by it (Pope and Mays, 2000; Arcury and Quandt, 1998; Boulton *et al*., 1996).

The goal of qualitative research is the development of concepts which help understand social phenomena in natural settings, giving due emphasis to the meanings, experience and views of the participants. Qualitative studies are concerned with answering questions such as 'what is X, how does X vary in different circumstances?' The methods used in qualitative research include direct observation, interviews, and analysis of text or documents and of recorded speech or behavior. The reasoning implicit in qualitative work is held to be prevailingly inductive, moving from data to hypotheses generation. The basic strategy to ensure rigor is systematic self-conscious design, data collection and interpretation. Persuasiveness of interpretation is seen as a main criterion for quality in qualitative research. Interpretations need to be meaningful and relevant also to the group whose actions, understandings and behavior are investigated (Pope and Mays, 2000; Pope and Mays, 1995; Mays and Pope, 1995; Denzin and Lincoln, 2000; Denscombe, 2005). The opportunity to corroborate findings by comparing the data produced by different methods can enhance the validity of qualitative data (Creswell, 2003).

Theoretical debate about the relative merits of quantitative and qualitative approaches' underlying premises has failed to establish any approach as the universally accepted 'best' for all situations. The choice of most appropriate research method should base on the aims of the particular research as well as on the practical considerations related to time, resources and access to the sources of data. A research problem is an issue that needs to be addressed first. The systematic, empirical standards of good research are often pursued in the name of four basic research goals: exploration, description, explanation and evaluation (Denscombe, 2005; Ruane 2005).

Exploratory research is conducted in the interest of increasing the understanding of new or little researched setting, group or phenomenon. This type of research usually produces qualitative data that best captures the research subject's genuine experience and understandings. Descriptive research offers a detailed picture of some phenomenon, setting, experience or group. It pays close attention to such issues as measurement and sampling, aligning naturally with quantitative methods. Explanatory research tries to identify the causes and effects of phenomena by establishing causal order and connections. Being closely related to the previous, practically oriented evaluation research seeks to judge the merits of efficacy of some specific program or policy. Explanatory and evaluation researches make a firm commitment to quantitative causal analysis (Ruane, 2005).

Qualitative and quantitative approaches to research tend to be portrayed as antithetical. However, it is increasingly emphasized that the two approaches should be regarded as complementary rather than competitive (Pope and Mays, 2000; Pope and Mays, 1995). Firstly, exploratory qualitative research is a prerequisite of good descriptive quantitative research, particularly in areas that have received little previous investigation (Pope and Mays, 2000; Pope and Mays, 1995). Qualitative techniques can be used to provide a description or understanding of a situation or behavior of interest. At their most basic, they can be used simply to discover the most comprehensible terms or words to use in a subsequent survey questionnaire. Secondly, qualitative methods can also be used to supplement quantitative work (Pope and Mays, 2000). This can be part of the validation process where different methods are used and results are compared for convergence, or as part of a multi-method approach, which examines a particular phenomenon or topic on several different levels. Thirdly, qualitative research can complement quantitative work by exploring complex phenomena or areas not amenable to quantitative research (Pope and Mays, 2000; Pope and Mays, 1995; Brannen, 1992; Tashakkori and Teddlie, 2003).

Despite the predominance of quantitative approach, there are several examples where qualitative methods have been applied for investigating patient-reported outcomes (Arcury and Quandt, 1998; Boulton *et al*., 1996; Ferrell *et al*., 2003; Gee *et al*., 2003; McEwan *et al*., 2004). Qualitative interviews have also been used during the development of patient-reported outcome measures, both to explore the construct of interest and to discover the comprehensible wordings (McKenna *et al*., 2004b; McKenna and Hunt, 1992; Doward *et al*., 2003). The dual-panel method of translation applies the qualitative interviewing technique in validating the new language versions of patient-reported outcome measures (Hunt *et al*., 1991; Swaine-Verdier, 2004).

Among different qualitative methodologies, the grounded theory's approach to analysis can be argued as being a currently preferred choice for conducting qualitative research in the field of health sciences. Due to a relatively short experience of usage of qualitative research in health sciences, its methods may be unfamiliar to health professionals and researchers with a biomedical or

natural science background. The grounded theory, characterized by rigor systematic approach and consistency, would be placed with most success in this field of traditionally quantitative thinking (Mays and Pope, 1995). Good acceptance of a method contributes to high persuasiveness of conclusions drawn by it enhancing their impact on the area's further development.

The grounded theory offers an approach for analyzing qualitative data using both inductive (open and axial coding, generation of core categories) and deductive (selective coding and theoretical/selective sampling) approaches in data processing (Glaser and Strauss, 1967; Strauss, 1987). The emerging cycle of induction and deduction, where hypotheses occurred in process of induction are instantly tested by deductive reasoning, fosters better sources of data and therefore better grounded inductions. The prerequisite for this inductive-deductive data handling is the simultaneous running of the processes of data collection, coding and analysis (Charmaz, 2002). Through the systematic gathering and analysis, the methodology enables to develop a theory that is grounded in the data and hence it is chosen to study previously under-investigated areas.

In particular, grounded theory has been adopted by those engaged in small-size projects using qualitative data for study of interactions, and by those whose research is exploratory and focused on particular settings (Denscombe, 2005). Because there is an emphasis on discovery and stress on the need to approach investigations without being blinkered by the concepts and theories of previous research, the grounded theory's approach fits neatly with the needs of researchers who are setting out to explore a new topic or a topic that has never been addressed with a certain sample or group of people. Due to its need for detailed data about activities and practice and its value for exploratory research, grounded theory's approach is well suited to small scale research conducted by individual researchers (Denscombe, 2005).

Thematic analysis is considered as a part of early procedures of data analysis in grounded theory (Ezzy, 2002; Denzin and Lincoln, 2000). However, thematic analysis can also be employed as a distinct method. One difference between these two approaches is that the grounded theory utilizes theoretical sampling in which emerging analysis guides the collection of further data, and this is not done in thematic analysis. Also the methodology for the creation of codes developed by the grounded theorists is not necessary applied in thematic analysis in its full sophistication. Thematic analysis aims to identify themes in the data. While the general issues of interest are determined prior to the analysis, the specific nature of the categories and themes to be explored are not predetermined. Similarly to the grounded theory, it may take the researcher into not anticipated issues and problems (Ezzy, 2002). The more straightforward procedures of thematic analysis may be appropriate for studies where developing a substantive theory is not specified as a research goal.

4.7. Epidemiology and burden of rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic, generally progressive disease that causes functional disability, significant pain and joint destruction, and leads to premature deaths. The established disease is distinguished from other forms of arthritis by multiple criteria; the set agreed by the American Rheumatism Association (ARA) in 1987 is usually used (Arnett *et al*., 1988). The estimated prevalence of RA in Europe and the US is between 0.5 and 1.0 %, with an incidence of about 0.03%. There is currently no known cure for the destruction of joints that occurs in RA (Hunsche *et al*., 2001; Abdel-Nasser *et al*., 1997; Kvien, 2004; Scott, 2004).

Most patients with the disease require ongoing medical treatment to retard progression or control disease flares. About one-third of patients will require surgery, often involving total joint replacement within 10 years of disease onset (Lubeck, 2001; Kvien, 2004). Two thirds of RA patients have reported to have mild-to-moderate disability, less than 10 % have severe disability (Hakala *et al*., 1994). Within 10 years of disease onset, at least 50% of patients in developed countries are unable to hold down a full-time job (Woolf and Pfleger, 2003; Scott *et al*., 2005; Lacaille, 2005). High levels of disability with consequent decreased ability to cope are commonly associated with depression, even in the early stages of the illness (Sharpe *et al*., 2001; Smedstad *et al*., 1997; Wright *et al*., 1996). Fatigue, which scores are significantly higher in patients with RA than in healthy individuals, also leads to debilitation of RA patients (Belza, 1995; Riemsma *et al*., 1998; Scott *et al*., 2005).

The magnitude of severe long-term economic consequences of RA has been underestimated in the past. The course of the disease has a significant impact on society as a whole, including individual sufferers, caregivers, employers and the health care system. The economic impact of RA can be measured in terms of both the direct and the indirect costs of the illness (Lubeck, 2001; Kvien, 2004). The direct costs of RA are those associated with the detection, treatment and/or prevention of the disease. Indirect costs are related to reduced productivity, including absences from work and diminished effectiveness, as well as losses attributable to the disease preventing individuals from taking better-paying or full-time jobs. Indirect costs also include the associated intangibles as pain, frustration, loss of Qol, and lack of self-esteem. Indirect costs incurred by RA patients and their caregivers are estimated to be higher than the direct costs. Although difficult to measure, indirect costs need to be taken into account when assessing the total disease burden on individuals and society; failure to do so carries the risk of underestimating the total impact of RA (Allaire *et al*., 1994).

Although the cost of drug treatment of RA constitutes a small proportion of the total costs of the disease to individuals and society, therapeutic interventions have the potential for significant economic benefit. Economic evaluations of RA therapies have become critically important in influencing decisions

regarding the role of costly but highly effective new therapies, particularly in settings where there are financial constraints on healthcare provisions (Emery, 2004).

4.8. Patient-reported outcome measures in RA

In the absence of cure and improvement of patients' Qol being a primary goal of therapy of RA, patient-reported outcomes are one of the few ways to measure therapeutic outcomes in addition to clinical or laboratory values (Emery, 2004; Lubeck, 2004).

The most commonly used instruments for assessment of patient-reported outcomes in RA studies are the generic Health Assessment Questionnaire, Nottingham Health Profile, Short-Form 36, EuroQol-5 Dimensions instrument and Sickness Impact Profile, and disease-specific Arthritis Impact Measurement Scale, McMaster-Toronto Arthritis Patient Preference Questionnaire and Rheumatoid Arthritis Specific Quality of Life Scale (Lubeck, 2004).

4.8.1. Generic instruments

The Health Assessment Questionnaire (HAQ) is the most frequently used assessment tool in RA studies (Ramey *et al*., 1992; Lubeck, 2004; Wolfe F, 2000). Although originating from rheumatology, the HAQ is currently accepted as a generic instrument for patient-reported outcome assessment. The HAQ Disability Index (HAQ-DI), the most widely utilized part of the instrument, assesses a patient's functional ability (Fries *et al*., 1980). The instrument consists of 20 statements about daily activities and 21 items asking about the requirement for assistive devices or help from another person. The questions are combined to form eight subscales — dressing, rising, eating, walking, hygiene, reach, grip and usual outside activities. According to the traditional scoring method, dependence on assistive devices or help from another person increases the subscale scores. The alternative rating model does not increase the scores if an assistive device or physical assistance is used (Nordenskiöld and Grimby, 1997).The HAQ-DI is self-administered in 5 to 10 minutes and it is simple to process. It is reliable, valid in various populations, sensitive to change and useful in patients with early disease. The HAQ-DI has been translated or culturally adapted into more than 60 languages and dialects (Bruce and Fries, 2003; Ramey *et al*., 1992; Lubeck 2004).

In an attempt to go beyond the evaluation of physical impairment and disability by addressing emotional and social aspects of the condition, the HAQ-DI was modified to form the Modified Health Assessment Questionnaire (MHAQ) by including questions concerning perceived patient satisfaction,

along with perceived change in the degree of difficulty (Pincus *et al*., 1983). Still, recently the original instrument was found to be more efficient at detecting change and assessing functional ability than the MHAQ (Wolfe, 2001).

The Nottingham Health Profile (NHP) is a generic instrument, developed to reflect lay rather than professional perceptions of health status (Hunt *et al*., 1985). The tool assesses perceived health by asking for 'yes' or 'no' responses to 38 statements in the following 6 sections: energy level, emotional reactions, physical mobility, pain, social isolation and sleep. The NHP has been found to be reliable and valid across a wide range of conditions; it may be self- or interviewer-administered and can be completed in 5 to 10 minutes (Coons *et al*., 2000). The NHP appears to be sensitive to changes in RA patients (Houssien *et al*., 1997). The Nottingham Health Profile Distress Index (NHPD) is a measure of patients' distress consisting of 24 items from the NHP, which unidimensionality has been suggested by preliminary Rasch analyses on UK samples (McKenna *et al*., 1993).

The Sickness Impact Profile (SIP) is a generic behavioral-based measure of ill-health or sickness-related dysfunction in performance, and may be completed by self-report or interview (Bergner *et al*., 1981). It contains 136 items that are scored in 12 categories, including 3 physical dimensions, 4 psychosocial dimensions, and 5 independent categories (sleep/rest, eating, work, home management and recreation/pastimes). Although it has shown good validity in patients with RA (Sullivan *et al*., 1990), the SIP has been criticized for not being responsive to clinical changes (MacKenzie *et al*., 1986). Also the SIP has been found to be time consuming to administrate (Damiano, 1996).

The Short-Form 36 (SF-36) is a generic health instrument which assesses 8 domains including physical function, role limitations due to physical problems, general health perceptions, vitality, pain, social function, general mental health and role limitations due to emotional problems (Ware and Sherbourne, 1992). It also includes summary scores for physical components and mental components. The instrument takes about 7 to 10 minutes to self-administer. The SF-36 is a valid and acceptable general health status questionnaire for a normal population and across a wide range of patient groups (Ware, 2000; Brazier *et al*., 1992), including those with RA (Talamo *et al*., 1997; Hagen *et al*., 1999; Kosinski *et al*., 2000). In comparison with the NHP, the SF-36 is more sensitive in a less severe disease, but less so in cases of more severe conditions. There is a growing concern that the elderly and disabled patients may find the instrument difficult to complete and its content inappropriate (Houssien *et al*., 1997; Fowler *et al*., 2000; Parker *et al*., 1998; Tennant *et al*., 1995; Brazier *et al*., 1996).

The EuroQol-5 Dimensions (EQ-5D) is a generic Qol instrument which identifies 243 possible health states based on five dimensions concerning mobility, self-care, usual activity, pain/discomfort, and anxiety/depression, and provides a single index for an individual health state (The EuroQol Group, 1990; Brazier *et al*., 1993; van Agt *et al*., 1994). The EQ-5D has been

advocated as a generic measure that could be applied to RA patients (Hurst *et al*., 1997). However, it is found that in patients with RA, the scoring properties and distributional aspects of the EQ-5D indicate substantial problems in capturing an accurate health status of this group (Wolfe and Hawley, 1997).

4.8.2. Specific instruments

The Arthritis Impact Measurement Scales (AIMS) specifically measures health outcomes in arthritis patients. It comprises nine scales measuring mobility, physical activity, dexterity, social role, social activity, activities of daily living, pain, depression and anxiety (Meenan *et al*., 1980). Nineteen additional questions address general health, health perceptions and demographic details. The AIMS takes 15 to 20 minutes to complete and is self-administered. It is reliable, valid and responsive to change (Meenan *et al*., 1982; Meenan *et al*., 1984). But yet, the AIMS fails to assess some important aspects of RA, particularly fatigue (Whalley *et al*., 1997; Lubeck, 2004).

The original AIMS questionnaire was revised by the original instrument developers to be more comprehensive and responsive to change. The AIMS2 includes three new sections that assess patient satisfaction with function, the attribution of the problem to arthritis, and a self-designation of three priority areas for improvement. It also includes three new scales: arm function, work, and support from family and friends. However, the elderly patients or those with low education level, may find the questionnaire too long to complete (Meenan *et al*., 1992). The shorter version of the AIMS2, the AIMS2-Short Form (AIMS2-SF) has been developed with demonstrated psychometric properties similar to those of the AIMS2 (Guillemin *et al*., 1997; Taal *et al*., 2004).

The McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) is RA-specific and includes a global questionnaire where patients list and rank activities affected by RA, and a patient preference questionnaire where patients prioritize functions which they most want to improve (Tugwell *et al*., 1987). The MACTAR assesses physical and social functioning by measuring mobility, self-care, work, and leisure; it does not assess psychological components of health status. A key advantage of using the MACTAR is the ability to detect the impact of treatment on functional outcomes that are of most importance to the patient. The MACTAR interview has been shown as being a valid and highly responsive instrument for assessing a change in the functional ability of patients with RA (Verhoeven *et al*., 2000). The administration of the MACTAR requires trained interviewers, therefore its feasibility is limited for standard clinical trials and clinical care. The instrument is not widely used and little information is available on other patients to be used for comparison (Lubeck, 2004).

The Rheumatoid Arthritis Quality of Life Scale (RAQoL) applies the needs-based model of Qol, which postulates that an individual's Qol is dependent on the extent to which his or her needs are fulfilled (Hunt and McKenna, 1992; Doward and McKenna, 1997). Its content was developed to assess effectiveness of a specific intervention from the patient's perspective (Whalley *et al*., 1997). The RAQoL is a 30-item measure where each item is in the form of a simple statement to which patients indicate whether or not it is true for them at that moment. The items cover a wide range of needs including self-care, activities and interpersonal relations. The self-administered instrument takes 5 to 6 minutes to complete and it is easy to score with total score expressing single index of disease-specific needs-based Qol. The RAQoL has been shown to have excellent reliability and validity (de Jong *et al*., 1997; Tijhuis *et al*., 2001; Neville *et al*., 2001; Thorsen *et al*., 2001).

4.8.3. Comparison of the instruments and general considerations

Comparison of the measures most often used in RA outcome studies is presented in Table 1.

Table 1. Patient-reported outcome instruments used in RA outcome studies (by Lubeck DP (Lubeck, 2004), modified).

Instrument	Domains	Time to complete	Advantages	Disadvantages
Generic				
HAQ-DI	Dressing/grooming Arising Eating Walking Hygiene Reach Grip Outside activity	5–10 min	Self-administered High reliability High validity High sensitivity Useful in early RA Most widely used	Does not evaluate psychological status
NHP	Energy level Emotional reactions Physical mobility Pain Social isolation Sleep	5–10 min	Self-administered Acceptable reliability Acceptable validity Acceptable sensitivity	Less sensitive to minor changes than SF-36
SIP	Physical function Psychosocial function Sleep/rest Eating Working Home management Recreation/pastimes	30 min	Self-administered Acceptable validity	Low sensitivity Lengthy to administer

Instrument	Domains	Time to complete	Advantages	Disadvantages
SF-36	Physical function Role limitations due to physical problems Health perception Vitality Pain Social function Mental health Role limitations due to emotional problems	7–10 min	Self-administered Acceptable validity Evaluates some aspects of function and well-being not covered by the HAQ-DI	Less sensitive in severe conditions than NHP Complex
EQ-5D	Mobility Self-care Usual activity Pain/discomfort Anxiety/depression	**Unknown**	Self-administered	Low validity Low sensitivity
Disease specific				
AIMS	Mobility Physical activity Dexterity Social role Social activity Activities of daily living Pain Depression Anxiety	15–20 min	Self-administered High reliability High validity High sensitivity	Does not evaluate fatigue Lengthy to administer
AIMS2	Mobility Physical activity Dexterity Social role Social activity Activities of daily living Pain Depression Anxiety Satisfaction with function Disease attribution Areas for improvement Arm function Work status Support from family/friends	23 min	Self-administered High reliability High validity Higher sensitivity than AIMS	Lengthy to administer
MACTAR	Mobility Self-care Work Leisure	10 min	Detects impact of treatment on functional outcomes of most importance to patient	Does not evaluate psychological status Requires trained interviewers Not widely used

Instrument	Domains	Time to complete	Advantages	Disadvantages
RAQoL	*Not specified by the questionnaire's authors, encompasses:* Mood/emotion Social life Hobbies Everyday tasks Personal/social relationships Physical contact	5–6 min	Self-administered High validity High sensitivity Physical contact not covered by other tools	*Have not been demonstrated*

Qualitative methods have also been applied in development of patient-reported outcome measures used in RA. The content and wording of the NHP and RAQoL were derived from qualitative interviews conducted with relevant patients (Whalley *et al*., 1997; Hunt *et al*., 1985) and observational study was used in validating the HAQ-DI (Fries *et al*., 1980).

It is clear that there is no perfect and all-encompassing instrument to assess patient-reported outcome in RA. Both generic and specific instruments for patient-reported outcome assessment in RA possess the strengths and weaknesses generally intrinsic for these approaches, and are not mutually exclusive. Each approach may be suitable for different circumstances, with the choice of tools depending on the research questions to be addressed. A combination of generic and disease-specific tools is currently recommended for RA clinical trials (Lubeck, 2004). The selection of particular instruments should be based on careful consideration of the purpose of research and conceptualization of the construct to be measured. The gathered evidence on instruments' reliability, validity and responsiveness has to be considered, as well as that of its acceptability by patients and feasibility — easiness of the instrument to be administered and processed. However, it should be taken into account that these are not fixed properties of a measure, but also depend on the specific setting of its application. Therefore, although highly informative, evidence on an instrument's performance in one particular setting cannot be automatically transferred to another context (Fitzpatrick *et al*., 1998; O'Connor, 2004).

4.9. Patient-reported outcome research in Estonia

The assessment of patient-reported outcomes is currently gaining its place as a recognized part of the clinical research in Estonia. The first attempts to expound the meaning and significance of patient-reported outcomes in Estonian professional literature have been done (Tammaru, 2004). Local versions of

different generic and disease-specific measures of patient-reported outcomes have been applied in Estonia (Herodes *et al*., 2001; Vibo *et al*., 2004; Krikmann *et al*., 2002). However, the instruments' adaptation process and psychometric properties are seldom reported (Herodes, 2002). The subject of the methodology of adaptation and psychometric properties testing of outcome measures has not been discussed in Estonian medical literature, and currently there are no recommendations on strategies for patient-completed instrument adaptations in Estonia.

Clinical and laboratory values have been traditionally used as indices for therapeutic outcomes in Estonian rheumatological practice (Veinpalu *et al*., 1992; Piai and Piai, 2000). Recently, papers that show the growing interest in evaluating patient-reported outcomes have been published. A study on the effect of health resort treatment on patients with osteoarthritis and RA used an unauthorized translation of the HAQ-DI to assess functional ability (Vaht and Birkenfeldt, 2004). Two studies where the SF-36 was applied to assess the health status of Estonian patients with chronic conditions including RA have been reported (Lai *et al*., 2001; Tender *et al*., 2003). However, the instrument's properties or rationale of its choice were not deeply discussed in either case.

4.10. Summary of literature

Health outcome research evaluates the quality of healthcare, specifying improvement of care and achievement of the optimum benefit from available resources as its overall goals. Patient-reported outcome research is an acknowledged and developing patient-centered approach in health outcome research. Patient-reported outcomes — Qol, the functional and health status and the satisfaction with the treatment — assessed via self-report scales have become the main measures for the prevailing quantitative approach in health outcome research.

The selection of the appropriate patient-completed instruments for any clinical research should follow careful consideration. Depending on research goals, generic or condition-specific questionnaires should be used. Indicators of psychometric properties, including reliability, validity, unidimensionality, responsiveness and acceptability, must be considered when choosing a research instrument. Due to economical feasibility and further opportunity to participate in inter-country comparisons, adaptation of existing international measures is often preferred to the development of new instruments. To reach the conceptual equivalence between the versions of the questionnaire, standard procedures of translation and validation must be followed. The forward-backward and dual-panel methods are the two main approaches for translation in the area of patient-reported outcome instrument adaptation. Besides the different translation

strategies, these approaches differ also in the extent and form of professionals' involvement in the adaptation process.

Qualitative methods, enabling a unique insight into patients' perspective, have been applied for investigating patient-reported outcomes either independently or in combination with quantitative approach. Qualitative methods have also proved their value as part of patient-reported outcome measures' development and validation processes.

RA is a chronic generally progressive disease that creates a considerable burden to the patients and their families and to society as a whole. In the absence of a cure and the improvement of patients' Qol being a primary goal of therapy, patient-reported outcomes are one of the few ways to measure therapeutic outcomes in RA. Several generic and condition-specific instruments have been applied in RA studies. However, there is no single perfect and all-encompassing instrument to assess patient-reported outcomes in RA; a combination of generic and disease-specific tools is recommended.

Although the evaluation of patient-reported outcomes is gaining popularity in Estonian rheumatology, currently there are no assessment tools with demonstrated psychometric properties available. Relying on the evidence published, the Health Assessment Questionnaire Disability Index, HAQ-DI and the Rheumatoid Arthritis Specific Quality of Life Scale, RAQoL can be considered as being the optimum measures to be adapted for outcome research on Estonian RA patients for the following reasons. Both measures have demonstrated excellent psychometric properties: high reliability, validity and responsiveness. The unidimensionality of the original RAQoL has also been proved. Neither the HAQ-DI nor RAQoL possesses considerable disadvantages, which could limit its applicability. Adaptation of these two instruments provides researchers with a set comprised of a general and a specific assessment tool that is useful for both inter- and intra-group comparisons. As these questionnaires evaluate different constructs, the HAQ-DI and RAQoL, being used together, allows to encompass a wide range of outcomes. Neither the HAQ-DI nor the RAQoL requires trained interviewers to administer and they are easily self-administered. Relatively short, even if administered as a battery, and simple-structured, these questionnaires should be well accepted also by respondents without previous experience with such a form of assessment. The HAQ-DI and the RAQoL are easy to handle and score, these qualities assure their convenience for a variety of RA outcome-research applications. Worldwide appreciation of the HAQ-DI and growing recognition for the RAQoL, will open broad opportunities for international comparative studies using these instruments.

Although qualitative methods have been applied at different phases of outcome research, no reports can be found on utilization of the qualitative approach for pre-adaptation evaluation of the suitability of a patient-reported outcome assessment tool. A method can be considered applicable for a

particular research if it allows to draw inferences that are important concerning achievement of research goals, that are impossible or complicated to derive using other methods, and that validity can be proved. While Estonian RA patients' disease-experience has received little previous investigation, qualitative research can be estimated as being applicable for revealing suitability of the HAQ-DI and RAQoL under local circumstances. A qualitative study, conducted prior to actual adaptation process is launched, would supposedly provide researchers with the opportunity to avoid resource-consuming procedure of adaptation of an instrument, unsuitable for local context. The potency of the qualitative approach to detect respondents' own perspective offers an opportunity to identify related topics that are relevant considering local interpretation of the measured construct but not included in the instrument, hence allowing to investigate the ability of the tool's items to cover the aspects important for the local RA patients. It is considered as especially valuable in the case of the RAQoL since understanding of Qol is highly influenced by social factors. The validity of drawn inferences can be evaluated during the successive process of instrument adaptation and by comparing with the relevant data derived by different methods.

Currently, there are no recommendations on strategies for patient-completed instrument adaptations in Estonia. Internationally published evidence can be found in support of either the forward-backward or dual-panel method; there are no comparative studies on these different approaches carried out. Due to the paucity of the literature on comparison of different adaptation methods, and in order to take into consideration characteristics of local context, the recommendations on the choice of patient-reported outcome instruments' adaptation methods in Estonia, should be based on evidence gathered during locally conducted research. On account of the lack of data supporting preferability of neither of these methods for instrument adaptation in Estonia, presently the choice of strategy for the HAQ-DI and the RAQoL adaptation for Estonia will be based on the instruments' authors recommendations and evidence on the translation strategies utilized so far for creation of the versions of these particular questionnaires. Subsequent assessment of psychometric properties of the adapted instruments and evaluation of the instruments' acceptability by the patients are meant to establish the applicability of the different translation strategies for instrument adaptation in Estonia. Adaptation method will be considered as applicable for a patient-reported outcome adaptation if it allows to create an instrument's version with acceptable reliability, validity and patients' acceptability values. A method with demonstrated applicability can be recommended for use under local circumstances.

5. AIMS OF THE RESEARCH

The aims of the present study are:

In the process of the development of a set of instruments for patient-reported outcome assessment in RA in Estonia, to investigate the applicability of qualitative methods and different translation strategies for a patient-reported outcome instrument adaptation.

1. To explore the applicability of qualitative methods for a patient-reported outcome instrument adaptation, including:
   1.1. to investigate RA patients' experience and views considering their everyday life quality (Paper I);
   1.2. to assess relevance of the RAQoL topics for Estonian patients (Paper I);
   1.3. to evaluate validity of the results by comparing the data with evidence produced during the consecutive adaptation procedure and with relevant data produced by different methods (Paper III, Paper IV).
2. To investigate the applicability of different translation strategies, forward-backward and dual-panel method, for a patient-reported outcome instrument adaptation, including:
   2.1. to assess reliability and validity of the Estonian versions of the HAQ-DI and the RAQoL created using these two methods (Paper II, Paper III);
   2.2. to evaluate acceptability of the Estonian versions of these instruments', relying on the opinions expressed by the patients during the reliability and validity testing (Paper II, Paper III).

The practical value of the present study is:

Development of a set of instruments comprising the Estonian versions of the HAQ-DI and the RAQoL for patient-reported outcome assessment in RA in Estonia.

6. METHODS

6.1. Applicability of qualitative methods for patient-reported outcome instrument adaptation (Papers I, III and IV)

6.1.1. Choice of methodology and structure of the interview (Paper I)

Qualitative methods were applied for testing the suitability of the RAQoL for Estonian RA patients. The design of the study, including theoretical basis, sampling, data collection and analysis strategies, was chosen in order to achieve deep and multifaceted insight into the investigated topic simultaneously with ensuring the rigor of the analysis. Thematic analysis along with the elements of the grounded theory was applied.

Data was collected by means of semi structured individual interviews. Four open-ended interview questions to investigate dimensions of everyday life reflected by the RAQoL were formed; two questions were intended to inquire to everyday problems/restrictions not covered by the RAQoL and to the problems' hierarchical importance for the patients.

6.1.2. Respondents and process of data collection (Paper I)

The interviewees were chosen from among the RA inpatients of Tartu University Hospital; patients' characteristics (age, gender, duration and severity of RA, education, working status, marital status, members of family unit, living conditions) were considered in order to get views from different perspectives. All interviews were conducted by one researcher. The discussion that was carried out simultaneously with the data collection, in accordance with the idea of theoretical sampling, allowed to determine the stage when the inflow of new data no longer added any essential information. No new topics relevant to the research questions came forth after the seventh interview and it was decided to stop at the tenth.

6.1.3. Coding and analyzing data (Paper I)

The interviews' transcriptions were read and discussed extensively. Tentative code families were chalked out as a notional framework for subsequent coding. Open coding of the transcripts was performed independently by two researchers. The open codes adhering to the previously identified code families were ascribed to expressions composed mostly of one or two sentences in order to distinguish their leading ideas. Axial codes were created through grouping and condensing of open codes. Side by side examination of affined axial codes

of different interviews formed the basis for creating core categories. In the analysis, the selective sampling of core categories was used.

6.1.4. Validation of results (Paper III, Paper IV)

Along with comparing the data with evidence produced during the consecutive adaptation procedure, two complementary validation strategies were applied. Firstly, transcripts of six focus-groups, that had been conducted to investigate RA patients' satisfaction with health care in a separate study, were used as comparison. Secondly, relevancy of the conclusions to a group of the Estonian RA patients was assessed by means of additional interviewing.

6.2. Applicability of different translation strategies for patient-reported outcome instrument adaptation for Estonia (Papers II and III)

6.2.1. Choice of methodology

The forward-backward and dual-panel translation methods were applied in adaptation processes of the HAQ-DI and the RAQoL respectively, as it is currently recommended by the developers of these questionnaires.

6.2.2. Linguistic validation of the HAQ-DI (Paper II)

Two native Estonian speakers, bilingual in English, independently translated the original instrument into Estonian. On the basis of these two translations, the pooled version was formed which was then translated back into English by a professional translator. The original questionnaire and the back translation were compared for coherence. Issues considering the clarity of meaning were discussed at a meeting of six RA outpatients of Tartu University Hospital. Subsequent processing of the instrument was attended by the involved translators and two experienced rheumatologists.

6.2.3. Linguistic validation of the RAQoL (Paper III)

Six native Estonians of non-clinical background worked as a group to translate the RAQoL into Estonian. The meeting was attended by one of the original authors of the RAQoL. Subsequent panel of five native Estonians of average or below average educational level discussed the clarity and immediacy of the

instructions and items. The face and content validity of the Estonian version of the RAQoL was tested by semi structured one-to-one interviews with fifteen RA outpatients of Tartu University Hospital.

6.2.4. Assessment of reliability and validity of the HAQ-DI and the RAQoL and unidimensionality of the RAQoL (Paper II, Paper III)

6.2.4.1. Procedure

Fifty RA outpatients of Tartu University Hospital were recruited. The evaluation comprised of the American College of Rheumatology (ACR) 1993 preliminary core set of disease activity measures for RA, including patient's assessment of pain on a visual analogue scale (VAS), patient's global assessment of disease activity on a VAS, the number of tender and swollen joints, and measurement of erythrocyte sedimentation rate and C-reactive protein values. (Felson *et al*., 1993; Wolfe *et al*., 2001). Radiography of hands was assessed for Larsen-Dale score (Larsen *et al*., 1977; Larsen, 1995). The participants rated their perceptions of disease severity and whether they were having a good or bad day (severity of today). Estonian versions of the HAQ-DI and the RAQoL, and the NHP (Hunt *et al*., 1986) as a comparator instrument were self-administered by all participants. In order to assess instrument's acceptability, the patients were asked to feel free to comment on the HAQ-DI and the RAQoL both verbally and in writing. The assessments (with the exception of the assessment of radiological stage) were performed and patients completed the HAQ-DI, RAQoL and NHP twice with a 2-week interval (Time 1 and Time 2). For the HAQ-DI both the traditional and the alternative scoring methods were applied.

6.2.4.2. Statistical analysis

Test-retest reliability was estimated as the correlation between Time 1 and Time 2 scores of the HAQ-DI and between Time 1 and Time 2 scores of the RAQoL; Spearman rank correlation coefficients were employed. Internal consistency was assessed using Cronbach's Alpha coefficient. For estimating the known group validity, groups were formed by self-perceived severity of RA and severity of today; Mann-Whitney U Test was applied to test differences between groups in the HAQ-DI and RAQoL scores. Convergent and discriminant validity were estimated by testing correlations between the HAQ-DI and RAQoL scores, and correlations between the adapted instruments scores and other measures. The associations were assessed by Spearman rank correlation coefficients. The statistical analysis was performed using Statistical Package for the Social Sciences (SPSS) 12.0 for Windows.

The unidimensionality of the Estonian version of the RAQoL was assessed using the Rasch analysis. The fit of the individual items was evaluated through Chi^2 fit statistics. For evaluating the unidimensionality of the Estonian RAQoL a total Chi^2 fit statistic was applied. The consistency of the pattern of responses to individual items was also assessed by standardized residuals. The items were calibrated and logit coverage was found to assess the capacity of the scale to cover persons of different levels of ability. The PSI was calculated in order to estimate reliability of differentiation of persons on the measured variable. Analyses were conducted using the Rasch Unidimensional Measurement Model (RUMM) software (Andrich, 1988).

6.3. Summary of methods

Table 2. Assessment of the suitability of the RAQoL for use in Estonia

Phase	**Purpose**	**Method of data collection**	**Participants**	**Main approach to analysis**
Interviewing	To investigate patients' experience considering Qol; To assess the importance of the instrument's topics	Individual semi structured in depth interviews	10 patients	Qualitative, thematic analysis
Validity assessment I	To assess the validity of interviewing results	Semi structured focus group interviews, independent study	6 groups, 28 patients	Discussion
Validity assessment II	To assess the validity of interview results	Individual semi structured interviews	4 patients	Discussion
Validity assessment III	To assess the validity of interviewing results	Patients' opinions expressed during psychometric properties' testing	50 patients	Discussion

Table 3. Adaptation of the HAQ-DI

Phase	**Purpose**	**Method of data collection**	**Participants**	**Main approach to analysis**
Forward translation	To translate the instrument	Individual translations	2 translators	Discussion
Back translation	To assess the coherence of translation and original instrument	Individual translation	1 translator	Discussion
Patients' panel	To assess clarity of the translation	Panel discussion	6 patients	Discussion
Specialists' group	To decide on the final wording of the translation	Group discussion	5 participants (3 translators, 2 rheumato-logists)	Discussion
Testing of psychometric properties	To assess reliability and construct validity of the translation	Questionnaires, laboratory tests	50 patients	Quantitative: statistical tests for reliability and validity

Table 4. Adaptation of the RAQoL

Phase	**Purpose**	**Method of data collection**	**Participants**	**Main approach to analysis**
Bilingual panel	To translate the instrument	Panel discussion	6 participants (healthy bilingual people)	Discussion
Lay panel	To asses clarity of the translation	Panel discussion	5 participants (healthy lay people)	Discussion
Field test interviews	To assess face and content validity of the translation	Individual semi structured interviews	15 patients	Discussion
Testing of psychometric properties	To assess reliability, construct validity and unidimensio-nality of the translation	Questionnaires, laboratory tests	50 patients	Quantitative: statistical tests for reliability and validity, Rasch analysis

7. RESULTS

7.1. Applicability of qualitative methods for patient-reported outcome instrument adaptation (Papers I, III and IV)

7.1.1. Importance of the RAQoL topics for patients (Paper I)

The interviewees were willing to share their experience and talked openly about their lives with the disease. The analysis of the gathered data highlighted three groups of items: items whose importance was demonstrated by the data; items that could conditionally be considered to be important; items whose significance could not be shown on the basis of the interview data.

Most topics assessed by the instrument were essential for the interviewed patients. For 22 of 30 items, the interview data provided evidence that allowed considering them applicable for the evaluation of the Estonian RA patients' Qol. Examples of patients' utterances supporting the significance of the content and also the appropriateness of the format of each of these items could be given from two or more interviews. Motifs related to four of the remaining eight items emerged a number of times from the interviews. However, before adding these items to the list of the appropriate ones, certain aspects of wording and interpretation should be taken into consideration. The relevance of four RAQoL items for the interviewees could not be shown.

These results allowed to state that the RAQoL could be adapted into Estonian for usage in research projects. The interviews also highlighted the difficulties in translating some specific items which were set apart for reckoning with during the adaptation process of the RAQoL.

7.1.2. Aspects that patients also found to be significant in connection with their everyday life quality (Paper I)

Three Qol aspects that were important for Estonian RA patients but were not evaluated by the RAQoL — the issues concerning changes in role performance, safety and stability of communication with the health care system, as well as some issues of procuring the necessities of life — were revealed by the analysis of interviews.

7.1.3. Validity of results (Paper III, Paper IV)

The validity of the results was proved by the subsequent successful adaptation of the questionnaire (Paper III). The potential difficulties of translation emphasized by the qualitative study were also confirmed in the course of the

adaptation. Applied complementary validation strategies corroborated validity of the conclusion drawn on analysis of the interviews, regarding the three Qol topics important for Estonian RA patients, but not evaluated by the RAQol (Paper IV). Uncertainty surrounding communication with the health care system was the central theme of all investigated focus-groups; reduction in income and increase in expenditure resulting from RA were discussed in most groups. All four RA patients, who participated in additional interviewing, agreed on the three highlighted topics being important issues of their Qol.

7.2. Applicability of different translation strategies for patient-reported outcome instrument adaptation for Estonia (Papers II and III)

7.2.1. Linguistic validation of the HAQ-DI and the RAQoL (Paper II, Paper III)

No considerable problems were observed during any of the three phases — translation, back translation, and patient testing — of the linguistic validation of the HAQ-DI. No major obstacles were experienced also in translating the RAQoL by the 'bilingual' and 'lay' panels, neither in the course of the field-test interviewing of RA patients. The wording of the translation of the HAQ-DI remained almost unchanged during the process of linguistic validation; only a minor change was introduced into the pooled version of the HAQ-DI by the patient testing panel. Into the translation of the RAQoL three changes were introduced as a result of 'lay' panel, five changes were made following the field-test interviews.

7.2.2. Reliability and validity of the HAQ-DI and the RAQoL and unidimensionality of the RAQoL (Paper II, Paper III)

During the conventional tests for psychometric properties on a sample of RA patients, both instruments showed excellent test-retest reliability and internal consistency. For the Estonian HAQ-DI, the test-retest reliability between Time 1 and Time 2 was 0.91 ($p < 0.01$) for both scoring methods, the Cronbach's Alpha was 0.96 at Time 1 and 0.97 at Time 2. For the Estonian RAQoL, the test-retest reliability was 0.91 ($p < 0.01$), and Cronbach's Alpha was 0.92 at Time 1 and 0.93 at Time 2. Both instruments also demonstrated the ability to distinguish between different groups of patient-perceived disease severity and patients' ratings of today, p of difference between groups <0.05 were shown at all occasions. Correlations between scores of adapted questionnaires and scores of the comparator measure, the NHP, and measures of disease activity and

radiological stage, demonstrated predicted convergent and discriminant validity. Neither the Estonian HAQ-DI nor the Estonian RAQoL demonstrated strong correlations with the activity measures and radiological stage. There were moderately strong correlations between the Estonian RAQoL and the HAQ-DI scores. For the NHP, the highest level of association with the Estonian HAQ-DI score was observed in the physical mobility section. The Estonian RAQoL scores were most closely related to the pain, physical mobility and energy level sections of the NHP. The study revealed no substantial differences between the psychometric properties of the original and alternative scoring methods of the Estonian HAQ-DI.

Rasch analysis indicated that the Estonian RAQoL fitted the Rasch model ($Chi^2 = 0.12$). The mean item fit was -0.025 (SD = 0.913) and the person fit was -0.070 (SD = 0.907). No items were significantly misfitted. The logit coverage was of sufficient scope to cover most patients. The PSI value was excellent (0.915) ensuring reliability of separation of the measured persons.

7.2.3. Acceptability of the HAQ-DI and the RAQoL — patients' comments and opinions (Paper II, Paper III)

Additional information that was gathered from patients' comments on the instruments during the psychometric properties survey pointed out some imperfections of the adaptations. The patients' opinions and comments revealed ambiguities of different nature in both the instructions and the items of the Estonian HAQ-DI, which were not detected during the adaptation process. For the Estonian RAQoL, only one misleading inaccuracy in the questionnaire format was discovered during the psychometric properties testing.

8. DISCUSSION

8.1. Applicability of qualitative methods for patient-reported outcome instrument adaptation (Papers I, III and IV)

Qualitative methods have been applied for investigating patient-reported outcomes by several researchers, either alone or in combination with quantitative methods (Arcury and Quandt, 1998; Boulton *et al*., 1996; Ferrell *et al*., 2003; Gee *et al*., 2003; McEwan *et al*., 2004). Qualitative interviews have been used during the development of patient-reported outcome measures, qualitative interviewing technique is utilized by the dual-panel method of translation (McKenna *et al*., 2004b; McKenna and Hunt, 1992; Doward *et al*., 2003; Hunt *et al*., 1991; Swaine-Verdier *et al*., 2004). The current research offers a new perspective for utilization of qualitative methods in outcome research, by demonstrating on the example of the RAQoL that qualitative methods can be successfully applied for testing the suitability of a patient-reported outcome measure before the actual adaptation process.

The choice of methodology of the study was based on theoretical and practical considerations and on evidence on the application of qualitative methods in health research currently available (Denzin and Lincoln, 2000; Pope and Mays, 2000; Pope and Mays, 1995; Mays and Pope, 1995) and it can be considered as befitting for investigating the Estonian patients' perspective. Thematic analysis allowed new motifs to spring up that can be considered as specific for local conditions. Failure to consider these specific topics will lead to underestimation of the impact of RA on the patients' Qol. The utilization of the theoretical sampling strategy enabled to gather data from a diversity of perspectives, relying upon the restricted number of interviews, which should be highly appreciated under conditions of resource limitation. The success of choice of individual interviewing as a data collection strategy was supported by the course of the interviews. All interviewees were highly willing to share their experience after overcoming some diffidence at the beginning of the interview, and talked openly about their lives with the disease.

The validity of the drawn conclusions was proved by the successful adaptation of the questionnaire. The findings were also corroborated by comparing the data with evidence gathered from RA patients in a separate focus-group study. Relevancy of the interpretations for the Estonian RA patients was demonstrated by additional interviewing.

The study showed that the utilization of a qualitative study as an introductory part of Qol assessment instrument adaptation provides possibilities for more thoroughly considered Qol research. The perspectives and voices of members of the potential target group of the instrument provide the researchers with unique and valuable information. Evaluating the significance of items in the particular context avoids the mechanical acceptance of instruments just

because they have performed well in other societies and cultures. Moreover, it offers an opportunity to identify topics that are not included in the instrument but are important for the local interpretation of Qol. These topics may otherwise be often overlooked (Hunt, 1998). Potential translation difficulties, highlighted during the qualitative analysis, can be overcome with enough preparation and attention in the course of the adaptation procedure. For researchers, qualitative studies offer a deeper understanding of the instrument in question and of the research topic itself, the Qol of the patients, by means of investigating the respondents' interpretations.

8.2. Applicability of different translation strategies for patient-reported outcome instrument adaptation for Estonia (Papers II and III)

Two basically different approaches to translation — the forward-backward and dual-panel translation methods — are accepted in the area of patient-reported outcome instrument adaptation (Guillemin *et al*., 1993; Guillemin, 1995; Beaton *et al*., 2000; Hunt *et al*., 1991). On the examples of the adaptation of the instruments for outcome research in RA, the current study demonstrates that both the forward-backward and dual-panel translation methods, allowing to produce instruments' versions with excellent reliability and validity figures, are applicable for patient-reported outcomes instrument validation for Estonia.

The considerable discrepancies between the adapted instruments' acceptability assessed during the psychometric properties survey are at least partly attributable to the differences in the course of adaptation procedures. The wording of the pooled version of the HAQ-DI, which relied on two independent translations, remained almost unchanged during the phases of linguistic validation. Only a minor change in one item that was suggested by the patient testing panel was introduced into the instrument before psychometric properties testing. Into the translation of the RAQoL, produced and scrutinized by the 'bilingual' panel, three changes were introduced as a result of 'lay' panel, five additional changes were made following the field-test interviews and subsequent elaboration.

These results accord with the critics which argue that the forward-backward method, due to its methodological peculiars, as being less able to achieve the conceptual equivalence of a translation and unambiguity of its wording (Hunt *et al*., 1991; Leplège and Verdier, 1995; da Mota Falcao *et al*., 2003). Both these qualities contribute to the comprehensibility of the instrument and acceptability of it by the patients.

The procedure of the forward-backward approach is oriented to achieving closeness in form and apparent content rather than conceptual equivalence between the original instrument and the adaptation. Formally affine translation

does not have to be conceptually close, and hence the conclusions, drawn from a back-translation, can be misleading. Relaying upon professional translators and medical professionals in the process of adaptation, results in generating expressions that may not be comprehendible to lay people, and the patients testing may fail to perform effectively enough to eliminate the ambiguities.

Ambiguous expressions missed during the linguistic validation process cannot be revealed by a regular postal survey for psychometric properties, where usually no informally expressed feedback is expected from the respondents. Even without intervening considerably with the results of conventional tests for reliability and validity, ambivalent wording may bring about an incline in acceptability of the instrument and a rise in the number of missed or wrongly interpreted items. Results of studies carried out using such instruments may be biased and confusing conclusions may be drawn from comparative studies where different language versions of an instrument are used.

Although both the forward-backward and dual-panel translation strategies are able to produce instruments with conventionally acceptable reliability and validity figures, on the base of the results of current study it can be argued that the dual-panel method may hold advantages in terms of achieving comprehensibility and unambiguity. Avoidance of the participation of professional translators, producing the translation during a group discussion of bilingual people without clinical background, inclusion of people of below average educational level — these qualities of the dual-panel method keep the wording close to everyday language. Deep exploration of every single item during one-to-one interviews with the patients assesses appropriateness of the expressions to describe the patients' experience. Reasoned corrections can be introduced and back-reflected to the panel members or patients during the further elaboration. Intensive communication with instrument authors or their representatives assures the conceptual equivalence of the translation. A carefully completed adaptation procedure familiarizes researchers thoroughly with the instrument and enables them to predict its future performance.

The successful application of the dual-panel method for adaptation of the RAQoL for Estonia was also clearly demonstrated by the Rasch analysis, which was the first time to apply this method of evaluating an instrument's unidimensionality in Estonian health sciences. The fitting of the instrument to the Rasch model confirmed the unidimensionality of its scale. Existence of a single underlying construct being measured indicates the items performing in the way pursued by the instrument's authors. This can be considered as referring to high conceptual equivalence between the original and the adapted questionnaires.

Highlighting the advantages of the two panel approach in achieving the instrument's comprehensibility for and acceptance by Estonian RA patients, the results of the study support the position of the researchers who suggest the dual-panel translation method as being preferred for instruments adaptation (Lovas *et al*., 2003).

The current study does not allow to draw conclusions on any adaptation method being preferable to apply. Rigorous design of comparative research did not meet the practical purpose of the study to eventuate in development of instruments for patient-reported outcome assessment in RA for Estonia. However, the present results clearly demonstrate that the acceptability by patients should be an area for elaborated investigation when the comparative value of the forward-backward and dual-panel methods is going to be established.

9. CONCLUSIONS

1. Application of qualitative methods allows testing the suitability of a patient-reported outcome measure before the actual adaptation process.
   1.1. The qualitative methodology — thematic analysis, theoretical sampling and individual interviews — allows disclosing patients experience and views in connection with their everyday life quality. The interviewees were willing to share their experience and talked openly about their lives with the disease; gathered data reflected the issues of Estonian RA patients' Qol.
   1.2. The qualitative methodology allows drawing inferences considering the relevance of the instrument's items for the patients including the items' ability to cover the aspects important for local interpretation of the measured construct. For most topics assessed by the instrument the evidence was provided that considered them applicable for the evaluation of the Estonian RA patients' Qol. Three Qol aspects that were important for Estonian RA patients but were not evaluated by the RAQoL were revealed.
   1.3. The validity of the drawn inferences can be corroborated by comparing with evidence produced during the consecutive adaptation procedure and with relevant data produced using different methods. The validity of the results was proved by the subsequent successful adaptation of the questionnaire. Both applied complementary validation strategies demonstrated validity of the conclusions drawn on analysis of the interviews.
2. Both the forward-backward and dual-panel translation methods are applicable for patient-reported outcomes instrument adaptation.
   2.1. Both methods are able to produce instruments' versions with acceptable reliability and validity figures. Excellent validity and reliability of Estonian versions of the HAQ-DI, produced using forward-backward method, and the RAQoL, produced using dual-panel method, were demonstrated.
   2.2. Both methods are able to produce instruments' versions that are well accepted by patients. However, the patients' opinions revealed ambiguities of a different nature in both the instructions and the items of the Estonian HAQ-DI; for the Estonian RAQoL, only one misleading inaccuracy in the questionnaire format was discovered. The acceptability by patients of the adapted instruments should be investigated to establish the comparative value of the translation methods.

10. PRACTICAL VALUE OF THE STUDY

A set of instruments comprising Estonian versions of the HAQ-DI and the RAQoL, both demonstrating excellent psychometric properties, was developed for patient-reported outcome assessment in RA in Estonia.

11. RECOMMENDATIONS

Based on the results of the study, recommendations can be made considering patient-related outcome instruments adaptation.

1. Utilization of a pre-adaptation qualitative study in order to test suitability of an instrument under local circumstances is recommended. Thematic analysis, theoretical sampling and individual interviews can be recommended as methods to explore patients' understanding in depth.
2. Adherence to accepted guidelines of translation in development of local versions of international instruments is recommended.
3. The applied set of determination procedures covering test-retest reliability, internal consistency and construct validity can be recommended for assessment of psychometric properties of patient-reported outcome instruments adapted for Estonia.
4. Introduction of the Rasch analysis into Estonia's adaptation practice is recommended.

12. REFERENCES

1. Abdel-Nasser AM, Rasker JJ, Valkenburg HA. Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis. Semin Arthritis Rheum 1997;27:123–40.
2. Adèr HJ, Mellenbergh GJ (eds). Research methodology: in the life, behavioural and social sciences. Thousand Oaks, London, New Delhi: Sage Publications; 1999.
3. Allaire SH, Prashker MJ, Meenan RF. The costs of rheumatoid arthritis. Pharmacoeconomics 1994;6:513–22.
4. Andrich D. Rasch models for measurement. London: Sage; 1988.
5. Arcury TA, Quandt SA. Qualitative methods in arthritis research: sampling and data analysis. Arthritis Care Res 1998;11:66–74.
6. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TS Jr, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31:315–24.
7. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.
8. Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. J Rheumatol 1995;22:639–43.
9. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981;19:787–805.
10. Boulton M, Fitzpatrick R, Swinburn C. Qualitative research in health care: II. A structured review and evaluation of studies. J Eval Clin Pract 1996;2:171–9.
11. Bowling A. Measuring disease. A review of quality of life measurement scales. Open University Press: Buckingham, Philadelphia; 1995.
12. Brannen J (ed). Mixing methods: qualitative and quantitative research. Aldershot: Avebury; 1992.
13. Brazier J, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 1993;2:169–80.
14. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;305:160–4.
15. Brazier JE, Walters SJ, Nicholl JP, Kohler B. Using the SF-36 and Euroqol on an elderly population. Qual Life Res. 1996;5:195–204.
16. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003;30:167–78.
17. Charmaz K. Qualitative interviewing and grounded theory analysis. In: Gubrium JF, Holstein JA (eds). Handbook of interview research: context & method. Thousand Oaks: Sage Publications; 2002:675–694.
18. Cheung YB, Thumboo J. Developing health-related quality-of-life instruments for use in Asia: The issues. Pharmacoeconomics 2006;24:643–50.
19. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000;17:13–35.
20. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;78:98–104.

21. Creswell JW. Research design: qualitative, quantitative and mixed methods approaches. Thousand Oaks: Sage; 2003.
22. da Mota Falcao D, Ciconelli RM, Ferraz MB. Translation and cultural adaptation of quality of life questionnaires: an evaluation of methodology. J Rheumatol 2003; 30:379–85.
23. Damiano AM. The Sickness Impact Profile. In: Spilker B (ed). Quality of life and pharmacoeconomics in clinical trials. Philadelphia (PA): Lippincott-Raven; 1996:347–354.
24. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and validity of the RAQoL. Br J Rheumatol 1997;36:878–83.
25. Denscombe M. The good research guide for small-scale social research projects. Open University Press: Maidenhead; Philadelphia 2005.
26. Denzin NK, Lincoln YS (eds). Handbook of Qualitative Research. Thousand Oaks, London, New Delhi: Sage Publications; 2000.
27. Doward L, McKenna SP. Defining patient-reported outcomes. Value Health 2004;7 Suppl1:4–8.
28. Doward LC, McKenna SP. Evolution of quality of life assessment. In: Rajagopalan R, Sheretz EF, Anderson RT (eds). Care management of skin diseases: life quality and economic impact. New York: Marcel Dekker; 1997:9–33.
29. Doward LC, Meads DM, Thorsen H. Requirements for quality of life instruments in clinical research. Value Health 2004;7 Suppl1:13–6.
30. Doward LC, Spoorenberg A, Cook SA, Whalley D, Helliwell PS, Kay LJ, McKenna SP, Tennant A, van der Heijde D, Chamberlain MA. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2003;62:20–6.
31. Emery P. Review of health economics modelling in rheumatoid arthritis. Pharmacoeconomics 2004;22 Suppl2:55–69.
32. Ezzy D. Qualitative Analysis. Practice and innovation. Australia: Allen & Unwin; 2002.
33. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, Furst D, Goldsmith C, Kieszak S, Lightfood R, Paulus H, Tugwell P, Weinblatt M, Widmark R, Williams JH, Wolfe F. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729–40.
34. Ferrell BR, Smith SL, Ervin KS, Itano J, Melancon C. A qualitative analysis of social concerns of women with ovarian cancer. Psychooncology 2003;12:647–63.
35. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i-iv,1–74.
36. Fowler RW, Congdon P, Hamilton S. Assessing health status and outcomes in a geriatric day hospital. Public Health 2000;114:440–5.
37. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
38. Gee L, Pearce E, Jackson M. Quality of life in schizophrenia: a grounded theory approach. Health Qual Life Outcomes 2003;1:31.
39. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. JAMA 1994;272:619–26.
40. Glaser BG, Strauss AL. The discovery of grounded theory. Chicago, IL: Aldine; 1967.

41. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417–32.
42. Guillemin F, Coste J, Pouchot J, Ghezail M, Bregeon C, Sany J. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. French Quality of Life in Rheumatology Group. Arthritis Rheum 1997;40:1267–74.
43. Guillemin F. Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol 1995;24:61–3.
44. Hagen KB, Smedstad LM, Uhlig T, Kvien TK. The responsiveness of health status measures in patients with rheumatoid arthritis: comparison of disease-specific and generic instruments. J Rheumatol 1999;26:1474–80.
45. Hakala M, Nieminen P, Koivisto O. More evidence from a community based series of better outcome in rheumatoid arthritis. Data on the effect of multidisciplinary care on the retention of functional ability. J Rheumatol 1994;21:1432–7.
46. Herodes M, Oun A, Haldre S, Kaasik AE. Epilepsy in Estonia: a quality-of-life study. Epilepsia 2001;42:1061–73.
47. Herodes M. Quality of life of people with epilepsy in Estonia. Tartu: Tartu University Press; 2002.
48. Houssien DA, McKenna SP, Scott DL. The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. Br J Rheumatol 1997;36:69–73.
49. Hunsche E, Chancellor JV, Bruce N. The burden of arthritis and nonsteroidal anti-inflammatory treatment. A European literature review. Pharmacoeconomics 2001; 19 Suppl1:1–15.
50. Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna SP. Cross-cultural adaptation of health measures. Health Policy 1991;19:33–44.
51. Hunt SM, McEwen J, McKenna SP. Measuring health status. London: Croom Helm; 1986.
52. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J R Coll Gen Pract 1985;35:185–8.
53. Hunt SM, McKenna SP. The QLDS: a scale for the measurement of quality of life in depression. Health Policy 1992;22:307–19.
54. Hunt SM. Cross-cultural issues in the use of quality of life measures in randomised controlled trials. In: Staquet M, Hays R, Fayers P (eds). Quality of life assessment in clinical trials. Oxford: Oxford University Press; 1998:51–68.
55. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997;36:551–9.
56. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478–87.
57. Krikmann Ü, Taba P, Asser T. Parkinsoni tõvega haigete hinnang oma elukvaliteedile. Eesti Arst 2002;81:328–32.
58. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 2004;22 Suppl2:1–12.
59. Lacaille D. Arthritis and employment research: where are we? Where do we need to go? J Rheumatol Suppl 2005;72:42–5.

60. Lai T, Kallikorm R, Salupere R, Kiivet R-A. Patsientide hinnangud oma tervisele krooniliste haiguste korral. Eesti Arst 2001;80:450–5.
61. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn 1977;18:481–91.
62. Larsen A. How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies. J Rheumatol 199522:1974–5.
63. Leplège A, Verdier A. The adaptation of health status measures: methodological aspects of the translation procedure. In: Shumaker S, Berzon R (eds). International use and performance of health-related quality of life instruments. Oxford: Oxford Rapid Communication; 1995:93–101.
64. Lovas K, Kalo Z, McKenna SP, Whalley D, Pentek M, Genti G. Establishing a standard for patient-completed instrument adaptations in Eastern Europe: experience with the Nottingham Health Profile in Hungary. Health Policy 2003;63:49–61.
65. Lubeck DP. A review of the direct costs of rheumatoid arthritis: managed care versus fee-for-service settings. Pharmacoeconomics 2001;19:811–8.
66. Lubeck DP. Patient-reported outcomes and their role in the assessment of rheumatoid arthritis. Pharmacoeconomics 2004;22 Suppl2:27–38
67. MacKenzie CR, Charlson ME, DiGioia D, Kelley K. Can the Sickness Impact Profile measure change? An example of scale assessment. J Chronic Dis 1986;39:429–38.
68. Marsella AJ, Dubanoski J, Hamada WC, Morse H. The measurement of personality across cultures: historical, conceptual, and methodological considerations. Am Behav Sci 2000;44:41–62.
69. Mays N, Pope C. Rigour and qualitative research. BMJ 1995;311:109–12.
70. McEwan MJ, Espie CA, Metcalfe J, Brodie MJ, Wilson MT. Quality of life and psychosocial development in adolescents with epilepsy: a qualitative investigation using focus group methods. Seizure 2004;13:15–31.
71. McKenna SP, Cook SA, Whalley D, Doward LC, Richards HL, Griffiths CE, Van Assche D. Development of the PSORIQoL, a psoriasis-specific measure of quality of life designed for use in clinical practice and trials. Br J Dermatol 2003;149:323–31.
72. McKenna SP, Doward LC, Meads DM, Patrick D, Tennant A. Summary of needs-based quality of life instruments. Value Health 2004a;7 Suppl1:39–40.
73. McKenna SP, Doward LC, Niero M, Erdman R. Development of needs-based quality of life instruments. Value Health 2004b;7 Suppl1:17–21.
74. McKenna SP, Hunt SM, Tennant A. The development of a patient completed index of distress from the Nottingham Health Profile: a new measure for use in cost-utility studies. British Journal of Medical Economics 1993;6:13–24.
75. McKenna SP, Hunt SM. A new measure of quality of life in depression: testing the reliability and construct validity of the QLDS. Health Policy 1992;22:321–30.
76. McKenna SP. Generic versus specific measures in clinical trials. British Journal of Medical Research 1993;6C:17–25.
77. Medical Outcomes Trust, Scientific Advisory Committee: Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.
78. Meenan RF, Anderson JJ, Kazis LE, Egger MJ, Altz-Smith M, Samuelson CO Jr, Willkens RF, Solsky MA, Hayes SP, Blocka KL, Weinstein A, Guttadauria M,

Kaplan SB, Klippel J.. Outcome assessment in clinical trials. Evidence for the sensitivity of a health status measure. Arthritis Rheum 1984;27:1344–52.
79. Meenan RF, Gertman PM, Mason JH, Dunaif R. The arthritis impact measurement scales. Further investigations of a health status measure. Arthritis Rheum 1982;25:1048–53.
80. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. Arthritis Rheum 1980;23:146–52.
81. 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.
82. Murphy KR, Davidshofer CO. Psychological testing. Principles and applications. Upper Saddle River (NJ): Pearson Prentice Hall; 2005.
83. Neville C, Whalley D, McKenna SP, Le Comte M, Fortin PR. Adaptation and validation of the rheumatoid arthritis quality of life scale for use in Canada. J Rheumatol 2001, 28:1505–10.
84. Nordenskiöld U, Grimby G. Assessments of disability in women with rheumatoid arthritis in relation to grip force and pain. Disabil Rehabil 1997;19:13–9.
85. O'Connor R. Measuring Quality of Life in Health. Churchill Livingstone; 2004.
86. Parker SG, Peet SM, Jagger C, Farhan M, Castleden CM. Measuring health status in older patients. The SF-36 in practice. Age Ageing 1998;27:13–8.
87. Perrin EB. Information systems for health outcome analysis. In: Detels R, Holland WW, McEwen J, Omenn GS (eds). Oxford textbook of public health. Oxford University Press: New York, Oxford, Tokyo; 1997:491–497.
88. Piai L, Piai S. Long-term drug therapy of rheumatoid arthritis and remission. Ter Arkh 2000;72:21–5.
89. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346–53.
90. Pope C, Mays N. Qualitative methods in health research. In: Pope C, Mays N (eds). Qualitative research in health care. London: BMJ Books; 2000:1–10.
91. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311:42–5.
92. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes 2003;1:27.
93. Ramey DR, Raynauld JP, Fries JF. The health assessment questionnaire 1992: status and review. Arthritis Care Res 1992;5:119–129.
94. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980.
95. Riemsma RP, Rasker JJ, Taal E, Griep EN, Wouters JM, Wiegman O. Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. Br J Rheumatol 1998;37:1042–6.
96. Ruane JM. Essentials of research methods: a guide to social science research. Blackwell Publishing; 2005.
97. Rychlik R. Strategies in pharmacoeconomics and outcomes research. Haworth Press: New York, London, Oxford; 2002.
98. Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Pract Res Clin Rheumatol 2005;19:117–36.

99. Scott DL. Pursuit of optimal outcomes in rheumatoid arthritis. Pharmacoeconomics 2004;22 Suppl2:13–26.
100. Sharpe L, Sensky T, Allard S. The course of depression in recent onset rheumatoid arthritis: the predictive role of disability, illness perceptions, pain and coping. J Psychosom Res 2001;51:713–9.
101. Smedstad LM, Vaglum P, Moum T, Kvien TK. The relationship between psychological distress and traditional clinical variables: a 2 year prospective study of 216 patients with early rheumatoid arthritis. Br J Rheumatol 1997;36:1304–11.
102. Smith RM. Fit analysis in latent trait measurement models. J Appl Meas 2000;1:199–218.
103. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. JAMA 1949;140:659–62.
104. Strauss AL. Qualitative analysis for social scientists. Cambridge: Cambridge University Press; 1987.
105. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 1995.
106. Sullivan M, Ahlmen M, Bjelle A. Health status assessment in rheumatoid arthritis. I. Further work on the validity of the sickness impact profile. J Rheumatol 1990;17:439–47.
107. Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. Value Health 2004;7 Suppl1:27–30.
108. 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.
109. Talamo J, Frater A, Gallivan S, Young A. Use of the short form 36 (SF36) for health status measurement in rheumatoid arthritis. Br J Rheumatol 1997;36:463–9.
110. Tammaru M. Elukvaliteedi hindamine ja käsitlused arstiteaduses. Eesti Arst 2004;83:47–50.
111. Tashakkori A, Teddlie C (eds). Handbook of mixed methods in social and behavioural research. Thousand Oaks: Sage; 2003.
112. Tender M, Kallikorm R, Rahula M, Lepiksoo M-L, Lember M. Reumatoidartriidi ja reaktiivse artriidiga patsientide elukvaliteet Eestis. Eesti Arst 2003;82:359–362.
113. Tennant A, Hillmann M, Chamberlain M, Fear J. Missing values and the MOS SF36: an over-55 year old sample with arthritis Arthritis Rheum 1995;38 Suppl9:178.
114. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004a;7 Suppl1:22–6.
115. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, Lawton G, Simone A, Carter J, Lundgren-Nilsson A, Tripolski M, Ring H, Biering-Sorensen F, Marincek C, Burger H, Phillips S. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004b;42 Suppl1:37–48.
116. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990;16:199–208.
117. Thorsen H, Hansen TM, McKenna SP, Sorensen SF, Whalley D. Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the

Rheumatoid Arthritis Quality of Life Scale (RAQoL). Scand J Rheumatol 2001;30:103–9.
118. Tijhuis GJ, de Jong Z, Zwinderman AH, Zuijderduin WM, Jansen L.M, Hazes JM, Vliet Vlieland TP. The validity of the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire. Rheumatology (Oxford) 2001;40:1112–9.
119. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. 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.
120. Vaht M, Birkenfeldt R. Hinnang 6- ja 10–12päevase sanatoorse ravi toimele osteoartroosi ning reumatoidartriidi haigetel. Eesti Arst 2004;83:9–13.
121. van Agt HM, Essink-Bot ML, Krabbe PF, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Soc Sci Med 1994;39:1537–44.
122. van Alphen A, Halfens R, Hasman A, Imbos T. Likert or Rasch? Nothing is more applicable than good theory. Adv Nurs 1994;20:196–201.
123. Veinpalu EIu, Trink RF, Veinpalu LE, Pyder KhA. The therapeutic action of the low-water bulk of therapeutic sea mud. Vopr Kurortol Fizioter Lech Fiz Kult 1992;(5–6):54–7.
124. Verhoeven AC, Boers M, van der Liden S. Validity of the MACTAR questionnaire as a functional index in a rheumatoid arthritis clinical trial. The McMaster Toronto Arthritis. J Rheumatol 2000;27:2801–9.
125. Vibo R, Teor A, Kõrv J. Ajutüveinsult ning sellega seotud elukvaliteet. Eesti Arst 2004;83:154–9.
126. 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.
127. Ware JE Jr. SF-36 health survey update. Spine 2000;25:3130–9.
128. Weiner EA, Stewart BJ. Assessing individuals. Boston: Little Brown; 1984.
129. Whalley D, McKenna SP, de Jong Z, van der Heijde D. Quality of life in rheumatoid arthritis. Br J Rheumatol 1997;36:884–8.
130. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. Br J Rheumatol 1997;36:786–93.
131. Wolfe F, O'Dell JR, Kavanaugh A, Wilske K, Pincus T. Evaluating severity and status in rheumatoid arthritis. J Rheumatol 2001;28:1453–62.
132. Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000;43:2751–61.
133. Wolfe F. Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ), and a rescored 20 item HAQ (HAQ20): analyses in 2,491 rheumatoid arthritis patients following leflunomide initiation. J Rheumatol 2001;28:982–9.
134. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003;81:646–56.
135. Wright BD, Masters GN. Rating Scale Analysis: Rasch measurement. Chicago: MESA Press; 1982.
136. Wright GE, Parker JC, Smarr KL, Schoenfeld-Smith K, Buckelew SP, Slaughter JR, Johnson JC, Hewett JE. Risk factors for depression in rheumatoid arthritis. Arthritis Care Res 1996;9:264–72.

13. SUMMARY IN ESTONIAN

Patsiendikeskse tulemi hindamine reumatoidartriidi-haigetel

Tulemiuuringud on meetod, millega igapäevases kliinilises töös hinnatakse erinevate menetluste tõhusust. Majandusliku hindamisega tihedalt seotud tulemiuuringute eesmärgiks on parandada tervishoiu kvaliteeti ja saavutada tõhus ressursikasutus. Traditsiooniliste objektiivsete näitajate registreerimise kõrval keskendutakse tulemiuuringutes üha enam patsiendikesksele tulemile: patsientide antud hinnangutele elukvaliteedile, terviseseisundile, funktsionaalsele seisundile ja rahulolule raviga.

Kvantitatiivsetes tulemiuuringutes, kus hinnatakse patsiendikeskset tulemit, on saanud juhtivaks käsitlusviisiks patsientide enesehinnanguküsimustike kasutamine. Uuringu edukuse üheks aluseks on hindamisvahendi (küsimustiku) kaalutletud valik. Uuringu eesmärgist lähtuvalt on vajalik kasutada üldisi või haigusspetsiifilisi küsimustikke. Valikul on oluline arvestada hindamisvahendi psühhomeetrilisi omadusi: reliaablust ehk usaldusväärsust, valiidsust, ühemõõtmelisust, tundlikkust muutuste suhtes ja aktsepteeritavust patsientide poolt. Originaalküsimustiku loomisele on sageli eelistatud rahvusvahelise hindamisvahendi kohandamine, sest see annab võimaluse osaleda rahvusvahelistes võrdlusuuringutes. Küsimustiku keeleliste versioonide võrreldavuse tagab nende kontseptuaalne samaväärsus, s.o omadus kirjeldada sama tulemit nii, et see on erinevates sihtkultuurides sarnaselt mõistetav. Hindamisvahendi keeleversioonide kontseptuaalse samaväärsuse saavutamiseks on oluline järgida standardseid tõlkimis- ja valideerimismenetlusi. Patsiendikeskse tulemi hindamise küsimustike kohandamisel on kasutusel tõlke-tagasitõlke ja kahe paneeli meetodid, mis erinevad nii tõlkestrateegiate kui ka professionaalide haaratuse poolest kohandamisse. Praegu puuduvad uuringud, kus oleks võrreldud nende kahe meetodi rakendamist.

Sotsiaalteadustest pärineva ja meditsiinis üha laiemalt kasutatava kvalitatiivse käsitlusviisi eesmärgiks on kirjeldada ja mõista sotsiaalseid nähtusi, asetades rõhu tähendustele, kogemustele ja arusaamadele. Tulemiuuringutele annab kvalitatiivne meetod ainulaadse võimaluse tuua välja patsientide vaatenurk. Kvalitatiivsed uuringud on olnud patsiendikeskse tulemi uurimisel kasutuses nii iseseisva meetodina kui ka kombineeritult kvantitatiivse käsitlusega. Kvalitatiivset meetodit on edukalt kasutatud patsiendikeskse tulemi hindamise küsimustike loomisel ja valideerimisel.

Reumatoidartriit (RA) on krooniline progresseeruv haigus, mille koormus nii haigetele ja nende perekondadele kui ka ühiskonnale tervikuna on märkimisväärne. Tervistava ravi puudumisel on RA ravi üheks peamiseks eesmärgiks tagada haigete igapäevane toimetulek ja säilitada nende elu kvaliteet. Seetõttu on RA-haigetel oluline uurida ravi tulemuslikkust patsiendikeskse tulemi alusel. Selleks on kasutatud mitmeid üldisi ja haigusspetsiifilisi küsimustikke. Siiski puudub hindamisvahend, mis lubaks võrdselt hästi uurida kõiki RA korral

olulisi patsiendikeskse tulemi valdkondi. Et leida uurimisküsimusele ammendav vastus, on sageli õigustatud üldiste ja spetsiifiliste küsimustike kombineerimine.

Hiljuti Eestis tehtud reumatoloogiauurimustest nähtub, et vajadus tulemit hindavate küsimustike järele on olemas. Samas puudub Eestis RA-patsiendikeskse tulemi hindamiseks küsimustik, mille töökindlus ja sobivus meie oludes oleks uuringutes kinnitust leidnud. Tuginedes seni avaldatud andmetele, võib sobivamateks patsiendikeskse tulemi hindamisvahenditeks Eesti RA-haigetel pidada tervisehinnangu küsimustiku HAQ funktsionaalset osa HAQ-DI (*Health Assessment Questionnaire Disability Index*) ja RA-spetsiifilist elukvaliteedi küsimustikku RAQoL (*Rheumatoid Arthritis Quality of Life Scale*). Eelistus põhineb nimetatud küsimustike omadustel. Mõlemad hindamisvahendid on uuringutes näidanud suurt usaldusväärsust, valiidsust ja tundlikkust muutuste suhtes; kummalgi küsimustikul pole kirjeldatud puudusi, mis võiksid piirata nende kasutamist. Üldine HAQ-DI ja spetsiifiline RAQoL võimaldavad edukalt korraldada nii rühmadevahelisi kui ka rühmasiseseid võrdlusi. Hinnates muutusi erinevates tulemites, lubavad HAQ-DI ja RAQoL koos kasutatuna anda hinnangu laiale tulemite ringile. Küsimustikud on patsiendile hõlpsasti täidetavad ning nende kasutamine ei nõua väljaõppinud intervjueerijaid. Võib eeldada, et suhteliselt väikesemahulised ja lihtsa ülesehitusega HAQ-DI ja RAQoL leiavad hea vastuvõtu ka patsientide seas, kellel puudub enesehinnanguküsimustike täitmise varasem kogemus. Hindamisvahendeid on kerge käsitleda ja skoorida ning see tagab nende sobivuse kasutamiseks erineva ülesehitusega RA-tulemiuuringutes. HAQ-DI ülemaailmne kasutatavus ja RAQoL keeleversioonide suurenev arv loovad avarad võimalused osaleda neid küsimustikke kasutavates rahvusvahelistes uurimustes.

Kuigi kvalitatiivset käsitlust on kasutatud tulemiuuringute eri etappides, puudavad seni andmed, kas kvalitatiivset meetodit saab enne kohandamist küsimustiku sobivuse hindamiseks rakendada. Meetodit võib pidada uuringus rakendatavaks, kui selle kasutamise tulemusel on võimalik teha uurimiseesmärgi kohta olulisi järeldusi, milleni on raske või pole võimalik teiste meetodite abil jõuda ning mille valiidsus on kinnitatav. Kuna Eesti RA-patsientide haiguskogemust on seni vähe uuritud, võib eeldada, et kvalitatiivne meetod pakub väärtuslikku teavet küsimustike sobivuse kohta Eesti oludes. Planeeritava kohandamise eel korraldatud kvalitatiivne uurimus annab võimaluse vältida kohalike olude jaoks sobimatu küsimustiku töömahukat ja kulukat adapteerimist. Kvalitatiivse käsitluse omadus tuua esile uuringus osalejate tõlgendusi ja vaatenurki lubaks uurida teemasid, mida kohandatav küsimustik ei käsitle, kuid mis on seotud huvipakkuva valdkonnaga, ning seeläbi anda hinnang kohandatava hindamisvahendi võimele haarata Eesti RA-patsientide jaoks olulisi uuritava tulemi aspekte. Eriti väärtuslik oleks selline teave RAQoL korral, kuna arusaama elukvaliteedist mõjutavad olulisel määral sotsiaalsed tegurid. Kvalitatiivse uurimuse tulemuste põhjal tehtud järelduste valiidsust saaks hinnata järgneva kohandamise, samuti muude meetodite abil saadud asjakohaste andmetega kõrvutamise käigus.

Seni pole uuritud erinevate adapteerimismeetodite rakendatavust Eestis. Rahvusvahelisest kirjandusest võib leida seisukohti mõlema juhtiva adapteerimismeetodi — nii tõlke-tagasitõlke kui ka kahe paneeli meetodi — kasutamise toetuseks. Arvestades erinevaid meetodeid kõrvutavate uuringute puudumist ning vajadust võtta arvesse kohalike olude võimalikke eripärasid, peaksid kohandamisstrateegia valiku soovitused Eestis põhinema kohalikes oludes korraldatud uurimuste käigus saadud andmetel. Kuna praegu puudub piisav tõendus kindla kohandamismeetodi eelistamiseks, tuleks RAQoL ja HAQ-DI adapteerimismeetodi valikul Eestis lähtuda küsimustike autorite soovitustest ja olemasolevatest andmetest eri strateegiate edukuse kohta nende hindamisvahendite uute keeleversioonide loomisel. Hinnang kohandatud küsimustike psühhomeetrilistele omadustele ning nende küsimustike aktsepteeritavus patsientide poolt lubavad otsustada, kas patsiendikeskse tulemi hindamisvahendite kohandamise meetodid on Eestis rakendatavad. Kohandamismeetodit saab rakendada, kui see võimaldab luua küsimustike versioone, mille valiidsuse, reliaabluse ja aktsepteeritavuse näitajad on vastuvõetavad.

Uurimistöö eesmärgid

Uurimistöö **eesmärgiks** oli kohandada patsiendikeskse tulemi hindamisvahendid HAQ-DI ja RAQoL kasutamiseks Eesti RA-haigetel ning anda hinnang kvalitatiivsete meetodite ja erinevate tõlkestrateegiate rakendatavusele nende küsimustike kohandamisel.

1. Uurida, kas patsiendikeskse tulemi hindamisvahendite sobivuse hindamisel saab rakendada kvalitatiivseid meetodeid, sealhulgas
   - 1.1. uurida RA-patsientide kogemusi ja arusaamu seoses nende igapäevase elu kvaliteediga;
   - 1.2. hinnata RAQoL-i teemade olulisust Eesti RA-patsientidele;
   - 1.3. hinnata tulemuste valiidsust, kõrvutades saadud andmeid küsimustiku edaspidise kohandamise käiguga ja teiste meetodite abil kogutud asjakohaste andmetega.
2. Anda hinnang mõlema adapteerimismeetodi — tõlke-tagasitõlke ja kahe paneeli meetodi — rakendatavusele patsiendikeskse tulemi hindamisvahendite kohandamisel, tuginedes
   - 2.1. nende kahe meetodi järgi kohandatud küsimustike HAQ-DI ja RAQoL reliaabluse ja valiidsuse hinnangutele;
   - 2.2. nende kahe meetodi järgi kohandatud küsimustike aktsepteeritavuse hinnangule, võttes aluseks reliaabluse ja valiidsuse hindamise käigus patisentide avaldatud arvamused.

Uurimistöö praktiliseks väljundiks oli RA-patsiendikeskse tulemi hindamise küsimustike HAQ-DI ja RAQoL eesti versioonide loomine.

Meetod

Kvalitatiivsete meetodite rakendatavus patsiendikeskse tulemi hindamisvahendite sobivuse kindlakstegemiseks

Küsimustiku RAQoL kohandamise sobivust Eestis uuriti, kasutades kvalitatiivset meetodit: temaatilist analüüsi koos põhistatud teooria elementidega. RA-patsientidega korraldati kümme poolstruktureeritud individuaalset intervjuud. Avatud lõpuga intervjuuküsimused käsitlesid igapäevaelu valdkondi, mida RAQoL hõlmas, ning samas andsid intervjueeritavatele võimaluse rääkida ka elukvaliteedi mõjuritest, mis olid neile olulised, kuid mida RAQoL ei käsitlenud. Intervjuutekstide etapiviisi kodeerimise tulemusena moodustati tuumkategooriad, mis võeti andmete analüüsimisel aluseks. Järelduste valiidsust hinnati, kõrvutades neid järgnenud kohandamisprotsessi käiguga. Samuti võrreldi saadud tulemusi RA-patsientide rahulolu-uuringu fookusgruppide andmetega; järelduste paikapidavuse kohta küsiti arvamust RA-haigete rühmalt.

Erinevate adapteerimismeetodite rakendatavus patsiendikeskse tulemi hindamisvahendite kohandamisel

Küsimustiku HAQ-DI eestindamisel juhinduti tõlke-tagasitõlke meetodil põhinevast kohandamisjuhendist. HAQ-DI eestikeelse versiooni sõnastuse arusaadavust ning küsimustele vastamise hõlpsust hindasid RA-patsiendid rühmaarutelu käigus. RAQoL tõlgiti eesti keelde kahe paneeli meetodil, tõlke arusaadavuse ning sõnastatud küsimuste olulisuse hindamiseks korraldati RA-haigetega 15 poolstruktureeritud individuaalset intervjuud.

Tõlgitud küsimustike usaldusväärsuse ja valiidsuse hindamisse kaasati 50 RA-patsienti. Uuritavad täitsid küsimustikud HAQ-DI, RAQoL ja NHP (*Nottingham Health Profile*) kahenädalase intervalliga kahel korral (esimesel ja teisel visiidil). Mõlemal visiidil hinnati ka haiguse aktiivsust. Osalejatel oli võimalik küsimustikke kirjalikult või suuliselt kommenteerida. Mõlemal visiidil andsid patsiendid hinnangu oma haiguse raskusele 4-astmelisel skaalal. Labakäte röntgenogrammi hinnati Larseni-Dale'i skoori alusel.

HAQ-DI ja RAQoL kordustesti usaldusväärsust hinnati, korreleerides esimesel ja teisel visiidil saadud skoorid. Sisereliaablust hinnati Cronbachi alfa abil. Hindamaks tuntud gruppide valiidsust moodustati rühmad patsientide endi poolt hinnatud RA raskusastme järgi. Konvergentse ja diskrimineeriva valiidsuse hindamiseks korreleeriti HAQ-DI ja RAQoL skoore omavahel ning vaadeldi korrelatsioone nimetatud skooride ja teiste uuritud näitajate vahel. RAQoL eesti versiooni ühemõõtmelisust hinnati, kasutades Raschi analüüsi.

Tulemused ja järeldused

Enamik RAQoL teemasid olid kvalitatiivse uurimuse käigus küsitletud patsientidele olulised. Intervjuude käigus ilmnes ka kolm elukvaliteedi aspekti (rolli

täitmine, suhtlemine tervishoiusüsteemiga ja eluks vajaliku muretsemine), mida küsimustik ei hõlmanud, kuid mis olid Eesti RA-patsientidele olulised. Küsimustiku adapteerimise käik ja täiendavad uuringud kinnitasid saadud tulemuste paikapidavust.

Küsimustikud RAQoL ja HAQ-DI osutusid kergest eesti keelde tõlgitavaks ning reliaabseks ja valiidseks. Psühhomeetrilisel uurimisel osalenute kommentaarid ja arvamused tõid välja enam puudusi HAQ-DI tõlkes.

1. Kvalitatiivsete meetodite kasutamine lubas enne kohandamist anda hinnangu patsiendikeskse tulemi hindamisvahendi sobivusele sihtkultuuris.
   1.1. Rakendatud kvalitatiivne metodoloogia võimaldas välja tuua RA-patsientide igapäevaelu kogemusi ja arusaamu elukvaliteedist.
   1.2. Kvalitatiivne uuring võimaldas hinnata küsimustikus RAQoL käsitletud teemade olulisust patsientidele, sealhulgas küsimustiku võimet haarata Eesti RA-patsientide jaoks olulisi elukvaliteedi aspekte.
   1.3. Kvalitatiivse uuringu põhjal tehtud järelduste valiidsus leidis kinnitust edasisel küsimustiku kohandamisel ja teistel meetoditel saadud asjakohaste andmetega kõrvutamisel.
2. Nii tõlke-tagasitõlke kui ka kahe paneeli meetod olid rakendatavad patsiendikeskse tulemi hindamisvahendite kohandamisel.
   2.1. Mõlemad meetodid võimaldasid luua küsimustike reliaabsed ja valiidsed versioonid.
   2.2. Mõlemad meetodid võimaldasid luua küsimustike versioonid, mille patsiendid võtsid hästi vastu; kahe paneeli meetodil võib olla eeliseid tõlke-tagasitõlke meetodi ees hindamisvahendi versiooni parema aktsepteeritavuse saavutamisel.

Uurimistöö praktiline väljund

Käesoleva uurimistöö tulemusena valmisid RA-patsiendikeskse tulemi hindamise küsimustike HAQ-DI ja RAQoL eesti versioonid.

Soovitused

Uuringu tulemuste alusel saab patsiendikeskse tulemi hindamisvahendite kohandamiseks sõnastada järgmised soovitused:

1. Enne kohandamist on soovitatav korraldada kvalitatiivne uurimus, et hinnata hindamisvahendi sobivust.
2. Rahvusvaheliste küsimustike kohandamisel on soovitatav järgida tunnustatud juhtnööre.
3. Reliaabluse ja valiidsuse hindamine on soovitatav korraldada kõigil Eestis kasutamiseks kohandatud instrumentidel.
4. Eesti tarvis kohandatud küsimustike ühemõõtmelisust on soovitatav hinnata, kasutades Raschi analüüsi.

14. ACKNOWLEDGEMENTS

This study was carried out at the Department of Internal Medicine of University of Tartu, Estonia. The study was supported by the Estonian Science Foundation, grant number 5239.

Above all I thank my friend and tutor, Stephen McKenna, for the support and guidance he provided.

I thank my supervisor Professor Margus Lember.

I thank my co-authors Gurkirpal Singh, Judit Strömpl, David Meads, Kadri Maimets, Ele Hanson, and Kaja Põlluste.

I thank Riina Kallikorm for introducing me to research in rheumatology.

I thank the following colleagues and friends for their contribution:

Methodological advices — Zuzana de Jong, Mati Rahu, Krista Fischer;

Radiological assessment — Helve Säre;

Assistance with the assessment of patients' status and the execution of laboratory tests — Kristi Lupkina and Asta Loogus, Lea Miidla-Leitu and Monika Kolpakov, Urve Kärtner, and Aune Avik;

Technical support — Eve Möls, Lea Tops, Silver Pukk, Ilmar Part, Talvi Adamson, Urve Pirso, and Casey Collins.

I thank the doctors and staff of the Department of Internal Medicine of Tartu University Clinics and Tartu City Polyclinic.

I thank the reviewers Anu Aluoja and Ruth Kalda.

I thank all the patients who participated in this study.

15. PUBLICATIONS

The adaptation, reliability and validity testing of the Estonian version of the Health Assessment Questionnaire's Disability Index

M. Tammaru, MD, Postgraduate Student of Internal Medicine, Department of Internal Medicine, Faculty of Medicine, University of Tartu

G. Singh, MD, Adjunct Clinical Professor of Medicine, Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Stanford, California, USA

E. Hanson, Undergraduate Student of Medicine, Faculty of Medicine, University of Tartu, Estonia

K. Maimets, Undergraduate Student of Medicine, Faculty of Medicine, University of Tartu, Estonia

Abstract

The aim of our study was the adaptation of the Health Assessment Questionnaire's Disability Index (HAQ-DI) for Estonia along with the assessment of its psychometric properties. The linguistic validation included phases of translation, back translation and testing on patients. Reliability and validity were tested on a sample of 50 rheumatoid arthritis patients by administering the HAQ-DI and the comparator instruments at two visits; disease activity and radiological stage were assessed. The participants were asked to comment on the questionnaires. The HAQ-DI was easily translatable into Estonian. It showed good test-retest reliability, internal consistency, and ability to discriminate between different levels of self-perceived severity. Comparison with the comparators and performed assessments demonstrated expected convergent and divergent validity. Still, the participants' comments revealed issues disregarded during the adaptation process. We recommend the Estonian HAQ-DI for rheumatoid arthritis clinical studies and trials; some aspects of the adaptation process are highlighted for further discussion.

Introduction

So far there have been no disability assessment tools with demonstrated psychometric properties for rheumatoid arthritis (RA) patients in Estonia. Now, when looking for an instrument to put to use in Estonian rheumatology, our preference was the adaptation of a well-known questionnaire rather than the development of an original one. Due to Estonia's small population (about 1.3 million), studies involving only our patients may suffer from a lack of statistical power, thus the future of our outcome research may be seen in international cooperative projects. Among existing instruments, our choice fell unreservedly on the Health Assessment Questionnaire's Disability Index as a prerequisite for participation in research of transnational level.

The Health Assessment Questionnaire (HAQ) has been available for researchers and clinicians for nearly a quarter of a century (1). Although originating from rheumatology, the HAQ is currently accepted as a generic instrument for patient oriented outcome assessment. The HAQ Disability Index (HAQ-DI), the most widely used part of the instrument, assesses a patient's functional ability. The instrument consists of 20 statements about daily activities (each with four possible response options) and 21 items asking about the requirement for assistive devices or help from another person. The questions are combined from eight subscales — dressing, rising, eating, walking, hygiene, reach, grip and usual activities.

In rheumatology the HAQ-DI is the most commonly used assessment tool (2) and it is rare for RA clinical studies not to report HAQ disability (3). A recently published overview of the HAQ and the HAQ-DI (4) summarizes the current knowledge on the psychometric properties of the instrument and its adaptations. Validity of the HAQ-DI has been demonstrated by a number of studies; test-retest correlations have ranged from 0.87–0.99. Disability assessed by the HAQ-DI has repeatedly shown correlation with mortality rates, progression of ageing, and health care resource utilization. The considerable number of adaptations of the HAQ-DI also serves as an indicator of the value of the instrument. It has been translated or culturally adapted into more than 60 languages and dialects; culturally adapted HAQ-DI instruments have proved to be as valid and reliable as the original questionnaire.

Basic guidelines for cross-cultural adaptation and validation of health status measures have been published (5, 6, 7) and are widely accepted. Although they have attracted some criticism (8, 9, 10), they underlie the methodology of the most of the HAQ-DI adaptations. Therefore for the HAQ-DI adaptation for Estonia, we chose to follow the principles introduced by the aforementioned guidelines.

In this paper we will describe the process of linguistic validation of the Estonian HAQ-DI and report the instrument's psychometric properties tested on a sample of RA patients. We also will introduce and discuss our experiences

acquired in the course of communication with the patients participating in the study.

Our study was approved by the Committee on Ethics in Human Research of Tartu University.

Methods

Linguistic validation

The linguistic validation of the HAQ-DI consisted of three phases. In the first phase, two native Estonian speakers, bilingual in English, independently translated the original instrument into Estonian. On the basis of these two translations we formed the pooled version, and, to ensure linguistic quality, it was reviewed by an Estonian philologist. In the second phase, the pooled version was back translated into English by a professional translator of American background. We compared the original questionnaire and the back translation for coherence and then formulated the initial Estonian version of the instrument for patient testing. The third phase, the patient testing panel, was attended by six rheumatoid arthritis (RA) outpatients of Tartu University Hospital. The participating five women and a man, aged from 42 to 67, were native speakers of Estonian. Every item of the initial Estonian version was read out loud; at the same time participants could follow the text on typed copies of the instrument. During the group discussion the patients were asked to answer two questions: 'What does this statement mean for you?' and 'Is there any other wording that enables this meaning to be expressed more clearly?' Subsequent processing took the position of the patient testing panel into account and was attended by specialists from the rheumatology department of Tartu University Hospital. It resulted in the elaboration of the Estonian version of the HAQ-DI for the subsequent psychometric properties assessment.

Assessment of reliability and validity

The procedure. Instead of a postal survey, more traditional for psychometric properties testing, we preferred to meet every patient personally in order to get the possibility to assess disease activity. The additional advantage was the opportunity to learn about patients' reactions and opinions by means of their comments on the instrument.

Fifty RA patients who had attended the outpatient and/or inpatient clinic of Tartu University Hospital from 1999 to 2003 and who were not currently hospitalized were recruited. The purpose of the study was explained to the participants. The evaluation comprised the ACR 1993 preliminary core set of disease activity measures for RA (11). We included patients' assessment of pain on the visual analog scale (VAS) and patients' global assessment of disease activity on VAS. The number of tender and swollen joints was evaluated using the 18-joint count (12), which we considered sufficiently detailed for our

adaptation study. Moreover, it had the advantage of being less burdensome for the respondents. Erythrocyte sedimentation rate and C-reactive protein values were measured. Radiography of hands performed not more than a year before was assessed for Larsen-Dale score (13, 14). Estonian versions of the Rheumatoid Arthritis Quality of Life Scale (RAQol) (15, 16), the Nottingham Health Profile (NHP) (17) and the HAQ-DI were self-administered by all participants in the aforementioned sequence. The patients were asked to feel free to comment on all three questionnaires both verbally and in writing. The assessments (with the exception of the assessment of radiological stage) were performed and all patients completed the RAQol, the NHP and the HAQ-DI twice with a 2-week interval between visits (Time 1 and Time 2). The participants were also questioned about their perceptions of disease severity at both visits, as was previously done during the adaptation of the Danish version of the HAQ-DI by Thorsen et al. (18).

Scoring the HAQ. The highest item score on the four-level difficulty scale (0 to 3) defines the score of the respective subscale. The average of the eight subscale scores is transferred to the overall HAQ-DI score on a 25-value scale from 0 (no disability) to 3 (complete disability). By the traditional scoring method, dependence on assistive devices or help from another person increases the lower subscale scores to 2. However, the traditional method has been criticized for not reflecting patients, perceptions of disability (19; 18). An alternative rating model that does not increase the score if an assistive device or physical assistance is used was suggested by Nordenskiöld and Grimby 1997. In our study we applied both the traditional and alternative methods of HAQ-DI scoring.

Statistical analysis. Test-retest reliability was estimated as the correlation between Time 1 and Time 2 scores of the HAQ-DI; Spearman rank correlation coefficients were employed. It was agreed that correlation coefficients of 0.85 or above would indicate an acceptable level of measurement error (20). Internal consistency was assessed using Cronbach's alpha coefficient; value 0.7 or above was required to indicate acceptable relatedness of items (21). Construct validity was determined by two methods. For estimating the known group validity, groups were formed by self-perceived severity of RA; Mann-Whitney U Test was applied to test differences between groups in the HAQ-DI scores. Convergent and divergent validity were estimated by testing correlations between the adapted instrument scores and other valid measures (activity assessment, radiographic stage, and the RAQol and the NHP scores); the associations were assessed by Spearman rank correlation coefficients. Statistical analysis was performed using SPSS 12.0 for Windows.

Results

Linguistic validation

No major obstacles were reported by the translators during translation of the HAQ-DI into Estonian. The two independent translations coincided in the majority of expressions; most differences resulted from the usage of synonyms, in which Estonian is well endowed. In forming the pooled version we tried to rely equally on both translations, preferring wording closer to lay Estonian language. We also accepted the suggestion of the translators to replace the original '5-pound object' with '2-kg object' in the Estonian version because the pound is not a well-known unit in Estonia. We used the help of an Estonian philologist to spruce up our translation linguistically in order to minimize the risk of misunderstandings during the back translation.

The pooled version turned out to be easily back translatable into English; no hesitations or obstacles were reported by the translator. We discovered no essential discrepancies between our back translation and the original HAQ-DI. Some differences in the expressions could be again attributed to the usage of equivalent words. For example, 'are you capable of' in the back translation corresponds to 'are you able to' in the original; 'with a degree of difficulty' to 'with some difficulty'; 'level ground' to 'flat ground'; 'everyday activities and housework' to 'errands and chores'. We considered these minor differences to be inevitable consequences of consecutive forward and back translation. No changes were implemented in the initial Estonian version of the instrument as a result of the back translation process prior to proceeding to the patient testing panel.

The testing panel worked in a tension-free atmosphere; the participants expressed high willingness to discuss the items. The questionnaire was accepted well by the test patients. The participants' comments on the majority of items revealed no ambiguity in comprehending the issues. To reach a higher level of comprehensibility, a minor change in the wording of one item was suggested by the panel; the modification was introduced to the instrument. At a following discussion that involved two experienced rheumatologists we concluded that the Estonian version of the HAQ-DI was ready for psychometric properties testing without any additional changes.

Assessment of reliability and validity

Sample. All patients who had completed the Time 1 visit attended the Time 2 visit as well. Forty five (90%) of 50 participants were female. The age of the patients ranged from 26 to 81 years, with a mean age of 57.0 years. The disease duration, counted from the actual occurrence of the symptoms, ranged from 1 to 37 years, mean 15.0 years. The relatively long mean duration of disease can be attributed to the specific sample of attendants of the university hospital, and also to the failure to include patients with a short duration of joint complaints whose diagnosis of RA had not been definitely confirmed. At Time 1, 18 patients

(36%) assessed their disease to be mild or moderate and 32 (64%) to be quite or very severe; for Time 2 the values were 27 (54%) and 23 (46%) respectively. In addition to any possible true changes in perception, the changes in assessments may partly be explained by the effect of reactivity, which is specific to the test-retest method. Reactivity refers to the situation in which the experience of being questioned for the first time makes respondents think thoroughly about the issue in focus, and in some cases, to change the initial assessment by the time of the second questioning (22). However, only in five cases the difference was greater than one point on the four point Likert scale. Moreover, the difference between disease severity assessments at the two visits was statistically not significant (Chi-square test, $p>0.05$).

Descriptive scores. Summary statistics for the Estonian HAQ-DI are presented in Table 1. The data revealed absent ceiling and minimal floor effects; this result can be considered to be as expected for an outpatient sample.

Test-retest reliability. The test-retest reliability between Time 1 and Time 2 of the Estonian version of HAQ-DI was as high as 0.91 ($p < 0.01$) for both scoring methods. Figure 1 and Figure 2 present the scatter plots illustrating scores at each time point. These results indicate the low level of random measurement error produced by the instrument.

Internal consistency. The Cronbach's alpha was 0.96 at Time 1 and 0.97 at Time 2; obtained values indicate the adequate interrelation between items.

Known groups validity. Table 2 demonstrates the Estonian HAQ-DI's ability to distinguish between patient-perceived severity groups. The instrument was able to differentiate the groups regardless of the scoring method (original or alternative) being used.

Convergent and divergent validity. The levels of association between the Estonian HAQ-DI scores and activity measures, radiological stage, the RAQol, and the six NHP sections' scores are shown in Table 3 and Table 4.

The Estonian HAQ-DI scores demonstrated no notably strong correlations with most activity measures and with radiological stage. This result is expected considering the complex nature of disability in the case of RA.

Moderately strong correlation between the Estonian HAQ-DI and the RAQol scores indicated that disability is an important, but not the only, determinant of quality of life of RA patients. For the NHP the highest level of association with the HAQ-DI score was observed in the physical mobility section; this result is expected.

Patients' comments and opinions. Mostly the Estonian HAQ-DI was received well by the participants. Some of the patients admitted that after filling the

quality of life and general health status questionnaires, they *'felt relaxed'* when working through the HAQ-DI, an instrument that had *'clear and concrete questions'*.

The concreteness of items and their functional nature were also a source of problems. *'I do not know, I have not done this for weeks!'* was the reason for a number of missing responses. Two patients just stood up and tried to perform as many tasks as it was possible in the physician's office.

Quite confusing was the time limit (over the past week) in the instructions. *'How should I answer if I felt well for three days and bad for four? Should the answer be about mornings or afternoons – they are so different?'* Confused patients resolved the problems in different ways. The majority of them chose the answers that were true at the moment of filling in the questionnaire. Some patients whose condition had changed dramatically during the past week preferred to describe their worst moments.

Patients seldom reported general problems with the response choices. There were patients who asked: *"Should I answer as it would be true for normal people or as it is true for me? I am so used to my condition that things that would be problematic for a healthy person are just normal for me."* A patient whose disease precluded him from performing most of his everyday tasks and who was assisted by his family members answered 'without any difficulty' to a number of questions. '*I cannot say that I have any difficulties or am unable to do something if I just do not do this!*'

Some items were more problematic than others.

'Open a new milk carton' was translated into Estonian as 'open a milk pack'. There is no equivalent word for 'carton' in Estonian; 'milk pack' was chosen as a widespread colloquialism. Two problems arose concerning this item, *'I do not drink milk!'* being the (already expected) first one. The second problem was more serious and indicated our carelessness during the translation process. Some patients reported different levels of difficulties for performing this task even if their condition had been stable between visits. A patient who marked 'without any difficulty' at Time 1 and 'unable to do' at Time 2 explained: *'But my wife bought a different milk pack now.'* We had forgotten about the two different types of packs: it is also possible to buy milk in plastic bags here; the effort required to open them cannot be compared. (We changed the wording to 'open a tetra pack' from the 36th patient. 'Tetra pack' is a quite widely used term in Estonia, but it could be more difficult to understand for the older generation.)

In addition to the patient's level of disability, the answer to the question about being able to 'walk outdoors on flat ground' depended on the season when the questionnaire was completed. In the winter in Estonia, which can be snowy and very icy, even a perfectly healthy person may have 'much difficulty' when walking outside.

'Take a tub bath' was the item that gave the largest number of missing responses. The reason given was unanimous: *'I do not have a tub bath'*. In the

towns people are more used to taking a shower; saunas continue to be popular in the countryside. Participants without bathing experience mostly skipped the question.

The item 'get on and off the toilet' often provoked the question: *'If I had difficulties at home and no problems at work (or when visiting friends or vice-versa) how should I answer the question?'*

'Turn faucets on and off' was one of the most problematic items. A notable number of participants had modern faucets with handles that must be raised to open the tap. Sometimes the old-fashioned faucets had been replaced because of the difficulty the patient had in making turning movements with his/her wrist, sometimes in order to make their use more comfortable for whole family, and in some modern flats new faucets were a part of the built-in design. All four responses were possible for those patients. '*Yes, I could not turn and therefore replaced the faucets*' explained those who answered 'unable to do'. *'No, after the modernization I have no problems with faucets any more'* said others and chose the answer 'without any difficulty'. *'I do not know exactly'* was the third comment and the answers accordingly were 'with some difficulty' or 'with much difficulty'. Most patients still preferred to answer something instead of skipping the item.

The item 'do chores such as vacuuming or yard work' was confusing for some patients who did not know which part of the question they should answer: *"I have no problems with vacuuming but yard work is too much for me!"*

The question about usage of long-handled appliances in the bathroom evoked comments like: *'Every decent person has such appliances in the bathroom'* or *'I would have them anyhow, with my disease or without it.'* The usage of long-handled appliances in the bathroom seemed to be rather a function of luxurious lifestyle than disability.

We translated the phrase 'help from another person' into Estonian as 'secondary help'. This expression is acceptable as equivalent to the original in Estonian but is maybe more of an 'official term'. Some patients who, according to our observations, were able to manage by themselves reported the need for secondary help for arising from bed or for hygiene. When asked to explain, they were surprised: *'But you saw, I leaned on the table when standing up.'* We learned that for some people 'secondary help' includes the 'help' of inanimate objects as well. (The wording was changed to an unequivocal exact equivalent of 'help from other people'.)

Discussion

The HAQ-DI was found to be translatable into Estonian without difficulties. No considerable problems were observed during the three phases – translation, back translation, and patient testing – of a regular linguistic validation process. Conventional tests for test-retest reliability and construct validity on a sample of

RA patients demonstrated the success of adaptation of the HAQ-DI into Estonian. The instrument showed excellent test-retest reliability and was able to distinguish between different levels of self-perceived disease severity. Moderately strong correlation between the Estonian HAQ-DI and the RAQol scores revealed in our study corresponded to our expectations considering disability as important, but not the only, determinant of RA patients' quality of life. Among the NHP sections, the HAQ-DI score demonstrated the highest level of association with physical mobility; this result can also be viewed as positive evidence of the instrument's validity. Absence of any notable strong correlation between the Estonian HAQ-DI score and most activity measures and radiological stage was expected. This consistent with the idea that the RA disability is a result of a combination of various impairments rather than a reflex of any specific measurable change. (As an exception, the patient's assessment of disease activity showed quite a high level of association with the HAQ-DI score. In our opinion it could be argued whether this measure, as proposed by ACR in 1993, estimates the disease activity, or rather the patient's perception of disease impact.) The results of psychometric properties testing were similar to those reported for other language versions. Our study revealed no substantial differences between the psychometric properties of original and alternative scoring methods of the instrument. It allows us to expect that the Estonian version of HAQ-DI is suitable for applying in RA clinical studies and trials using either scoring methods, as dependent on researchers' preference, and study design.

Still, some additional information was gathered from patients' comments on the instrument during the psychometric properties survey. Although this information came too late to have considerable influence on the process of adaptation, we would like to discuss the three lessons we learned from it.

The first lesson we learned: patients should be involved in adaptation process more intensively and earlier. In reviewing our study we should face the issue of previously not having comprehended the patients well enough. Strictly following the cross-cultural adaptation guidelines, we involved six patients in a discussion panel. A wording inaccuracy was discovered by the panel, but in general the Estonian HAQ-DI was accepted unanimously. In our opinion, the method of patient participation could be a point of further discussion here. We also have experience with the adaptation of the RAQol, where we followed different adaptation guidelines that had been proposed by the questionnaire's authors (8) and whose applicability for instrument adaptation procedure in Eastern Europe had been demonstrated (23). The main differences were the inclusion of a 'lay panel', where healthy people of below average educational level discussed the wording of items and helped to bring them down to earth, and a different form of patient participation. Instead of a panel we conducted semi-structured interviews with patients, where we discussed the meaning and format of items in depth. Later the summarization of patients' opinions allowed us to make changes in the questionnaire before entering the reliability and

construct validity survey. Though this did not preclude all misunderstandings during the survey, in general we were ready for them: instead of surprising us, patients' comments gave us the possibility to test our ideas about the strengths and weaknesses of the translation.

The changes in the Estonian HAQ-DI wording carried out during the psychometric properties survey described above can not affect the reliability and construct validity scores significantly. Nevertheless, if we had carried out the changes beforehand, the reported results would have been more precise.

The second lesson we learned: we should try to make use of every possible way to learn more about the instrument. We are used to trusting statistics, but numbers can not capture all significant aspects of the usage of a questionnaire, which can be understood as a form of communication as well. Despite excellent reliability and construct validity scores, we encountered some unexpected questions from patients and comprehension problems when observing the instrument's administration process.

Our possibilities to change the wording or design of the questionnaire are limited (indeed, we need the HAQ-DI, not any home-built instrument), but having more varied information allows us to handle it and understand the results of surveys in a more flexible way. In some cases the new information we receive about the questionnaire permits us to prepare ourselves for patients' questions, for which we should find answers for ourselves first. Shared knowledge on the probable obstacles and a consensus of opinion on how to overcome them should confirm the researcher's competence in the eyes of patients and contribute to the preciseness of the results of further surveys. (Some very useful suggestions can currently be found on the Stanford University website http://aramis.stanford.edu/HAQ.html.)

The third lesson also originates from the previous – there should be closer communication between researchers working on the instrument. The nature of disability has not changed during past 25 years but some devices (faucets, tube baths, vacuum cleaners) described in the HAQ-DI have. It would be totally wrong to claim that the HAQ-DI is out of date but the relevance of some items has diminished. These items have a different meaning than they had years ago and might be a little confusing for the patients. Although it may be not significant enough to alter the coefficients of reliability and construct validity, we could lose in face and content validity and gain in missing responses and impreciseness if we continue using these items in their original form. Some reports of adaptations discuss the problems arising from the meaning of concrete items and describe the changes carried out in the wordings (for further references see Bruce 2003). But locally executed changes, although improving face and content validity for one language version, could diminish the comparability of the results of international surveys. How can we ensure the comparability of the translations of different items? Differential Items Functioning (DIF) analysis is used as a mechanism to evaluate the cross-cultural validity of outcome measures. This approach is set within the unidimensional

Rasch measurement model, which assumes the probability of an item to be affirmed by a person, being only a function of the difference between the person's ability and the difficulty of the item (24, 25). Hence, individuals with the same level of ability will have the same probability to affirm an item irrespectively of any group membership. This assumption allows us to examine the cross-cultural validity by contrasting the item response functions across cultures. The adequate fit to the Rasch model of the HAQ-DI both at the item and subscale levels for RA patients has been demonstrated (26); DIF analysis has been successfully applied for assessment of cross-cultural validity of the eight subscales of the HAQ-DI (27). Still, this approach does not enable issues relevant for the adaptation process to be resolved. It points out the differences in the functioning of items of different adaptations, but it does not offer any explanations for their existence. Also the prerequisite for application of DIF analysis is the existence of an adapted version of the questionnaire. It deprives us of the opportunity to learn about potential hindrances beforehand, in order to avoid them in the process of adaptation. We believe that gathering of evidence, open discussion during the adaptation processes and shared knowledge could lead to decisions which would contribute to preserving comparability of the different HAQ-DI adaptations.

It can be concluded that the adaptation of the Health Assessment Questionnaire's Disability Index into Estonian is complete; the instrument showed good psychometric properties on a sample of RA patients. Still, our study revealed some aspects of the adaptation process that we consider to merit further discussion.

Acknowledgements

The authors wish to thank Mrs. Talvi Adamson, Mr. Silver Pukk and Mr. Ilmar Part for linguistic consultation and Dr. Helve Säre for the help with radiological assessment. The study was supported by the Estonian Science Foundation grant number 5239.

References

1. Fries JF, Spitz P, Kraines RG, Holman HR (1980) Measurement of patient outcome in arthritis. Arthritis Rheum 23:137–145
2. Ramey DR, Raynauld JP, Fries JF (1992) The health assessment questionnaire 1992: status and review. Arthritis Care Res 5:119–129
3. Wolfe F (2000) A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 43:2751–2761
4. Bruce B, Fries JF (2003) The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 30:167–178

5. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 46:1417–1432
6. Guillemin F (1995) Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol 24:61–63
7. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–3191
8. Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna SP (1991) Cross-cultural adaptation of health measures. Health Policy 19: 33–44
9. Leplège A, Verdier A (1995) The adaptation of health status measures: methodological aspects of the translation procedure. In: Shumaker S, Berzon R, editors. International Use and Performance of Health-Related Quality of Life Instruments. Oxford: Oxford Rapid Communication; 93–101
10. da Mota Falcao D, Ciconelli RM, Ferraz MB (2003) Translation and cultural adaptation of quality of life questionnaires: an evaluation of methodology. J Rheumatol 30:379–385
11. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, Furst D, Goldsmith C, Kieszak S, Lightfood R, Paulus H, Tugwell P, Weinblatt M, Widmark R, Williams JH, Wolfe F (1993) The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 36:729–740
12. Wolfe F, O'Dell JR, Kavanaugh A, Wilske K, Pincus T (2001) Evaluating severity and status in rheumatoid arthritis. J Rheumatol 28:1453–1462
13. Larsen A, Dale K, Eek M (1977) Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn 18:481–491
14. Larsen A (1995) How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies. J Rheumatol 22:1974–1975
15. Whalley D, McKenna SP, de Jong Z, van der Heijde D (1997) Quality of life in rheumatoid arthritis. Br J Rheumatol 36: 884–888
16. de Jong Z, van der Heijde D, McKenna SP, Whalley D (1997) The reliability and validity of the RAQoL. Br J Rheumatol 36:878–883
17. Hunt SM, McEwen J, McKenna SP (1986) Measuring Health Status. London: Croom Helm
18. Thorsen H, Hansen TM, McKenna SP, Sorensen SF, Whalley D (2001) Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the Rheumatoid Arthritis Quality of Life Scale (RAQoL). Scand J Rheumatol 30:103–109
19. Nordenskiöld U, Grimby G (1997) Assessments of disability in women with rheumatoid arthritis in relation to grip force and pain. Disabil Rehabil 19:13–19
20. Weiner EA, Stewart BJ (1984) Assessing individuals. Boston: Little Brown
21. Steiner DL, Norman GR (1995) Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press
22. Murphy KR, Davidshofer CO (2005) Psychological testing. Principles and applications. Sixth edition. Upper Saddle River, New Jersey: Pearson Educational International

23. Lovas K, Kalo Z, McKenna SP, Whalley D, Pentek M, Genti G (2003) Establishing a standard for patient-completed instrument adaptations in Eastern Europe: experience with the Nottingham Health Profile in Hungary. Health Policy 63:49–61
24. Rasch G (1980) Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press
25. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, Lawton G, Simone A, Carter J, Lundgren-Nilsson A, Tripolski M, Ring H, Biering-Sorensen F, Marincek C, Burger H, Phillips S (2004) Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 42: I37–48
26. Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA (1996) Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 35:574–578
27. Kucukdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A (2004) Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 51:14–19

Table 1. Summary statistics for the Estonian HAQ-DI scores.

	HAQ-DI, original scoring		HAQ-DI, alternative scoring	
	Time 1 (n = 50)	Time 2 (n = 50)	Time 1 (n = 50)	Time 2 (n = 50)
Mean	1.42	1.42	1.17	1.18
Std. Deviation	0.72	0.72	0.69	0.69
Median	1.50	1.50	1.25	1.13
Interquartile range	0.88 – 1.91	1.00 – 1.88	0.75 – 1.63	0.75 – 1.66
Range	0.00 – 2.75	0.00 – 2.75	0.00 – 2.75	0.00 – 2.75
% scoring minimum (0)	6.0	6.0	8.0	6.0
% scoring maximum (3)	0.0	0.0	0.0	0.0

Table 2. Known group validity: ability of the Estonian HAQ-DI to distinguish self-perceived severity groups. The results are presented for both scoring methods at Time 1 and Time 2.

		HAQ-DI, original scoring		HAQ-DI, alternative scoring	
		Time 1	Time 2	Time 1	Time 2
Severity group Mild/Moderate	n	18	27	18	27
	Median	0.81	1.15	0.69	0.93
	Interquartile range	0.22 – 1.50	0.50 – 1.88	0.13 – 1.13	0.25 – 1.38
Severity group Quite/Very Severe	n	32	23	32	23
	Median	1.88	1.73	1.50	1.47
	Interquartile range	1.31 – 2.00	1.38 – 2.00	1.00 – 1.97	1.00 – 1.75
Groups' difference	Mann-Whitney U	91.5	180.5	113.5	173.0
	p	< 0.01	< 0.05	< 0.01	< 0.01

Table 3. Convergent and divergent validity: correlations between the Estonian HAQ-DI scores and activity measures, radiological stage, the RAQol scores and the NHP section scores, Time 1.

	Spearman rank correlation coefficients	
	HAQ-DI, original scoring	HAQ-DI, alternative scoring
Patient's assessment of disease activity (n = 50)	0.68***	0.69***
Patient's assessment of pain (n = 50)	0.58***	0.57***
SR (n = 47)	0.17	0.25*
CRP (n = 45)	0.21	0.25*
Tender joint count (n = 50)	0.45***	0.47***
Swollen joint count (n = 50)	0.22	0.30**
Larsen-Dale score (n = 50)	0.21	0.19
RAQol (n = 50)	0.63***	0.64***
energy level, NHP (n = 49)	0.39***	0.41***
pain, NHP (n = 47)	0.55***	0.58***
emotional reactions, NHP (n = 49)	0.16	0.19
sleep disturbance, NHP (n = 49)	0.19	0.24*
social isolation, NHP (n = 47)	0.26*	0.24*
physical mobility, NHP (n = 49)	0,76***	0,73***

*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.1$

Table 4. Correlations between the Estonian HAQ-DI scores and activity measures, radiological stage, the RAQol scores and the NHP section scores, Time 2.

	Spearman rank correlation coefficients	
	HAQ-DI, original scoring	HAQ-DI, alternative scoring
Patient's assessment of disease activity (n = 50)	0.64***	0.62***
Patient's assessment of pain (n = 50)	0.54***	0.53***
SR (n = 47)	0.23	0.23
CRP (n = 47)	0.17	0.18
Tender joint count (n = 50)	0.53***	0.53***
Swollen joint count (n = 50)	0.35***	0.39***
Larsen-Dale score (n = 50)	0.21	0.23
RAQol (n = 50)	0.67***	0.65***
energy level, NHP (n = 49)	0.54***	0.59***
pain, NHP (n = 49)	0.56***	0.49***
emotional reactions, NHP (n = 47)	0.27*	0.28*
sleep disturbance, NHP (n = 49)	0.03	0.03
social isolation, NHP (n = 47)	0.39***	0.45***
physical mobility, NHP (n = 46)	0.77***	0.71***

*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.1$

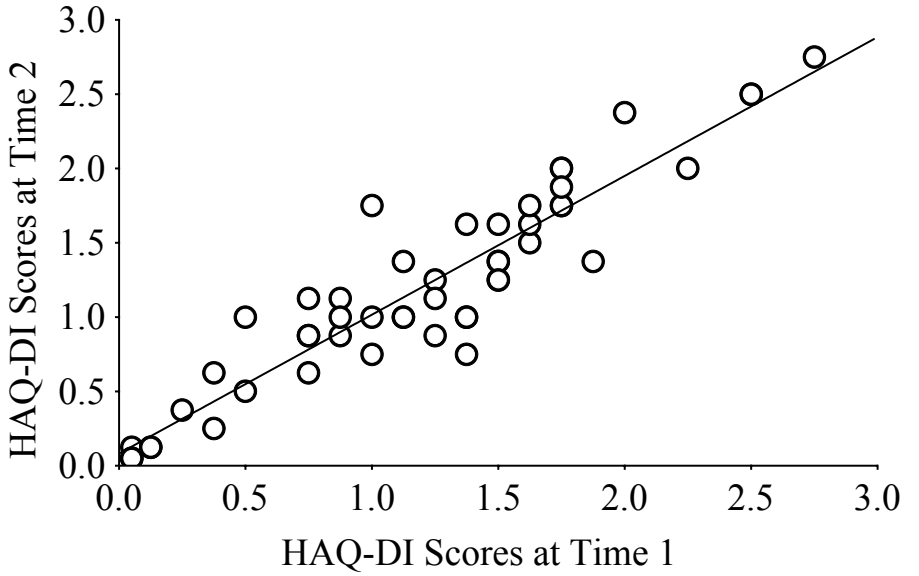

Figure 1. Scatter graph illustrating test-retest reliability of the Estonian HAQ-DI (original scoring).

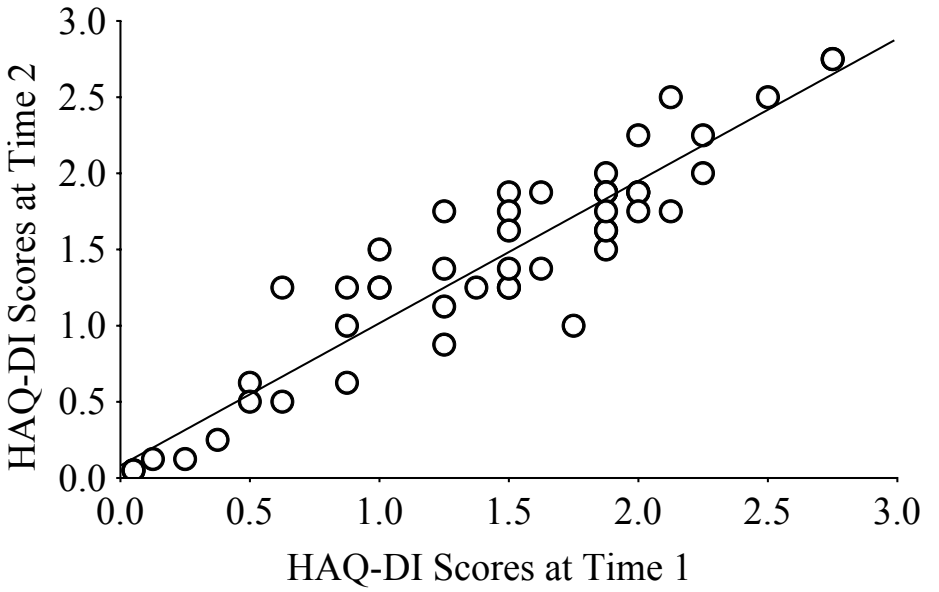

Figure 2. Scatter graph illustrating test-retest reliability of the Estonian HAQ-DI (alternative scoring).

Cultural adaptation of quality of life measures

Marika Tammaru[1], Margus Lember[1], Kaja Polluste[1], Stephen P McKenna[2]

[1] Department of Internal Medicine, Faculty of Medicine, University of Tartu
[2] Galen Research, Manchester, UK

Dear Sir,

We read with interest the article 'The Rheumatoid Arthritis Quality of Life (RAQoL) for Sweden: adaptation and validation' by Hedin et al, that was published in the *Scandinavian Journal of Rheumatology* (1). A new language version of this internationally accepted outcome measure opens opportunities to broaden multinational research on the quality of life (QoL) of rheumatoid arthritis (RA) patients. This official version of Swedish RAQoL, produced by the dual panel translation method, has shown psychometric properties similar to those demonstrated for the Estonian version of the RAQoL (2). Due to the geographical proximity and traditions of scientific cooperation, Sweden is one of the countries Estonia would be expected to join with in medical research projects. Therefore, the development of a Swedish version of the RAQoL that employed the approved adaptation procedure is of importance to rheumatological research in Estonia.

Despite the proximity of Sweden and Estonia there are considerable differences between the countries in terms of history and culture which should be considered when comparative QoL research is planned. Different cultures and societies have unique concepts of QoL. This diversity of concepts can confound QoL research in the health sciences. Before selecting a QoL tool developed in and for another country, it is necessary to determine how the instrument will work in the context of the proposed research study. An instrument that has performed excellently in the country of origin can loose validity when applied in a different social context. Standard validation processes are less likely to be of value where the local QoL concept differs from that of the culture where the measure was developed.

Ten semi-structured qualitative interviews with RA patients were conducted during the development of the Estonian version of the RAQoL to assess the suitability of the instrument. Thematic analysis conducted on the generated data indicated that the measure was highly relevant to patients; it was concluded that the RAQoL was an appropriate measure of QoL for use with Estonian RA patients (3).

However, three QoL issues that were important for Estonian RA patients were not covered by the RAQoL. These were changes in role performance, communication with the medical system and financial issues related to obtaining the necessities of life. An assumption was made that failure to assess

these specific issues might lead to an underestimation of the impact of RA on Estonian patients' QoL.

To determine the importance of the missing issues for Estonian RA patients, two additional validation strategies were applied. First, analysis was undertaken of data derived from a separate study that had employed six focus-groups to investigate RA patients' satisfaction with health care. The transcripts of these meetings were scrutinized for expressions supporting the significance of the three QoL issues independently by two researchers. Secondly, the results of these analyses were fed back to four RA patients. These patients were asked to express their opinion about the validity of the inferences drawn from the analysis and to describe their own experiences of these issues.

Both strategies clearly demonstrated the validity of the conclusions drawn from the semi-structured patient interviews. Insecurity and uncertainty surrounding communication with the medical system was the central theme of all six focus-groups. Lack of funds due to reduction in income and increase in expenditure resulting from RA were discussed in most groups. Fewer comments were found regarding the importance of changes in role function. All four interviewees agreed that these were important issues and offered relevant examples from their own experience.

It is recommended that qualitative assessments are employed prior to future instrument adaptations in order to provide a more thorough evaluation of the cultural suitability of potential QoL measures. It would also avoid the mechanical acceptance of instruments because they are well known or have performed well in other cultures.

While differences in culture are relatively minor between Western and Eastern Europe, the importance of qualitative studies of instrument face and content validity are even more necessary when adapting Western measures for use in Asian or African cultures. Despite this, many such adaptations are undertaken without adequate assessment of the suitability of the measure's content for the new culture. This can have serious consequences for the validity of the new language versions produced (4,5).

References:

1. Hedin PJ, McKenna SP, Meads DM. The Rheumatoid Arthritis Quality of Life (RAQoL) for Sweden: adaptation and validation. Scand J Rheumatol 2006;35:117–23.
2. Tammaru M, McKenna SP, Meads DM, Maimets K, Hanson E. Adaptation of the rheumatoid arthritis quality of life scale for Estonia. Rheumatol Int 2006;26:655–62.
3. Tammaru M, Strompl J, Maimets K, Hanson E. The value of the qualitative method for adaptation of a disease-specific quality of life assessment instrument: the case of the Rheumatoid Arthritis Quality of Life Scale (RAQoL) in Estonia. Health Qual Life Outcomes 2004;2:69.

4. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol. 1998;51:1037–44.
5. Cheung YB, Thumboo J. Developing Health-Related Quality-of-Life Instruments for Use in Asia: The Issues. Pharmacoeconomics 2006;24:643–50.

CURRICULUM VITAE

Marika Tammaru

Citizenship: Estonian
Born: April 29, 1973 in Tartu, Estonia
Address: Näituse 17, Tartu, 50409
Phone: +372 5192 4623
E-mail: marika.tammaru@kliinikum.ee

Education

1980–1991 Tartu Secondary School No 10
1992–1999 Medical Faculty, University of Tartu
1999–2000 Internship, Tartu University Clinics
2001–2006 PhD studies, Department of Internal Medicine, University of Tartu

Professional employment

2005–cont. Researcher, Department of Internal Medicine, University of Tartu

Scientific work

The main subject of the research work has concerned methodologies of adaptation of patient-reported outcome measures and applicability of qualitative methods for outcome research.

Member of Estonian Society of Rheumatology

ELULOOKIRJELDUS

Marika Tammaru

Kodakondsus: Eesti
Sünd: 29.aprillil 1973 Tartus
Aadress: Näituse 17, Tartu, 50409
Tel: +372 5192 4623
E-mail: marika.tammaru@kliinikum.ee

Haridus

1980–1991 Tartu 10. Keskkool
1992–1999 Tartu Ülikooli Arstiteaduskond, arstiteaduse eriala
1999–2000 Tartu Ülikooli Kliinikum, internatuur
2001–2006 Tartu Ülikooli Sisekliinik, doktorantuur

Teenistuskäik

2005– Tartu Ülikooli Sisekliinik, spetsialist

Teadustegevus

Peamiseks uurimisvaldkonnaks on patsiendikeskse tulemi hindamisvahendite kohandamise metoodika, kvalitatiivsete meetodite rakendatavus tulemiuuringutes.

Eesti Reumatoloogide Seltsi liige