Health-related quality of life in dermatology: measurement, interpretation and application

Prinsen, C.A.C.

Citation for published version (APA):
Measurement of health-related quality of life in dermatological research and practice: outcome of the EADV Taskforce on Quality of Life

CAC Prinsen
J de Korte
M Augustin
F Sampogna
MS Salek
MKA Basra
EA Holm
TEC Nijsten

on behalf of the EADV Taskforce on Quality of Life

ABSTRACT

In the last decade, the importance of the measurement of health-related quality of life (HRQoL) has grown significantly. Today, HRQoL measurement is generally considered to be important in clinical trials, in the assessment of disease severity, in patient management and in the field of health economics. Therefore, a good understanding of the concept of HRQoL and its measurement instruments is a prerequisite for both researchers and clinicians. The European Academy for Dermatology and Venereology (EADV) Taskforce on Quality of Life encourages the application of HRQoL instruments in research and clinical practice, and with this manuscript, the Taskforce aims to contribute to the quality of this application. In dermatology, a large number of HRQoL instruments exist and herewith, we summarize the most commonly used generic and dermatology-specific HRQoL instruments. Information is given on the most important psychometric characteristics of these instruments, including: scale structure, reliability, validity and responsiveness. Furthermore, a flow chart is provided to support researchers and clinicians in selecting an existing instrument or, in case an appropriate instrument does not exist, in finding alternative solutions. The present manuscript is the first of a series of manuscripts to be written on behalf of the EADV Taskforce on Quality of Life, aiming to contribute to the scientific knowledge and measurement of patient-reported outcomes in dermatological research and practice.
INTRODUCTION

The ultimate goal of treatment is to cure a disease. However, in many chronic skin diseases, such as atopic dermatitis and psoriasis, dermatological treatment can only offer a temporary suppression or remission of symptoms. Herewith, treatment and therapeutic efforts are increasingly directed towards both a decrease in disease severity and an increase in patients’ health-related quality of life (HRQoL).

In accordance with the definition of the World Health Organization, HRQoL is defined as patients’ evaluation of the impact of disease and treatment on their physical, psychological and social functioning and well-being.1-4 It has become an essential outcome parameter in clinical research, clinical practice and health care management.5,6 In randomized controlled clinical trials, HRQoL is often required by the regulatory authorities as a secondary outcome measure, and increasingly, it has been considered as a primary endpoint. The regulatory and reimbursement approval of a treatment are also based on indices that express gain in terms of HRQoL improvement, e.g., biologic treatment. Moreover, the European S3-Guidelines for the treatment of psoriasis (2009) recommend HRQoL measurement in all patients who are candidates for photo, (chemo), therapy and systemic drugs, to monitor HRQoL during treatment, and to consider it as an important outcome parameter.7 This evolution may also imply that HRQoL assessment will be integrated in measuring the quality of provided care. Therefore, a good understanding of the concept HRQoL has become essential.

The European Academy for Dermatology and Venereology (EADV) Taskforce on Quality of Life was established in 2008 by Professor A.Y. Finlay and Dr. T. Schaefer on invitation by Professor J. Ring, the former president of the EADV. This Taskforce aims to contribute to the scientific knowledge and measurement of patient-reported outcomes in dermatological research and practice. With this first manuscript, the Taskforce provides an introduction to the concept and methodology of HRQoL and its measurement instruments, including background information on psychometrics.

HEALTH-RELATED QUALITY OF LIFE INSTRUMENTS

HRQoL instruments are questionnaires consisting of a number of items (i.e., questions). Response options are most often on a multiple response scale, for instance ranging from ‘never’ to ‘all the time’.8 This results in one or more (domain) scores reflecting the impact of disease on HRQoL. In dermatology, HRQoL is most commonly assessed by means of (i) generic instruments and/or (ii) specific instruments, including dermatology-specific and disease-specific instruments.

Generic instruments

Generic instruments can be used for the measurement of HRQoL in all kinds of diseases. Examples of well-established generic HRQoL instruments are the EuroQol EQ-5D,9 the Medical Outcomes Study 36-item Short Form Health Survey (SF-36),10,11 the Nottingham Health Profile (NHP),12-14 the Sickness Impact Profile (SIP),15 and the World Health Organization Quality of Life assessment (WHOQOL).16,17
The advantage of generic instruments is that they are applicable to patients with various conditions. Previous research with the SF-36, for example by Rapp et al., has shown comparisons across diseases and against the general population. Over the years, these comparisons have been acknowledged, and are often referred to. Generic instruments, however, may not have been designed with reference to dermatology, may not focus on all areas of interest of a specific disease, and as a result, may not capture issues that are most important to patients with dermatological conditions. To this aim, specific instruments are needed.

**Specific instruments**

Among specific instruments, dermatology-specific and disease-specific instruments can be distinguished. Dermatology-specific HRQoL instruments, such as the Dermatology Life Quality Index (DLQI), the Dermatology Quality of Life Scales (DQOLS), the Dermatology-Specific Quality of Life instrument (DSQL), the Skinindex-29, and -17, are developed for the assessment of HRQoL in skin diseases. They assess domains and aspects of HRQoL particularly important to patients with skin diseases. In theory, dermatology-specific instruments are applicable in (chronic) skin diseases, thereby, allowing comparisons between them. However, patients with different skin diseases may interpret and respond to items in different ways. Therefore, one should interpret the results of such comparisons with caution.

Disease-specific instruments, such as the Melasma Quality of Life Scale (MELASQOL), the Psoriasis Disability Index, the Psoriasis Index of Quality of Life (PSORIQoL), the Quality of Life Index for Atopic Dermatitis (QoLIAD) and the RosaQoL, are developed for, and limited to, a specific skin disease. Disease-specific instruments can give a better insight into the specific constellations of particular skin diseases and may detect even more specific aspects on HRQoL or disability.

Today, a very large number of HRQoL instruments exist. In this manuscript, however, we restrict ourselves to the most commonly used generic- and dermatology-specific instruments in adults (Table 1). This table is based on De Korte et al. who systematically reviewed the quality of generic and dermatology-specific instruments. The EADV Taskforce is currently working on a series of manuscripts, and will pay specific attention to disease-specific instruments in a subsequent manuscript.

**PSYCHOMETRIC CHARACTERISTICS**

Psychometrics involves the application of statistical techniques to test the measurement properties of an instrument. Of the many psychometric methods that exist, the EADV Taskforce believes that at least basic knowledge of psychometrics is important, as well as of basic principles of the related statistical tests (Table 2).

**Classical test theory**

Classical test theory (CTT) is the most widely used and commonly known measurement theory to test an instrument. Within the CTT framework, the following psychometric characteristics are considered to be the minimal prerequisites: scale structure, reliability, validity and responsiveness.
Scale structure
The structure of a set of items refers to the extent to which items belong together, representing a certain construct (such as HRQoL), and can be tested by e.g., factor analysis. Factor analysis is based on item correlations: if the factor loading (i.e., correlation) of an item is >0.40 it can be considered to load sufficiently on a specific construct; items with factor loadings <0.40 can be removed from the instrument, or belong to another construct, as they do not cover the intended construct.6,20,37 Among factor analysis, confirmatory and exploratory factor analysis can be distinguished.

Confirmatory factor analysis is a hypothesis-testing technique. Items reflecting a certain construct should be tested on its uni-dimensionality.6

Exploratory factor analysis is a data-focused technique suitable for generating hypotheses about the structure of the data. It basically clusters items together on the basis of correlation tests that seem to relate to each other and represent a certain construct.6

Item response theory (IRT) analysis can be used to test the uni-dimensionality of a construct in a more sophisticated way as part of the modern test theory model approach (see below).

Reliability
Reliability is defined as measurement precision. Two main forms of reliability exist: internal consistency and test/retest reliability.

Internal consistency is the degree to which the items of a domain of an instrument are measuring the same construct. To examine internal consistency Cronbach’s $\alpha$ can be estimated. Cronbach’s $\alpha$ varies between 0 and 1.0 and should be between 0.70 and 0.90; $\alpha<0.70$ suggests that the items of a domain assess different constructs; $\alpha>0.90$ suggests item redundancy.6

Test/retest reliability is the extent to which scores of an instrument are stable over time. To examine test/retest reliability the instrument under study is administered on two separate occasions, with a time interval that is sufficiently short to assume that the underlying condition is unlikely to have changed, but long enough that patients do not remember their previous answers.36,38 The correlation between the two separate measurements can best be computed by the Intraclass Correlation Coefficient (ICC) and varies between 0 and 1.0. The closer the coefficient is to 1.0, the higher the reliability of the instrument under study. Ideally, it should be above 0.80, which is indicative for a high degree of reliability. Nevertheless, a correlation coefficient above 0.70 is generally considered to be sufficient.6,36

Validity
Validity refers to the degree to which an instrument actually measures what it is intended to measure (i.e., the accuracy of an instrument). Validity can be subdivided into three main aspects: construct validity, content validity and criterion validity.6,20,36

Construct validity is the extent to which an instrument measures an intended (hypothetical) construct, for example: patients with a higher degree of disease severity may have a lower degree of HRQoL than those with a lower degree of disease severity.20 This uni-dimensionality can be tested by factor analysis or IRT analysis (see below). Convergent validity (i.e., the extent to which the instrument under study is correlated with other instruments of the same
Table 1. Examples of most commonly used generic and dermatology-specific HRQoL instruments in dermatological research and practice.*

<table>
<thead>
<tr>
<th>Generic HRQoL instruments</th>
<th>Total number of items</th>
<th>Total number of domains</th>
<th>Main psychometric issues</th>
<th>Completion time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EuroQol Quality of Life Scale (EQ-5D)</strong></td>
<td>Five items, including a visual analogue scale (VAS)</td>
<td>Three domains: - Physical - Mental - Social functioning</td>
<td>The EQ-5D appeared to be not very sensitive to differences associated with minor morbidity, such as skin diseases and scores suffer from ceiling effects</td>
<td>1</td>
</tr>
<tr>
<td><strong>Medical Outcomes Study 36-item Short Form Health Survey (SF-36)</strong></td>
<td>36 Items, including a single item on perceived change in health (health transition)</td>
<td>Eight domains: - Physical functioning - Role limitations due to physical problems - Bodily pain - General health - Vitality - Social functioning - Role limitations due to emotional problems - Mental health</td>
<td>Internal consistency, test/retest reliability, construct validity and responsiveness were tested (although its scale structure and test/retest reliability may be somewhat controversial)</td>
<td>5</td>
</tr>
<tr>
<td><strong>Nottingham Health Profile (NHP)</strong></td>
<td>38 Items, including seven single items on daily function</td>
<td>Six domains: - Physical mobility - Sleep - Pain - Energy level - Emotional reactions - Social isolation</td>
<td>Internal consistency and test/retest reliability, construct validity and responsiveness were tested</td>
<td>5-10</td>
</tr>
<tr>
<td><strong>Sickness Impact Profile (SIP)</strong></td>
<td>136 Items</td>
<td>12 domains: Physical dimension: - Ambulation - Mobility - Body care and movement Psychosocial dimension: - Social interaction - Communication - Emotional behaviour - Alertness behaviour Independent categories: - Sleep and rest - Eating - Work - Home management - Recreation and pastime</td>
<td>Internal consistency and test/retest reliability, construct validity and responsiveness were tested (although construct validity tested by factor analysis has not been documented). Its interpretability is not well documented</td>
<td>20-30</td>
</tr>
<tr>
<td><strong>World Health Organization Quality of Life assessment (WHOQOL-100)</strong></td>
<td>100 Items, divided into 24 facets, with each four items and four general questions</td>
<td>Conceptually, the WHOQOL-100 has six domains but two (Level of independence and Spirituality) could not be confirmed by factor analysis: - Physical - Psychological - Level of independence - Social relationships - Environment - Spirituality</td>
<td>Internal consistency and test/retest reliability, construct validity and responsiveness were tested. The interpretability of the obtained scores is not documented, except for the normative data for the general Danish population.</td>
<td>30</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Dermatology-specific HRQoL instruments</th>
<th>Total number of items</th>
<th>Total number of domains</th>
<th>Main psychometric issues</th>
<th>Completion time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dermatology Life Quality Index (DLQI)(^1)</td>
<td>21 items, most commonly used instrument</td>
<td>No domains</td>
<td>Internal consistency and test/retest reliability, construct validity and responsiveness were tested, as well as the interpretability</td>
<td>2</td>
</tr>
<tr>
<td>Dermatology Quality of Life Scales (DQoLS)(^2)</td>
<td>22 items, developed to assess the impact of skin diseases on patients’ psychosocial state and everyday activities</td>
<td>Three domains: - Dermatologic symptoms - Physical activities (subscales: embarrassment, despair, irritableness, distress) - Psychosocial state (subscales: everyday, summer, social, and sexual activities)</td>
<td>Internal consistency and test/retest reliability, construct validity, and responsiveness were tested</td>
<td>5</td>
</tr>
<tr>
<td>Dermatology Specific Quality of Life instrument (DSQL)(^3)</td>
<td>52 and 53 items, the 52-item instrument was originally developed for contact dermatitis; the 53-item instrument for acne</td>
<td>Seven domains: - Physical symptoms - Activities of daily living - Social functioning - Work/school performance - Self-perception - General mental health (SF-36) - Vitality (SF-36)</td>
<td>Internal consistency and test/retest reliability, construct validity and responsiveness were tested</td>
<td>15</td>
</tr>
<tr>
<td>Skindex-29(^4)</td>
<td>29 items, a multi-dimensional HRQoL instrument, plus one extra item on adverse effects of treatment (item 18)</td>
<td>Three domains: - Symptoms - Emotions - Functioning</td>
<td>Internal consistency and test/retest reliability, construct validity, content validity and responsiveness were tested, as well as the interpretability</td>
<td>5</td>
</tr>
<tr>
<td>Skindex-17(^7)</td>
<td>17 items, a Rasch reduced version of the Skindex-29</td>
<td>Two domains: - Psychosocial - Symptoms</td>
<td>Internal consistency, test/retest reliability, construct validity and responsiveness were tested, using existing data of the Skindex-29</td>
<td>2</td>
</tr>
</tbody>
</table>

* With approval from the authors, this table was based on the reviews performed by De Korte et al. and Both et al.\(^{19,35}\)
Table 2. Important psychometric characteristics of HRQoL instruments.

<table>
<thead>
<tr>
<th>Psychometric characteristics</th>
<th>Definition</th>
<th>Examples of applicable statistical tests*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classical test theory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale structure*</td>
<td>The extent to which items belong together, representing a particular domain of a certain construct</td>
<td>Factor analysis is most commonly used to test the uni-dimensionality of the construct. Factor analysis can be distinguished into: - Exploratory factor analysis - Confirmatory factor analysis Factor loading should be &gt;0.40</td>
</tr>
<tr>
<td>Reliability*</td>
<td>Measurement precision</td>
<td>To establish the reliability of an instrument, internal consistency reliability and test/retest reliability can be assessed. Cronbach’s alpha (α) should be between 0.70 and 0.90</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>The extent to which items of a domain of an instrument are measuring the same construct</td>
<td>Intraclass Correlation Coefficient: &gt;0.80 is indicative for a high test/retest reliability &gt;0.70 is indicative for a good test-retest reliability</td>
</tr>
<tr>
<td>Test/retest reliability</td>
<td>The extent to which scores of an instrument are stable over time</td>
<td></td>
</tr>
<tr>
<td>Validity*</td>
<td>The degree to which the instrument actually measures what it is intended to measure</td>
<td>To establish the validity of an instrument, at least construct-, content-, and criterion validity should be assessed</td>
</tr>
<tr>
<td>Construct validity</td>
<td>The extent to which an instrument measures an intended hypothetical construct Two aspects of construct validity exist: - Convergent validity - Divergent validity</td>
<td>Factor analysis / IRT models</td>
</tr>
<tr>
<td>Content validity</td>
<td>The extent to which all relevant domains are captured Face validity is closely related to content validity</td>
<td>Content validity: respondents’ judgement Face validity: expert panel judgement</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>The extent to which the instrument correlates to other instruments measuring the same construct</td>
<td>The degree of agreement between measures of the same construct is assessed by calculating the correlation coefficient, which should be &gt;0.40</td>
</tr>
<tr>
<td>Responsiveness**</td>
<td>The extent to which the score of an instrument changes as a patient’s condition changes over time</td>
<td>Evaluated by longitudinal assessment of patients - Standardized Response Mean - Effect Size</td>
</tr>
<tr>
<td>Interpretability**</td>
<td>The ability to interpret the significance of the results of an instrument in terms of a qualitative meaning to quantitative results Norms: the extent to which standard comparative data are available and/or published, from the general population and/or dermatological patients Categorization: the extent to which categories and/or cut-off scores of scores are available MIN(C)D: the minimally important (clinical) difference; the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate a change in the patient’s health care management</td>
<td>- Distribution-based methods - Anchor-based methods</td>
</tr>
</tbody>
</table>
Table 2. Continued

<table>
<thead>
<tr>
<th>Psychometric characteristics</th>
<th>Definition</th>
<th>Examples of applicable statistical tests*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item response theory (IRT)</td>
<td>IRT analysis incorporates statistical tests for e.g., uni-dimensionality of an item set and the presence of item bias (also known as ‘differential item functioning’)</td>
<td>Several software programs are available for IRT analysis such as, among others, OPLM, RUMM2010, and Winsteps</td>
</tr>
</tbody>
</table>

* All statistical tests mentioned can be calculated by all respected statistical software packages, such as SPSS, SAS and R, with the exception of IRT analysis.

construct) and divergent validity (i.e., the extent to which the instrument under study should not relate to divergent instruments), are two aspects of construct validity. Discriminant validity is often regarded as interchangeable with divergent validity, although these are different: discriminant validity is the extent to which the instrument is able to discriminate between different groups of subjects.

Content validity is the extent to which an instrument captures all relevant items and adequately covers the construct intended. It involves the critical examination of the design and development of the instrument, to test the comprehensiveness, relevance, and understanding of the instrument among experts (e.g., via expert panels) and patients (e.g., via pilot tests and/or cognitive debriefing interviews). Face validity is closely related to content validity; does the instrument measure what it is supposed to measure at first sight? It is a critical review of the items of an instrument by experts.

Criterion validity is the extent to which the instrument correlates with an external criterion of the same construct, ideally, a ‘gold standard’. A correlation coefficient of >0.40 is generally considered to be acceptable. Criterion validity can be divided into concurrent validity (i.e., the correlation of the instrument under study and the criterion instrument, administered at the same time) and predictive validity (i.e., the extent to which the instrument under study is able to predict an outcome).

Responsiveness
Responsiveness is the extent to which the score of an instrument changes as a patient’s condition changes over time, and is useful to assess the interpretability of these changes in scores. Responsiveness can be evaluated by longitudinal assessment of patients, and the most commonly used measures of responsiveness are the standardized response mean (SRM), and the effect size (ES). The SRM is calculated by dividing the mean score change by the standard deviation of the change; and the ES is the degree of change measured in standard deviations. Responsiveness is sometimes referred to as sensitivity to change, although these are related but different. Responsiveness is the ability of an instrument to measure clinically important change within patients, whereas sensitivity to change refers to the ability of an instrument to detect any degree of change.

Reliability, validity and responsiveness are interrelated, yet each is independently important in assessing the psychometric characteristics of HRQoL instruments.
Additional psychometric characteristics of HRQoL instruments, such as floor and/or ceiling effects, item bias, cultural bias, response burden, administrative burden and alternative forms are described in detail by e.g., Lohr et al.\textsuperscript{36} and Both et al.\textsuperscript{35} In addition, more information about CTT can be found in text books, for instance the one by Lord and Novick.\textsuperscript{41}

**Modern test theory**

In addition to CTT models, modern test theory models, such as IRT,\textsuperscript{41-44} are advanced techniques to test the psychometric characteristics of an instrument.\textsuperscript{45} They provide further insight into the dimensionality of an instrument; the format of response categories can be tested, and information can be provided on item weights and item bias (also known as differential item functioning). Nevertheless, in this introductory article, we mainly focus on CTT. For information regarding modern psychometrics, we refer to the text book of Lord and Novick.\textsuperscript{41}

**INTERPRETABILITY**

A HRQoL score in itself has little or no direct meaning and, therefore, the interpretation of scores is not immediately straightforward.\textsuperscript{46-48} Two types of methods to establish a clinically meaningful interpretation of HRQoL scores exist: distribution-based and anchor-based methods.\textsuperscript{49} Distribution-based methods rely on the distribution of scores or their clustering within a dataset in statistically distinct subgroups.\textsuperscript{49} Recently, Nijsten et al. applied this method to categorize responses to the Skindex-29 into statistically distinct subgroups.\textsuperscript{50} Anchor-based methods examine the relationship between scores on an instrument and an independent, external measure or anchor.\textsuperscript{49} Hongbo et al.\textsuperscript{51} and Prinsen et al.\textsuperscript{46,47} used this method to support the interpretation of the clinical meaning of DLQI and Skindex-29 scores respectively.

Until now, relatively few data exist on the interpretation of HRQoL scores in dermatology, however, by gaining more insight into the interpretation of these scores, it will gradually become easier to apply HRQoL scores in clinical practice.\textsuperscript{48}

**SELECTION OF HRQOL INSTRUMENTS**

How should one select the most appropriate instrument for its use in clinical research and/or practice? Table 1 shows commonly used HRQoL instruments in dermatology that adequately can be used in a wide range of dermatological conditions. An instrument is chosen on the basis of the outcome of the study or the clinical activity to be performed. An instrument may focus on one aspect more than another, so it is important to know exactly what one wants to measure. If one of these instruments does not appear to fulfil a specific research or clinical objective, or if an instrument of choice does not exist in the preferred language, it can be decided (i) to translate an existing instrument, (ii) to look for an alternative HRQoL instrument, (iii) to add relevant questions to the most appropriate existing instrument, (iv) to combine (domains of) existing instruments or (v) to develop a new instrument, see Flowchart (Figure 1).\textsuperscript{37,52,53} With respect to alternative (iii) and (iv) permission from the author(s)/developer(s) should be sought. Furthermore, the psychometric characteristics of translated and/or newly created instruments should be tested. With respect to alternative (v) we would like to stress that the
development of a new instrument is a difficult and time-consuming process, and should not be given priority. In addition, development of a new instrument often does not contribute to the body of knowledge of existing instruments, and the data gathered with these instruments.

Figure 1. The process of selecting a HRQoL instrument for its use in clinical research and/or practice.

(*) Refer to paragraph Psychometric Characteristics and Table 2.

**DISCUSSION**

Although several reviews of generic and dermatology-specific HRQoL instruments exist, to date there is no univocal consensus as to which HRQoL instruments are to be preferred in dermatology. The choice for a suitable generic and/or specific HRQoL instrument remains a trade-off between a variety of methodological and practical pros and cons, and the selection of an appropriate HRQoL instrument in clinical research or practice depends on the research question and the target population in which the HRQoL instrument is administered. However, there are minimal prerequisites defined that should be taken into consideration before selecting an instrument. A useful HRQoL instrument should have all the properties described previously. Moreover, it is preferable to use multi-dimensional instruments, including uni-dimensional domains. These aspects are considered mandatory in establishing the usefulness of an instrument, and should be tested in a study population that is representative for the population in which the instrument will be used.

The most commonly used dermatology-specific instrument is the DLQI. Since its introduction by Finlay and Khan in 1994, it has played a major role in the measurement of HRQoL in dermatology. The DLQI is available in many languages, and may serve as a comparator instrument. The Skindex-29, introduced in 1997, and developed according to the CTT model,
was the first multi-dimensional dermatology-specific instrument that included three domains of HRQoL. Testing an instrument using CTT alone is nowadays considered to be insufficient. An important issue is that the item and instrument statistics apply to the study population in which the instrument is tested, and will not be equivalent in all circumstances. This means that, if an instrument is being tested in another (study) population, test statistics might be different. Therefore, the EADV Taskforce recommends that present and future instruments are being (re)analysed according to the above mentioned requirements of the CTT, and preferably also according to modern test theory models. In dermatology, for example, IRT analysis has been useful for evaluating instruments, such as the DLQI. In addition, it has also been used in testing the Skindex-29, that resulted in a reduced version: the Skindex-17. Both et al. and De Korte et al. concluded to use a combination of a generic instrument, the SF-36, and a dermatology-specific instrument, the Skindex-29. The EADV Taskforce supports the recommendation of using a combination of a generic- and a dermatology-specific instrument for HRQoL assessment in clinical research. For clinical practice purposes, however, the Taskforce recommends using a dermatology-specific instrument as generic instruments might fail in the assessment of important dermatology-specific aspects.

The EADV Taskforce on Quality of Life wishes to increase the scientific knowledge of HRQoL measurement and encourage researchers and clinicians in dermatology with a great interest in patient-reported outcomes research to conduct additional methodological studies in cooperation with the Taskforce. For instance, insight into psychometrics has evolved and information about dimensionality, response categories, and differential item functioning of several HRQoL instruments in dermatology has been investigated. From these studies we know that patients with different skin diseases may interpret and respond differently to items of a HRQoL instrument. From a theoretical perspective, the comparison of these scores is currently under debate and future research in this field is needed. In addition, it is recommended to fully review the existing disease-specific instruments used in dermatology, so one can adequately select the most appropriate disease-specific instrument for its purpose as well.

ACKNOWLEDGEMENTS

We thank Prof. Dr. M.A.G. Sprangers, Department of Medical Psychology, Academic Medical Center, University of Amsterdam, the Netherlands, and Dr. Ph.I. Spuls, Department of Dermatology, Academic Medical Center, University of Amsterdam, the Netherlands, for critically reviewing this manuscript.

APPENDIX

REFERENCES

Chapter 2.1


40. Terwee CB, Bot SD, Boer de MR, *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60:34-42.


