Health-related quality of life in dermatology: measurement, interpretation and application
Prinsen, C.A.C.

Citation for published version (APA):
HEALTH-RELATED QUALITY OF LIFE IN DERMATOLOGY: measurement, interpretation and application

C.A.C. Prinsen
HEALTH-RELATED QUALITY OF LIFE IN DERMATOLOGY: measurement, interpretation and application

Cecilia Anna Catharina Prinsen
HEALTH-RELATED QUALITY OF LIFE IN DERMATOLOGY:
measurement, interpretation and application

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 19 december 2013, te 12:00 uur

doors

Cecilia Anna Catharina Prinsen
geboren te Amersfoort
Iets is alleen onmogelijk als je er zelf niet in gelooft

(inzichtkaartje)
Promotiecommissie:

Promotores:  
Prof. dr. M.A. de Rie  
Prof. dr. M.A.G. Sprangers

Co-promotores:  
Dr. J. de Korte  
Dr. P.I. Spuls  
Dr. R. Lindeboom

Overige leden:  
Prof. dr. N.K. Aaronson  
Prof. dr. A.W.M. Evers  
Prof. dr. R.J. de Haan  
Prof. dr. T.E.C. Nijsten  
Dr. A. Wolkerstorfer

Faculteit der Geneeskunde
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>General introduction</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2</td>
<td><strong>Measurement of health-related quality of life in dermatology</strong></td>
<td>23</td>
</tr>
<tr>
<td>2.1</td>
<td>Measurement of health-related quality of life in dermatological research and practice: outcome of the EADV Taskforce on Quality of Life</td>
<td></td>
</tr>
<tr>
<td>Chapter 3</td>
<td><strong>Interpretation of health-related quality of life scores</strong></td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>Health-related quality of life assessment in dermatology: interpretation of Skindex-29 scores using patient-based anchors</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>J Invest Dermatol</em> 2010; 130 (5):1318-22</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Interpretation of Skindex-29 scores: cutoffs for mild, moderate, and severe impairment of health-related quality of life</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>J Invest Dermatol</em> 2011; 131(9):1945-7</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Interpretation of Skindex-29 scores: response to Sampogna and Abeni</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>J Invest Dermatol</em> 2012; 132(5):1500-1</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>Determining the meaningfulness of differences in Skindex-29 scores using item response theory modelling</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Submitted</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 4</td>
<td><strong>Health-related quality of life application in clinical practice</strong></td>
<td>85</td>
</tr>
<tr>
<td>4.1</td>
<td>Health-related quality of life assessment in dermatologic practice: relevance and application</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Dermatol Clin</em> 2012; 30(2):323-32</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: study protocol for a multicenter randomized controlled trial</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Trials</em> 2012; 13:236</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: results of a multicenter randomized controlled trial</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Submitted</em></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5  Summary and conclusions  135
Samenvatting en conclusies

Chapter 6  General discussion  151

Addendum
- Interpretation of quality-of-life scores  163
- Interpretation of Skindex-29 scores  171
- Skindex-29 (source questionnaire) / Skindex-29 (Nederlandse versie)  179
- List of contributing authors  185
- Acknowledgements / Dankwoord  189
- Bibliography  197
- Portfolio  203
- Curriculum Vitae  209
GENERAL INTRODUCTION
INTRODUCTION

The importance of the use of patient reported outcomes (PROs) in research, clinical practice, and health care management is increasingly being recognized. A PRO is a measurement of any aspect of a patient’s health status that is directly assessed by the patient without the interpretation of the patient’s response by anyone other than the patient. Examples of PROs include illness perceptions, satisfaction with care, and health-related quality of life (HRQoL). PROs are most commonly assessed by means of self-administered questionnaires, termed instruments or patient reported outcome measures (PROMs). One of the most commonly used PROs in health care research and clinical practice is HRQoL.

Health-related quality of life

Following the definition of the World Health Organization, HRQoL entails at least three domains and reflects patients’ evaluation of physical, mental and social functioning and well-being. Examples include physical discomfort, psychological stress, and social interaction. HRQoL is particularly relevant in patients with a chronic skin disease, such as acne, eczema, hidradenitis suppurativa, psoriasis, and vitiligo, where dermatological treatment can only offer a temporary suppression or remission of symptoms. When treatment is not expected to cure the disease and where patients’ well-being is adversely affected, dermatological treatment is increasingly directed towards both a decrease of disease severity and an increase of patients’ HRQoL.

Since HRQoL is generally considered to be an important PRO in research and clinical practice, understanding of the concept HRQoL and the measurement of HRQoL is essential. Reliable, valid, and interpretable instruments are needed to adequately measure HRQoL. In addition, knowledge of the interpretation of HRQoL data, as well as knowledge of the relevance and application of HRQoL measurement, are prerequisites for the routine use of a HRQoL intervention in clinical practice.

Measurement of health-related quality of life in dermatology

HRQoL instruments generally consist of a number of multiple-response questions or items. Response options often refer to frequency, intensity, or severity. The measurement of HRQoL with such instruments results in one or more (domain) scores reflecting patients’ HRQoL. A large number of HRQoL instruments exist. Among these instruments a distinction can be made between generic instruments and specific instruments. Generic instruments, such as the Medical Outcomes Study 36-item Short Form Health Survey (SF-36), can be applied to different kinds of disease populations as well as the general population. Specific instruments comprise dermatology-specific and disease-specific instruments. Dermatology-specific instruments, such as the Dermatology Life Quality Index (DLQI), the Skindex-29 and the Skindex-17, have been designed for various skin diseases. Disease-specific instruments have been developed for particular skin diseases, for example the Psoriasis Index of Quality of Life (PSORI-QoL) for psoriasis patients and the Quality of Life Index for Atopic Dermatitis (QoLiAD) for patients with atopic dermatitis. Specific instruments capture relevant domains and aspects of HRQoL that are considered to be important to patients with skin diseases that might not be captured with generic instruments.
The DLQI, introduced in 1994, was the first dermatology-specific instrument to measure HRQoL in skin diseases. This practical tool consists of ten items. An overall sum score can be calculated, ranging from 0-30, with higher scores indicating a greater impairment of patients’ HRQoL. The DLQI is the most commonly used instrument in dermatological research, and it has played a major role in the development of HRQoL measurement in dermatology over the years.

Knowledge of measurement properties has evolved. The Skindex-29, developed in 1997 by Professor M.M. Chren and colleagues, was the first multi-dimensional, dermatology-specific HRQoL instrument. It consists of 29 items that are combined to form three domains: Symptoms (consisting of 7 items), Emotions (10 items), and Functioning (12 items). Items include statements, and inquire about the frequency of these statements during the past week. Responses are given on a five-point response scale, ranging from ‘never’ (0) to ‘all the time’ (100), with higher scores indicating lower levels of HRQoL. Several reviews suggest that the Skindex-29 is considered to be the instrument of choice in dermatology.

Before a HRQoL instrument can be adequately used in research and clinical practice, the psychometric quality (i.e., measurement properties) needs to be tested to ensure the reliability (i.e., measurement precision) and validity (i.e., measuring what is intended) of the instrument. Psychometrics is the methodology concerning the theory and technique of the design, construction, validation, administration, analysis and interpretation of a measuring instrument that measures a certain construct such as HRQoL. Most HRQoL instruments are developed and psychometrically tested according to classical test theory models, and so are the DLQI and the Skindex-29.

To further contribute to the quality of HRQoL measurement in dermatology, basic knowledge of the concept HRQoL, its measurement instruments and the psychometric quality of these instruments is important, especially for clinicians or junior researchers who are relatively unfamiliar with this concept.

**Interpretation of health-related quality of life scores**

Once an instrument is found to be reliable and valid, an important next question is: what is the clinical meaningfulness of HRQoL scores and how can scores be interpreted by clinicians? Questions on the interpretation of HRQoL scores are challenging because, in the absence of a unit of measurement, a HRQoL score in itself has little or no direct meaning and the interpretation of scores is not immediately straightforward. In general, there are two types of methods to establish the clinically meaningfulness of HRQoL scores: anchor-based methods and distribution-based methods. Anchor-based methods examine the relationship between scores on a HRQoL instrument and an independent measure or anchor, whereas distribution-based methods rely on the score distributions of clinically distinct subgroups of patients.

Contrary to the classical test theory model, that is directed to sum scores to estimate patients’ HRQoL, modern test theory models, such as item response theory (IRT) models and Rasch measurement, focus on the properties of individual test items as the main source of information on patients’ HRQoL. Modern test theory models incorporate statistical tests for uni-dimensionality of an item set. In a study by Professor T.E.C. Nijsten and colleagues, for example, the Rasch measurement model was used in testing the Skindex-29. This resulted
in a reduced 17-item version. Modern test theory models can also be used to provide further insight into the clinical meaningfulness of HRQoL scores. They provide directions on how to score items, and inform whether weighting the items on their discriminative ability is needed. In addition, modern test theory models can be used to estimate ‘person ability’ and ‘item difficulty’ and can place items in the order of these estimates. The resulting hierarchy of items is an aid in understanding the differences in HRQoL scores. With these advantages over classical test theory models, modern test theory models are at present generally acknowledged as a step forward in instrument development and testing, and allow the use of the best available instruments in health care research and clinical practice.

To date, however, little is known about the clinically meaningfulness of Skindex-29 scores, or differences herein.

**Health-related quality of life application in clinical practice**

The measurement of HRQoL is considered to be specifically relevant for (i) patients with a chronic skin disease, (ii) who may require a long-term treatment, (iii) where social visibility plays an important role, and (iv) when the skin disease is severe. It is known that, for example, patients with psoriasis and eczema may suffer considerably from their skin disease in terms of HRQoL. Patients with other chronic skin diseases, such as acne, hidradenitis suppurativa, and vitiligo, may also experience a considerable negative impact on their HRQoL. In patients who are expected to experience a negative impact of their skin disease on HRQoL, it is important that dermatologists do not exclusively focus on the physical symptoms but also on the effects of the skin disease on HRQoL. Aspects such as ‘depression’, ‘social interaction’ and ‘shame’ often remain unknown to clinicians. However, the application of HRQoL measurement in dermatological practice is not yet customary. At busy dermatology clinics, clinicians might experience the measurement of HRQoL as time consuming or prohibitive. They may also experience practical burdens, such as extra paper work, and the need for assistance to instruct patients in completing the questionnaires. In order to effectively apply a HRQoL intervention in clinical practice, it is essential for clinicians to recognize the relevance of HRQoL measurement, to know which patients might benefit most from it, and to know how HRQoL measurement can be best applied in clinical practice.

Among others, Professor G. Velikova and colleagues and Dr S.B. Detmar and colleagues investigated the effects of HRQoL measurement in clinical oncology practice. They reported a positive impact on patients’ well-being and doctor-patient communication, and a clinically meaningful improvement in patients’ HRQoL. Dr J.M. Valderas and colleagues conducted a systematic review on the impact of PRO measurement in different clinical settings, such as internal medicine, oncology, and primary care. They reported that no apparent conclusion could be drawn because included studies were heterogeneous in types of setting, participants, intensity of intervention, and diversity of outcomes. Despite these findings, positive aspects of the impact of PRO measurement in clinical practice were found, such as the facilitation of doctor-patient communication and the detection of physical or psychological problems, and further research was suggested. At present, evidence on the effectiveness of a HRQoL intervention in dermatological practice is missing.
AIMS AND OUTLINE OF THE THESIS

The overall aim of this thesis is to contribute to the improvement of HRQoL of patients with a chronic skin disease. The focus of the thesis is on (i) the measurement of HRQoL in dermatology, (ii) the interpretation of HRQoL data, and (iii) the application of a HRQoL intervention in clinical practice.

The aim of the first part of the thesis is to further contribute to the quality of HRQoL measurement in dermatological research and clinical practice, by providing an introduction to the concept HRQoL and the related methodology of the measurement of HRQoL (chapter 2.1). The aims of the second part are to facilitate the interpretation of HRQoL scores and of score differences using the Skindex-29 (chapters 3.1 through 3.4). The aims of the third part are to support the application of HRQoL measurement in clinical practice (chapter 4.1), and to investigate the efficacy of a HRQoL intervention in dermatological practice (chapters 4.2 and 4.3).

In 2008, the European Academy for Dermatology and Venereology (EADV) Taskforce on Quality of Life was established by Professor A.Y. Finlay and Dr T. Schaefer (http://www.eadv.org/). The objective of this Taskforce is to contribute to the scientific knowledge and measurement of PROs in dermatology. The Taskforce is encouraging the application of HRQoL instruments in research and clinical practice. The aim of the first manuscript that was written on behalf of the Taskforce is to contribute to the overall quality of HRQoL measurement in dermatology. Chapter 2.1 is a review article that provides an introduction to the concept and methodology of HRQoL. An overview of the most commonly used HRQoL instruments in dermatology is given. Background information is provided on the psychometric quality of these instruments, as well as on the selection process of a HRQoL instrument. Furthermore, it discusses the lack of consensus regarding the preferred HRQoL instrument to be used in dermatology.

As indicated, little is known about the interpretability of Skindex-29 scores. Our aim is to facilitate the interpretation of Skindex-29 scores by identifying clinically meaningful cut-off scores. An anchor-based method is used to establish these cut-off scores. In chapter 3.1 we present the results of a multicenter survey study, determining the most optimal cut-off scores for the Skindex-29 domain and overall scores indicating patients with (very) severely impaired HRQoL. Patients with scores equal to or above the established cut-off scores are significantly affected by their skin disease.

In a commentary on the interpretation of Skindex-29 scores presented in chapter 3.1, Professor M.M. Chren emphasized the relevance of establishing Skindex-29 cut-off scores for mild and moderate impairment of HRQoL in addition to those for (very) severe impairment (Addendum I). Chapter 3.2 is a response to Professor M.M. Chren and presents additional empirical data on Skindex-29 cut-off scores that are indicative for mildly and moderately impaired HRQoL, in addition to the cut-off scores for (very) severe impairment.

In another commentary by Dr F. Sampogna and Dr D.D. Abeni, the categorization of Skindex-29 scores using an anchor-based method is compared with the categorization of scores using a distribution-based method. Differences and implications between these two categorization methods are discussed (Addendum II). Unfortunately, the authors
misinterpreted our findings, leading to an incorrect categorization of cut-off scores. Chapter 3.3 is a Letter to the Editor reflecting on this commentary, and a correct overview of the categorization of Skindex-29 cut-off scores is provided.

An important limitation in HRQoL research in dermatology is the lack of evidence concerning which difference in Skindex-29 scores is clinically meaningful. In addition to this, the reduction of the Skindex-29 into the 17-item version, by using the Rasch model, may have resulted in the deletion of valuable items, as the Rasch model does not permit items to differ in their level of discrimination. These items may provide important information to be used in the clinical management of dermatology patients. Our aim is to examine the discriminating capacity of the Skindex-29 items, and to determine the clinical meaningfulness of differences in Skindex-29 scores. Chapter 3.4 presents the results of an empirical study using a specific extension of the IRT model, called the one-parameter logistic model (OPLM). OPLM is used to identify clinically meaningful differences in Skindex-29 scores.

The application of HRQoL measurement in dermatological practice is not yet customary and many practical and attitudinal barriers need to be overcome. Our aim is to support the application of HRQoL measurement among clinicians. Chapter 4.1 is a didactic article that adds to this application. It describes the relevance of HRQoL measurement to dermatological practice, it illustrates which patients would benefit most from routine HRQoL measurement, and it reflects on how HRQoL measurement can be applied in clinical practice.

Evidence on the evaluation of the effectiveness of a HRQoL intervention in dermatological practice is missing. The aim of our study is to investigate the efficacy of a HRQoL intervention in clinical practice. Chapter 4.2 presents a study protocol describing the rationale and design of a multicenter randomized controlled trial (RCT) that investigates the efficacy of a HRQoL intervention in patients with moderate to severe psoriasis undergoing biologic treatment during a treatment period of 48 weeks.

Chapter 4.3 presents the results of this multicenter RCT. Psoriasis patients who received a biologic treatment with the HRQoL intervention (that is, the completion of the electronic Skindex-29 and communication about the resulting data with their dermatologists) are compared to patients who received a biologic treatment alone. The effect of the HRQoL intervention on, primarily, patients’ HRQoL and doctor-patient communication, and secondary on health status and disease severity, is examined after 24 and 48 weeks respectively.

Chapter 5 is a summary of chapters of the thesis.

The thesis concludes with a general discussion reflecting on the main findings and limitations, and future perspectives are discussed (chapter 6).
REFERENCES


MEASUREMENT OF HEALTH-RELATED QUALITY OF LIFE IN DERMATOLOGY
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. It has become an essential outcome parameter in clinical research, clinical practice and health care management. 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. 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’. 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, the Medical Outcomes Study 36-item Short Form Health Survey (SF-36), the Nottingham Health Profile (NHP), the Sickness Impact Profile (SIP), and the World Health Organization Quality of Life assessment (WHOQOL).
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 Skindex-29, and the Skindex-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. and Both 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. 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.

**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.

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.

**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. 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.

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.

**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. 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>EuroQol Quality of Life Scale (EQ-5D)</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>Medical Outcomes Study 36-item Short Form Health Survey (SF-36)</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>Nottingham Health Profile (NHP)</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>Sickness Impact Profile (SIP)</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>World Health Organization Quality of Life assessment (WHOQOL-100)</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>
<tr>
<td>Instrument</td>
<td>Total number of items</td>
<td>Total number of domains</td>
<td>Main psychometric issues</td>
<td>Completion time (min)</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>DLQI</td>
<td>10 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>DQOLS</td>
<td>41 Items, developed to assess the impact of skin diseases on patients’ psychosocial state and everyday activities</td>
<td>Three domains: - Dermatologic symptoms - Physical activities (subscale: embarrassment, despair, irritability, distress) - Psychosocial state (sub-scale: 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>DSQL</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</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</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><strong>Scale structure</strong></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 unidimensionality 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><strong>Reliability</strong></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><strong>Validity</strong></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. Factor analysis / IRT models</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></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><strong>Responsiveness</strong></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 MI(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><strong>Modern test theory</strong></td>
<td></td>
<td></td>
</tr>
<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 textbooks, 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 textbook 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.

**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.\textsuperscript{28,55} In addition, it has also been used in testing the Skindex-29, that resulted in a reduced version: the Skindex-17.\textsuperscript{27} 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.\textsuperscript{19,35} 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


INTERPRETATION OF HEALTH-RELATED QUALITY OF LIFE SCORES
Health-related quality of life assessment in dermatology: interpretation of Skinex-29 scores using patient-based anchors

CAC Prinsen
R Lindeboom
MAG Sprangers
CM Legierse
J de Korte

J Invest Dermatol 2010; 130 (5):1318-22
ABSTRACT
In dermatology, the clinical use of health-related quality of life (HRQoL) scores is impeded by lack of empirically and clinically based interpretation of these scores. We aimed to facilitate the interpretation of Skindex-29 domain and overall scores by identifying clinically meaningful cut-off scores, using patient-based anchors. Consecutively included dermatology outpatients completed the Skindex-29 and four sets of anchor-based questions, such as questions on the impact of skin disease on HRQoL, on global disease severity, and on psychiatric morbidity. Pearson’s correlations and receiver operating characteristic analysis were used to identify the optimal Skindex-29 cut-off scores corresponding to severely impaired HRQoL. A total of 339/434 patients completed the questionnaires (response rate 78%), of which 322 could be used for data analysis. Cut-off scores associated with the patient-based anchors on the impact of skin disease on HRQoL showed the highest accuracy (area under the curve ranged from 0.83 to 0.91). The corresponding Skindex-29 cut-off scores for severely impaired HRQoL were as follows: ≥52 points on symptoms, ≥39 on emotions, ≥37 on functioning, and ≥44 on the overall score. The estimated cut-off scores can be used in clinical practice to identify patients with (very) severely impaired HRQoL.
INTRODUCTION

Health-related quality of life (HRQoL) reflects patients’ evaluation of the impact of disease and treatment on their physical, psychological, and social functioning and well-being.¹² In clinical practice, HRQoL is considered to be an aid for clinical decision making, monitoring the therapeutic process, communicating with the patient, and evaluating treatment outcome.³⁴ A good understanding of the concept of HRQoL and a correct interpretation of HRQoL scores are essential.

The well-established Skindex-29 is a three-dimensional, dermatology-specific HRQoL questionnaire. Twenty-nine items are combined to form three domains: symptoms, emotions, and functioning. The domain scores and an overall score are expressed on a 100-point scale, with higher scores indicating lower levels of quality of life.⁵⁻⁸ However, a score in itself has little or no direct meaning and cannot be interpreted in a straightforward manner.

Two types of methods to establish a clinically meaningful interpretation of HRQoL scores exist: distribution-based and anchor-based methods. Distribution-based methods rely on the score distributions of clinically distinct subgroups of patients.⁴ A categorization of Skindex-29 scores using this method was recently published.⁹ However, in this study, patient-based anchors were not used and further research was suggested. Anchor-based methods examine the relationship between scores on an HRQoL instrument and an independent measure or anchor.⁴ This method was developed to estimate clinically meaningful cut-off scores for HRQoL instruments to allow clinicians to interpret scores more straightforward (e.g., a score of ≥44 indicates severely impaired HRQoL). An anchor should be itself interpretable and should at least moderately correlate with the HRQoL instrument under study.⁴,¹⁰ With respect to dermatology-specific questionnaires, the interpretation of scores by using a patient-based anchor was previously studied for the Dermatology Life Quality Index.¹¹,¹² In this study, we aim to determine Skindex-29 domain and overall cut-off scores using different patient-based anchors.¹³

METHODS

Setting and study population

We conducted a multi-center, cross-sectional study in dermatology outpatients with unselected chronic skin disease. Patients were consecutively recruited at nine dermatology outpatient clinics during a predetermined period of 4 weeks (14 April to 9 May 2008). Patients eligible for this study had a chronic skin disease and were 18 years or older. Excluded were patients who were mentally and/or physically unable to complete the questionnaires and patients with insufficient mastery of the Dutch language. Patients who gave their written informed consent during their visit at the dermatology outpatient clinics were asked to complete the questionnaires independently and to return the completed package by using a stamped return envelope. The central Ethics Committee AMC (EC AMC) exempted this study for ethical approval. For non-interventional questionnaire research, this is common policy in the Netherlands. A written confirmation of this policy was given by the EC AMC. The study was conducted according to the Declaration of Helsinki Principles.
**Measurements**

As it is strongly recommended to use multiple independent anchors to examine cut-off scores, the questionnaires comprised of the Skindex-29 and four sets of anchor-based questions: (i) four General Questions (GQs), evaluating the impact of the skin disease on the three domains of the Skindex-29 (GQ1-3) and on overall impairment of HRQoL (GQ4); (ii) one question on patients’ perception of the degree of global severity of the skin disease; (iii) seven questions on patients’ treatment needs; and (iv) the 12-item General Health Questionnaire (GHQ-12) consisting of 12 items and designed to measure psychiatric morbidity, usually depressive or anxiety disorders. A pilot study among seven patients of the Academic Medical Center was performed to test whether there was any difficulty or ambiguity in the wording of the anchor-based questions, with the exception of the standardized GHQ-12.

**Statistical analysis**

Sample size calculations were based on the precision of the estimates of the receiver operating characteristic - area under the curve (ROC-AUC) statistic. The ROC-AUC statistic expresses the strength of the relation between the Skindex-29 domain and overall scores, and the anchor-based question dichotomizations (see below) to indicate severe impairment of HRQoL. A total of 300 patients were needed to have ± 10% error margins around the validity coefficients under conventional study reliability requirements (5% type I error rate, 80% power), assuming a 3:1 ratio between the test negative and test positive groups and a correlation of 0.60 between tests. Patients were excluded from data analysis in case ≥25% of the items of the Skindex-29 were missing. In case of < 25%, the missing values were imputed with the mean of the domain scores per patient.

To assess whether the anchor-based questions met the formal characteristics of a patient-based anchor (at least a moderate degree of correlation: $r \geq 0.40$), Pearson’s correlations were calculated between Skindex-29 domain and overall scores and the anchor-based questions: GQ1, GQ2, and GQ3 were related to the Skindex-29 domain scores for symptoms, emotions, and functioning, respectively, whereas GQ4 was related to the overall Skindex-29 score. The score on Global Disease Severity (GDS), the scores on the seven questions with regard to patients’ treatment needs, and the GHQ-12 score were related to both the Skindex-29 domain and overall scores. All questions referred to the past week.

ROC-curve analysis was then used to determine optimal Skindex-29 cut-off scores for the selected anchors. The ROC-AUC indicates the overall accuracy of the Skindex-29 cut-off scores; a higher value indicates a better discriminating capacity of a given Skindex-29 cut-off score to distinguish patients, for instance, with and without impaired HRQoL. For the construction of the ROC-curves, the five-category anchor variables of the Skindex-29, namely, (1) never, (2) seldom, (3) sometimes, (4) often, and (5) all the time, were dichotomized using ratings 1-3 vs 4-5 for severe and very severe impairment of HRQoL. For the GHQ-12, the presence of psychiatric morbidity was indicated by a score of five points or more. Cut-off scores were rounded to zero decimal places. The Youden Index was used to determine the optimal balance between sensitivity (true positive rate) and specificity (true negative rate) in the estimation of the Skindex-29 cut-off scores. All analyses were run under SPSS, (Chicago, IL), version 16.0.
RESULTS

Study population
At nine outpatient dermatology clinics in the Netherlands, 434 patients were asked to complete the questionnaires after informed consent was obtained. A total of 339 patients returned the questionnaires (response rate 78%). Seventeen patients were excluded from data analysis as ≥25% of the Skindex-29 items were missing, leaving 322 patients for analysis. In these patients, only 0.4% of the Skindex-29 items had to be imputed. The 95 non-respondents did not significantly differ from respondents with regard to gender, but they were younger (45.2 vs 49.5 years). Table 1 shows the characteristics of the study population, their Skindex-29 scores, their scores on disease severity, and their scores on the GHQ-12 at baseline. In this study population, the prevalence of psychiatric morbidity was 24.4%. In case of more than one dermatological condition, the diagnosis that had bothered the patient the most during the past week was taken as the diagnosis.

Patient-based anchors
Correlations were calculated for the Skindex-29 domain and overall scores versus four sets of anchor-based questions. Five anchor-based questions (the four GQs and the question on GDS), and the GHQ-12 had a correlation of ≥0.40 (range 0.42-0.79) with the relevant Skindex-29 domain and overall scores, and thus met the requirements for a patient-based anchor (see Supplementary Table S1). Low correlations were found for seven anchor-based questions with regard to patients’ treatment needs. Therefore, these questions were excluded from further analysis.

Skindex-29 cut-off scores
We established cut-off scores for “severe to very severe” impact of disease on HRQoL (further referred to as “severe”). Table 2 shows the estimated cut-off scores associated with severely impaired HRQoL, severe disease severity, and psychiatric morbidity for the Skindex-29 domain and overall scores. The Skindex-29 domain and overall cut-off scores associated with the patient-based anchors relating to the impact of disease on HRQoL showed the highest accuracy: the AUC ranged from 0.83 to 0.91. The AUC for the anchor on disease severity ranged from 0.69 to 0.76, and for psychiatric morbidity from 0.73 to 0.83. The optimal and, according to the AUC statistic, most accurate Skindex-29 cut-off scores for severely impaired HRQoL were as follows: symptoms ≥52, emotions ≥39, functioning ≥37, and for the overall score ≥44 points.

Subgroup analyses
We have performed subgroup analysis for psoriasis patients (N=138) and patients with eczema (N=76). As psoriasis patients were overrepresented (> 40%), we also examined the cut-off scores of this subgroup and the cut-off scores of the entire study population without this subgroup. The cut-off scores of the subgroup analyses did not significantly differ from the presented cut-off scores of the entire study population. The resulting cut-off scores for eczema patients on the anchors relating to the impact of disease on HRQoL and on disease severity are higher than the study results presented, but those for psychiatric morbidity were similar (data not shown).
Table 1. Baseline characteristics of the study population (N=322).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male gender</td>
<td>146</td>
<td>45.3</td>
</tr>
<tr>
<td>Mean age in years (SD, range)</td>
<td>49.5</td>
<td>17.4</td>
</tr>
</tbody>
</table>

Diagnoses, n (%)

- Acne, other disorders of sebaceous, apocrine, or eccrine glands: 19 (5.9)
- Autoimmune disorders: 5 (1.6)
- Benign pigmented lesions and naevi: 1 (0.3)
- Benign skin and vascular tumors: 2 (0.6)
- Decubitus: 1 (0.3)
- Eczematous lesions: 76 (23.6)
- Genetic disorders: 4 (1.2)
- Genital skin disorders: 3 (0.9)
- Granuloma annulare: 1 (0.3)
- Hair and scalp disorders: 3 (0.9)
- Infection of skin transplant after trauma: 1 (0.3)
- Jessner-Kanoff lymphocytic infiltrate: 1 (0.3)
- Lichen sclerosus: 9 (2.8)
- Non-melanoma skin cancers and premalignant lesions: 17 (5.3)
- Pigmentary disorders: 6 (1.9)
- Pityriasis lichenoides chronic: 1 (0.3)
- Pruritus: 4 (1.2)
- Psoriasis: 138 (42.9)
- Reactive skin disorders and drug reactions: 4 (1.2)
- Skin malignancies not otherwise specified: 3 (0.9)
- Superficial fungal infections: 1 (0.3)
- Ulcers: 4 (1.2)
- Urticarial disorders: 9 (2.8)
- Varicose veins: 1 (0.3)
- Viral skin lesions: 8 (2.5)

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean</th>
<th>SD (minimum-maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms</td>
<td>46.7</td>
<td>22.3 (0-96.4)</td>
</tr>
<tr>
<td>Emotions</td>
<td>37.6</td>
<td>22.0 (0-100.0)</td>
</tr>
<tr>
<td>Functioning</td>
<td>26.4</td>
<td>21.2 (0-97.9)</td>
</tr>
<tr>
<td>Overall</td>
<td>35.2</td>
<td>19.0 (0-97.9)</td>
</tr>
</tbody>
</table>

Global Disease Severity (GDS) (N=312): 3.2 (NA)
GHQ-12 score (N=303): 2.5 (NA)
Psychiatric morbidity (N=74): 8.2 (NA)

Abbreviation: NA, not applicable.

1 The domain scores and the overall score are expressed on a 100-point scale, with higher scores indicating a lower level of quality of life.
2 Global Disease Severity (GDS): one question on patients’ perception of the degree of global severity of the skin disease.
3 Different sample sizes are because of missing values.
4 GHQ-12: the 12-item General Health Questionnaire designed to measure psychiatric morbidity.
**DISCUSSION**

We aimed to facilitate the interpretation of Skindex-29 scores by determining clinically important cut-off scores. We found robust Skindex-29 cut-off scores with all three patient-based anchors expected to indicate severely impaired HRQoL: GQs on the impact of the skin disease on HRQoL, patients’ assessment of disease severity, and the presence of psychiatric morbidity as measured with the GHQ-12. Except for the Skindex-29 functioning domain using the GHQ anchor, the cut-off scores were highly comparable.

In clinical practice, the primary focus should be on the profile of the three domain scores, as these scores will provide clinicians with information on which domain of HRQoL bothered the patient the most. The overall score of the Skindex-29 should be interpreted with some caution, as the validity of the overall score as such is debatable. Patients with scores equal to or above the presented cut-off scores in at least one of the three domains are significantly affected by their skin disease. These scores may signal a need for (adjustment of current) treatment and/or for additional care or support. However, they do not automatically indicate what kind of treatment, care or support is appropriate: the specific needs of an individual patient should be explored in direct contact with the patient. HRQoL scores may also facilitate doctor-patient communication and mutual decision making. With the formal external anchors on disease severity and psychiatric morbidity, we were able to evaluate and confirm the robustness of the given cut-off scores on impaired HRQoL.

<table>
<thead>
<tr>
<th>Patient-based anchors:</th>
<th>Cut-off score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impact on HRQoL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptoms (r=0.54)</td>
<td>≥52</td>
<td>0.67</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>Emotions (r=0.73)</td>
<td>≥39</td>
<td>0.72</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>Functioning (r=0.79)</td>
<td>≥37</td>
<td>0.83</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>Overall (r=0.75)</td>
<td>≥44</td>
<td>0.82</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td><strong>Disease severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptoms (r=0.42)</td>
<td>≥52</td>
<td>0.70</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>Emotions (r=0.46)</td>
<td>≥35</td>
<td>0.70</td>
<td>0.67</td>
<td>0.74</td>
</tr>
<tr>
<td>Functioning (r=0.44)</td>
<td>≥42</td>
<td>0.93</td>
<td>0.45</td>
<td>0.74</td>
</tr>
<tr>
<td>Overall (r=0.51)</td>
<td>≥39</td>
<td>0.81</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>Psychiatric morbidity</strong> 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptoms (r=0.42)</td>
<td>≥55</td>
<td>0.64</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>Emotions (r=0.55)</td>
<td>≥39</td>
<td>0.78</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>Functioning (r=0.57)</td>
<td>≥28</td>
<td>0.80</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td>Overall (r=0.60)</td>
<td>≥42</td>
<td>0.74</td>
<td>0.81</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Abbreviation: AUC, area under the curve.
1 The domain scores and the overall score are expressed on a 100-point scale, with higher scores indicating a lower level of quality of life.
2 Five or more points on the 12-item General Health Questionnaire.
The low correlation between the Skindex-29 domain and overall scores and the patient-based anchors with regard to patients’ treatment needs indicates that HRQoL and treatment needs are likely to be two different constructs.

Interestingly, our results with respect to psychiatric morbidity are consistent with the results of a previous study, thereby giving further evidence for a relatively high prevalence of psychiatric morbidity among dermatological patients.20 Earlier results on the categorization of Skindex-29 scores by Nijsten et al. (2009), using a distribution-based method to establish a clinically meaningful interpretation of Skindex-29 scores, were similar with respect to the results of our study for the functioning domain and the overall score, but different for the symptoms and emotions domains. This may, in part, be the result of differences in the distribution of diagnoses and disease severities of the samples and also of the statistical methods used to derive cut-off scores.

Two limitations of this study merit attention. First, an unexpectedly high number of psoriasis patients (>40%) were included in the sample. However, the results of the subgroup analyses did not significantly differ from the presented cut-off scores. Nevertheless, further research using similar techniques in other dermatoses is recommended.

Second, owing to the relatively small sample sizes per diagnostic category, subgroup analysis could only be meaningfully performed for psoriasis patients (N=138) and patients with eczema (N=76). Further research on the generalizability of the established cut-off scores, particularly in specific diagnostic categories, is recommended.

We conclude that the estimated cut-off scores of the Skindex-29 can be used in clinical practice to identify patients with (very) severely impaired HRQoL.

ACKNOWLEDGEMENTS

We thank all dermatologists whose collaboration made the study possible: MTW Gaastra, MD, Flebologisch Centrum Oosterwal, Alkmaar; DB de Geer, MD, Diakonessenhuis, Zeist; AY Goedkoop, MD, PhD, St Antonius Hospital, Nieuwegein; CLM van Hees, MD, Reinier de Graaf Group, Voorburg; WJA de Kort, MD, Amphia Hospital, Breda; MCG van Praag, MD, PhD, St Franciscus Gasthuis, Rotterdam; MLA Schuttelaar, MD, University Medical Center Groningen, Groningen; AME Visser-Van Andel, MD, Gelderse Vallei Hospital, Ede, The Netherlands. Furthermore, we acknowledge the contributions of FJ Oort, PhD, and B King-Kallimanis, MSc, from the Academic Medical Center, Department of Medical Psychology, Amsterdam, for their contribution to this study.

REFERENCES


Table S1. Established patient-based anchors: four General Questions evaluating the impact of the skin disease on the three domains of the Skindex-29 and on overall impairment; one question on Global Disease Severity; and the 12-item General Health Questionnaire. All questions referred to the past week.

<table>
<thead>
<tr>
<th>General Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My skin condition caused physical complaints.</td>
</tr>
<tr>
<td>2. My skin condition negatively affected my mood or feelings.</td>
</tr>
<tr>
<td>3. My skin condition negatively affected my daily activities or interaction with others.</td>
</tr>
<tr>
<td>4. My skin condition negatively affected my overall quality of life.</td>
</tr>
<tr>
<td>Response: 1. Never, 2. Seldom, 3. Sometimes, 4. Often, 5. All the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Global Disease Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In your opinion, how severe was your skin condition?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12-item General Health Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you recently...</td>
</tr>
<tr>
<td>1. Been able to concentrate on what you’re doing?¹</td>
</tr>
<tr>
<td>2. Lost much sleep over worry?¹</td>
</tr>
<tr>
<td>3. Felt you are playing a useful part in things?¹</td>
</tr>
<tr>
<td>4. Felt capable of making decisions about things?¹</td>
</tr>
<tr>
<td>5. Felt constantly under strain?¹</td>
</tr>
<tr>
<td>6. Felt you couldn’t overcome your difficulties?¹</td>
</tr>
<tr>
<td>7. Been able to enjoy your normal day-to-day activities?¹</td>
</tr>
<tr>
<td>8. Been able to face up to your problems?¹</td>
</tr>
<tr>
<td>9. Been feeling unhappy and depressed?¹</td>
</tr>
<tr>
<td>10. Been losing confidence in yourself?¹</td>
</tr>
<tr>
<td>11. Been thinking of yourself as a worthless person?¹</td>
</tr>
<tr>
<td>12. Been feeling reasonably happy, all things considered?¹</td>
</tr>
<tr>
<td>Response: ¹ better than usual, same as usual, less than usual, much less than usual</td>
</tr>
<tr>
<td>² not at all, no more than usual, rather more than usual, much more than usual</td>
</tr>
<tr>
<td>³ more so than usual, same as usual, less useful than usual, much less than usual</td>
</tr>
</tbody>
</table>
3.2

Interpretation of Skindex-29 scores: cutoffs for mild, moderate, and severe impairment of health-related quality of life

CAC Prinsen
R Lindeboom
J de Korte

J Invest Dermatol 2011; 131(9):1945-7
TO THE EDITOR

Health-related quality of life (HRQoL) is commonly assessed by means of standardized questionnaires and expressed in domain and overall HRQoL scores. An important challenge is to interpret these scores correctly. What does a given score really mean? Although there is no standard approach, several methods exist to facilitate the interpretation of HRQoL scores. In a recently published study, we identified clinically meaningful domain and overall cutoff scores for the Skindex-29 by using an anchor-based method.4 We related patient responses on the Skindex-29 to anchor questions, and we established cutoff scores by using receiver-operating characteristic analysis. As a result, we were able to determine cutoffs for severely impaired HRQoL (Table 1).

In a commentary on the interpretation of HRQoL scores, Chren (2010) stressed the relevance of Skindex-29 cutoff scores for mild and moderate degrees of effect in addition to the scores we presented for a severe degree of effect.5 In this letter, we will provide these additional cutoff scores.

We analyzed the data of our sample of 322 patients to identify optimal cutoff scores. Again, the Skindex-29 domain scores, and the overall score, were related to three types of patient-based anchors: (i) four global questions on the impact of disease on HRQoL; (ii) a question on disease severity as perceived by the patient; and (iii) the results on the 12-item General Health Questionnaire, a standardized instrument to measure psychiatric morbidity. For complete methods, we refer to the original article.1

The four global questions relating to the impact of disease on HRQoL showed the highest correlation with the domain and overall scores of the Skindex-29 (range 0.54-0.79). Cutoff scores associated with these anchors also showed the highest accuracy, as measured by the area under the curve receiver-operating characteristic statistic for mildly impaired HRQoL (range 0.76-0.91) as well as for moderately impaired HRQoL (range 0.75-0.91). On the basis of the results of these analyses, the optimal and most accurate Skindex-29 cutoff scores for mildly and moderately impaired HRQoL could be determined (Table 1).

Table 1. Skindex-29 cutoff scores for mildly, moderately, and severely impaired HRQoL.

<table>
<thead>
<tr>
<th>Impact of disease on HRQoL for Skindex-29 domain and overall scores²</th>
<th>Correlation</th>
<th>Skindex-29 cutoff scores</th>
<th>AUC³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mild</td>
<td>Moderate</td>
</tr>
<tr>
<td>Symptoms (r = 0.54)</td>
<td>39</td>
<td>42</td>
<td>52</td>
</tr>
<tr>
<td>Emotions (r = 0.73)</td>
<td>24</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>Functioning (r = 0.79)</td>
<td>21</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>Overall (r = 0.75)</td>
<td>25</td>
<td>32</td>
<td>44</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the curve; HRQoL, health-related quality of life.
1The domain scores and the overall score are expressed on a 100-point scale, with higher scores indicating a lower level of quality of life.
²The number of patients in each severity category varies, as it largely depends on the required number of responses to the Skindex-29 domains. The number of patients with mildly impaired HRQoL ranged from 144-195; the number of patients with moderately impaired HRQoL ranged from 73-104; and the number of patients with severely impaired HRQoL ranged from 49-74.
³AUC: 0.50 indicates chance categorization and 1.00 indicates perfect categorization of a given Skindex-29 score to correctly classify mild, moderate, or severe impairment of HRQoL.
⁴Cutoff scores and AUC coefficients for severely impaired health-related quality of life as presented in the original article.¹
The relatively similar cutoffs of ≥39 and ≥42 points for mildly and moderately impaired HRQoL, respectively, on the symptom domain, result from the lower correlation of that particular anchor question with the corresponding Skindex-29 domain score ($r = 0.54$). This is also visible in the lower accuracy of the symptom domain and, thereby, the lower discriminating capacity between patients who perceive mildly or moderately impaired HRQoL (area under the curve = 0.76 and 0.75, respectively). From an analytical point of view, there is no apparent explanation for this lower correlation. We assume that the patients experienced a difference between the meanings of “symptoms” as worded in the anchor question and “symptoms” as worded in the seven questions representing the symptoms domain of the Skindex-29.

The presented Skindex-29 cutoff scores for mildly, moderately, and severely impaired HRQoL are generally higher than those presented in a study by Nijsten et al. (2009), who used a distribution-based method. In our anchor-based study, the Skindex-29 cutoff scores were determined by patients’ assessments on their HRQoL, whereas Nijsten et al. (2009) capitalized on the distribution of HRQoL levels in the sample. Cutoff scores established by anchor-based methods depend on the particular anchor questions and their wording, but they are less dependent on the distribution of HRQoL levels in the sample. As patients are grouped by their scores on anchor questions (i.e., mild, moderate, and severe impairment), the obtained cutoff scores are likely to show invariance across samples. This is one of the reasons for the popularity of anchor-based methods in HRQoL research, particularly in determining minimal important differences and/or change in scores on a HRQoL instrument.

By providing these additional cutoff scores, we hope to contribute to a meaningful interpretation of HRQoL scores. To facilitate the application of the identified cutoff scores in clinical practice, it might be helpful, as a rule of thumb or memory aid, to round off the cutoffs for mild, moderate, and severe impairment to ≥20, ≥30, and ≥40 points, respectively, for the domain and overall scores, with the exception of the symptoms domain.

As expressed in our original study, we recommend further research on the generalizability and, thereby, on the robustness of the cutoff scores of the Skindex-29.

ACKNOWLEDGEMENTS

The original study was performed in nine dermatology outpatient clinics in the Netherlands. We thank all the dermatologists whose collaboration made the study possible.

REFERENCES

Cutoffs for mild, moderate, and severe impairment of health-related quality of life

Interpretation of Skindex-29 scores: response to Sampogna and Abeni

CAC Prinsen
R Lindeboom
J de Korte
TO THE EDITOR

Until recently, little was known about the interpretability of scores of the Skindex-29, a well-established, dermatology-specific health-related quality of life (HRQoL) instrument. Nijsten et al. (2009) and Prinsen et al. (2010, 2011) were the first to identify the clinical meaningfulness of Skindex-29 scores by estimating a categorization of Skindex-29 scores, denoting mildly, moderately, and (very) severely impaired HRQoL.

In their thoughtful commentary in the Journal of Investigative Dermatology, 131, (9) September 2011, Sampogna and Abeni persuasively showed how different methods, a distribution-based and an anchor-based method, respectively, result in different categorizations of scores. They applied the distribution-based ranges of scores found by Nijsten et al. and the anchor-based cutoff scores found by Prinsen et al. to an Italian sample of inpatients diagnosed with psoriasis, and to another Italian sample of dermatological outpatients. By means of this comparison, differences between the two categorizations were shown; in general, the ranges of scores presented by Nijsten et al. were lower than the cutoff scores presented by Prinsen et al. Sampogna and Abeni also explored the clinical implications of these differences, for instance the consequence of using different categories in determining patient’s eligibility for systemic treatment.

Unfortunately, a misinterpretation leading to an incorrect categorization of scores was made. To illustrate, according to Prinsen et al., the cutoff scores for mildly, moderately, and severely impaired HRQoL on the emotions domain were ≥24, ≥35, and ≥39, respectively, meaning that a patient with a score ≥24 can be categorized as having a mildly impaired HRQoL on this domain, a score ≥35 as “moderate”, etc. However, Sampogna and Abeni categorized “mild” as having a score between 0 and 23.9 and, as a consequence, misclassified all cutoff scores. Therefore, we would like to provide a correct overview of the categorization of Skindex-29 scores (Table 1).

Having said this, we fully agree with Sampogna and Abeni on the limitations of both methods, such as dependence on the distribution of HRQoL scores in estimation samples and biases when using prospective anchors. Nevertheless, we believe that, under the condition that

<table>
<thead>
<tr>
<th>Categorization</th>
<th>Symptoms</th>
<th>Emotions</th>
<th>Functioning</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very little</td>
<td>-</td>
<td>&lt;3</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Mild</td>
<td>≥39</td>
<td>4-10</td>
<td>≥24</td>
<td>6-24</td>
</tr>
<tr>
<td>Moderate</td>
<td>≥42</td>
<td>11-25</td>
<td>≥35</td>
<td>25-49</td>
</tr>
<tr>
<td>Severe</td>
<td>≥52</td>
<td>26-49</td>
<td>≥39</td>
<td>&gt;50</td>
</tr>
<tr>
<td>Very severe</td>
<td>-</td>
<td>&gt;50</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1 The domain scores and the overall score are expressed on a 100-point scale, with higher scores indicating a lower level of quality of life.
2 Skindex-29 cutoff scores are derived from the original articles.
3 Categorization of Skindex-29 scores are derived from the original article.
the same scale or anchor question is being used, anchor-based methods may lead to less variant estimates of cutoff scores than distribution-based methods. In addition, anchor-based methods are less dependent on the sociocultural and clinical characteristics of the estimation sample. For example, patients in one sample, scoring themselves as having a severely impaired HRQoL on a global rating scale or anchor question (for instance, an anchor question such as “In your opinion, how severe is your skin condition?”), are likely to have Skindex-29 scores in the same range of scores as patients of another sample who also score themselves as having a severely impaired HRQoL. Nevertheless, the phrasing of an anchor question is a great source of variation in the comparison of different cutoff scores. We therefore advocate the use of standardized anchors.

A clinically meaningful interpretation of Skindex-29 scores is of great value. At present, two studies on this intriguing subject are available. As already expressed by Sampogna and Abeni, the combination of an anchor-based and a distribution-based method in a subsequent study would allow an objective comparison of the results within one study population. In addition to this, we recommend including standardized anchors, and to conduct such a study on an international level. Eventually, such efforts will contribute to reaching consensus on the categorization of scores so that they can be applied in clinical practice.

ACKNOWLEDGEMENTS

We thank MAG Sprangers (Academic Medical Center, University of Amsterdam, Department of Medical Psychology) and Phi Spuls (Academic Medical Center, University of Amsterdam, Department of Dermatology) for critically reviewing this letter.

REFERENCES

Determining the meaningfulness of differences in Skindex-29 scores using item response theory modelling

CAC Prinsen
J de Korte
Phl Spuls
MAG Sprangers
MA de Rie
R Lindeboom
ABSTRACT

**Importance:** An important aspect in the understanding of HRQoL data is the interpretation of scores. Although information on the interpretation of Skindex-29 scores is available, an important limitation is the lack of evidence concerning what difference in scores represents a clinically meaningful difference.

**Objective:** To examine the discriminating capacity of the Skindex-29 items, and to determine scale score cut-points that mark clinically meaningful differences in Skindex-29 scores using item response theory (IRT) modelling.

**Design:** Data from two previously conducted prospective, multicenter survey studies were used for new analyses.

**Setting:** Nine dermatology out-patient clinics in the Netherlands.

**Patients:** Skindex-29 data, completed by 291 dermatology outpatients with skin diseases were applied to a specific extension of the IRT model, called the one-parameter logistic model (OPLM). The OPLM model was used to establish item weights reflecting the discrimination capacity of Skindex-29 items, and to investigate the meaning of Skindex-29 score differences. A cross-validation sample of 183 patients was used for validation purposes.

**Main outcome measures:** Clinically meaningful score differences of the Skindex-29, a dermatology-specific health-related quality of life (HRQoL) instrument.

**Results:** Preliminary forced two-factor analysis indicated a two-dimensional structure of the Skindex-29, denoting a Psychosocial and a Symptoms domain. After dichotomizing the response options and weighting the items according to their discrimination, overall fit to the OPLM model was demonstrated. Cross-validation on an independent sample confirmed these results. The observed scores on the Skindex-29 items, in both domains, closely agreed with those expected by the OPLM model; the agreement ranged between 73-98% (median 88%). Fifteen and six distinct scale cut-points could be determined for the Psychosocial and Symptoms domains respectively marking a one-increment step of one or more items of the Skindex-29.

**Conclusions:** Using a more flexible IRT model enabled us to retain all 29 Skindex items. Items clearly differed in their discriminative capacity and thus needed weighting before summation. The meaningfulness of Skindex-29 score differences was greatly enhanced using IRT modelling and may aid in the interpretation and, herewith, in the application of HRQoL data in research and clinical practice.
INTRODUCTION

In dermatology, the Skindex-29 developed and tested according to the classical test theory model, is nowadays considered to be the instrument of choice to measure the impact of a skin disease on patients’ health-related quality of life (HRQoL). To date, psychometric evaluation of HRQoL instruments is increasingly directed to item response theory (IRT) based methods, also in dermatology. IRT is an approach to unidimensional scaling in order to test whether test items measure HRQoL. Contrary to the classical test theory based method that is directed to sum scores to estimate HRQoL levels, IRT analysis focuses on the properties of individual test items as the main source of information for patients’ HRQoL. IRT analysis, for example Rasch analysis, compares the response patterns of individuals to the entire sample to estimate ‘person ability’ and ‘item difficulty’ that are both expressed on a common interval scale using the logit (log-odds probabilities) as unit of measurement. This information can be combined to make predictions about a patient’s score on the test items, given his/her level of overall HRQoL. IRT analysis is also used to assess various other psychometric characteristics, including the appropriateness of the rating scale categories, the estimation of item weights and, for item reduction purposes, to measure the concept with requisite accuracy.

The Rasch measurement model was used in testing the Skindex-29, resulting in a reduced 17-item version. The retained items of the Skindex-17, remain identical to the original items. However, the scoring system has been revised by transforming the 5-point scoring system into a 3-point scoring system, and the original three domains as well as the overall score have been substituted by two domains, namely a Psychosocial and Symptoms domain, with separate sum scores. Especially in clinical research, where test burden for patients is often high, the reduced Skindex-17 version appeared to be a valid alternative.

A major drawback, however, is that Rasch measurement models do not permit items to differ in their discriminative capacity between trait levels; the model is often too restrictive to fit the data. There may be both clinical and psychometric arguments against the reduction of the Skindex-29. First, only items with a similar level of discrimination are maintained in the questionnaire. In the development of the Skindex-17, this may have resulted in the deletion of items with an above average level of discrimination. Second, omitted items may provide valuable information for clinical use. For example, the deleted item 20 is asking patients about their ‘interaction with others’, item 12 about ‘shame’, and item 7 about ‘burns or stings’ because of the skin disease. Such aspects could be of great influence on a patient’s HRQoL, but often remain unknown to clinicians.

Therefore, the first aim of this study was to examine the discriminating capacity of the Skindex-29 items. This would allow us to investigate whether it is desirable to discard the 12 items of the Skindex-29 that has resulted in the 17-item version. To this aim, we used a more flexible extension of the IRT model, called the one-parameter logistic model (OPLM), in a sample of Dutch patients. The OPLM model allows items to differ in their level of discrimination and examines whether there are differences in discrimination capacity of individual Skindex-29 items.

To aid in the application of HRQoL data in research and clinical practice, an important aspect in the understanding of HRQoL scores is the interpretation of scores. In general,
the interpretation of scores and score differences is not straightforward when it comes to instruments that measure concepts, such as HRQoL. Although information on the interpretation of Skindex-29 scores is available,\textsuperscript{11-13} an important limitation is the lack of evidence concerning what difference in scores represents a clinically meaningful difference. For example, what does a difference of a score of 23 points mean in comparison with a score of 34 points? Therefore, our second aim was to determine clinically meaningful differences in Skindex-29 scores. The OPLM model was used to identify distinct Skindex-29 scale cut-points that mark meaningful differences in HRQoL. An improved interpretation may aid in the clinical management of patients and, also, in the understanding of the effects of dermatological treatment on patients’ HRQoL in comparative effectiveness research.

In addition, our study offered the possibility to obtain additional information on the psychometric characteristics of the 29-item and 17-item versions of the Skindex. For this reason, our final aim was to test the reliability and validity of both Skindex versions. Based on evidence for a relatively high prevalence of psychiatric morbidity among dermatological patients,\textsuperscript{12,14} an external anchor that measures psychiatric co-morbidity, the 12-item General Health Questionnaire (GHQ-12),\textsuperscript{15} was used to test the known-groups validity. The strength of the association of both versions of the Skindex with this external anchor was examined.

In summary, the aims of this study were i) to examine the discriminating capacity of the Skindex-29 items; ii) to determine clinically meaningful differences in Skindex-29 scores; and iii) to test the reliability and validity of both Skindex versions.

\textbf{MATERIALS AND METHODS}

The Skindex-29 is a commonly used and well-established, dermatology-specific HRQoL instrument. It combines 29 items that form three domains: Symptoms, Emotions, and Functioning.\textsuperscript{1,2} Responses are given on a five-point response scale (‘never’, ‘rarely’, ‘sometimes’, ‘often’, and ‘all the time’), with an average completion time of less than five minutes. The domain scores and overall score are expressed on a 100-point scale, with higher scores indicating a more impaired HRQoL. Scores equal to or above 52, 39, and 37 points for the Symptoms, Emotions, and Functioning domain respectively, and 44 points for the overall score indicated severely impaired HRQoL.\textsuperscript{12,13}

\textbf{Data collection}

Collected data of Skindex-29 scores of dermatology outpatients were used (item calibration sample, \(N=291\)). These data have been described previously.\textsuperscript{12} Patients were consecutively recruited at nine dermatology outpatient clinics in the Netherlands (two academic and six non-academic clinics, and one private clinic), during a predetermined period of four weeks between April 14 and May 9, 2008. Data of another sample (cross-validation sample, \(N=183\)) were assembled for validation purposes. These patients were consecutively recruited at the outpatient clinic of the Department of Dermatology and the Institute of Pigmentary Disorders of the Academic Medical Center (AMC) in Amsterdam, the Netherlands, between February and March 2004. All patients were informed about the content of the study, orally as well as in writing, and gave their written informed consent before their participation in either one of the studies.
Consecutive included patients were diagnosed with a skin disease, were 18 years or older, and were mentally and/or physically able to complete the questionnaires. All patients were asked to complete the Skindex-29 at the time of inclusion. The central Ethics Committee AMC exempted this study for ethical approval. For non-interventional questionnaire research, this is common policy in the Netherlands. A written confirmation of this policy was given by the EC AMC.

Analysis
The data analyses comprised of two main components: i) factor analysis, and ii) OPLM analysis. In addition, we compared the reliability and known-groups validity of the 29 and 17-item Skindex versions.

Factor analysis
A preliminary forced two-factor structure exploratory factor analysis with varimax rotation was performed to test the dimensionality of the Skindex-29 in our Dutch sample by examining the proposed two-factor structure that was found by Nijsten and colleagues. 6

OPLM analysis
Subsequently, OPLM analysis was used (i) to first test the unidimensionality of the proposed Skindex-29 domains by examining their fit to the OPLM model; (ii) to obtain item weights indicating the precision by which an item can differentiate between levels of HRQoL (α) and item difficulty estimates (β); and (iii) to identify distinct Skindex-29 scale cut-points that mark clinically meaningful differences in HRQoL. For details of the OPLM analysis, see Supplementary Material and Methods, S1.

Skindex data was fitted to the OPLM model using a dichotomized response scale (see below). Subsequently, item weights and item difficulty estimates were identified on the basis of the OPLM model. Additionally, after fitting the Skindex-29 data to the OPLM model, we examined the percentage of agreement between the observed and the OPLM’s expected item scores (i.e., to examine to what extent the score patterns of patients could be predicted from their total revised Skindex-29 score). To facilitate the clinical interpretation of Skindex-29 score differences, the following procedure was used to identify Skindex-29 scale cut-points that mark meaningful differences in HRQoL. When the HRQoL estimate (i.e., level of HRQoL, also denoted as ‘θ’) for a Skindex-29 score exceeded the difficulty estimate (β) for an item, we expected an ‘often’ or ‘all the time’ response (score 1) to be present with more than 50% chance. If not, we expected a ‘never’, ‘rarely’, or ‘sometimes’ response (score 0). We examined what increase in a Skindex-29 domain score is needed to induce a score change on one or more Skindex-29 items, considering every one-increment step on any item (i.e., complaint) to be a meaningful difference in HRQoL. 16

Statistical analysis was done using the statistical software program SPSS, version 19.0 [SPSS inc. Chicago, ILL] and the OPLM software package for analyzing item responses (http://www.cito.com/research_and_development/psychometrics/psychometric_software.aspx).

Reliability and validity of the revised Skindex-29 and reduced Skindex-17
To obtain additional information on the psychometric quality of the revised Skindex-29 and the reduced 17-item version, we examined the reliability of both instruments by calculating...
the Cronbach’s α coefficient. The known-groups validity of both Skindex versions was examined by the strength of their association with psychiatric co-morbidity as determined by the 12-item General Health Questionnaire (GHQ-12) cut-off score (i.e., a score of 5 points or more).\textsuperscript{15} We constructed receiver operating characteristic (ROC) curves to compare the performance of the sum scores of both Skindex versions in discriminating patients with and without psychiatric co-morbidity. We expected that patients with higher Skindex scores (i.e., a more impaired HRQoL) were more likely to have psychiatric co-morbidity.

RESULTS

Based on complete data for the Skindex-29 (\(N=291\); 31 patients had missing values on one or more Skindex-29 items), the calibration sample consisted of 158 female patients (54.3%) with a mean of 49.1 years (SD=17.0; range=18.3\,-\,84.9). In the cross-validation sample (\(N=183\); 19 patients had missing values), 113 patients were female (61.7%) and the mean age was 53.1 years (SD=15.3; range=18.0\,-\,83.0). Table S2 shows the demographic and disease characteristics of both study populations (see Supplementary Material and Methods, S2). Sample characteristics were generally similar, although in the calibration sample there were more patients with psoriasis (43.3\% vs. 12.6\%).

Results of the factor analysis

Factor analysis indicated that the Skindex-29 consisted of two uni-dimensional domains: a Psychosocial domain (Eigenvalue=12.4) and a Symptoms domain (Eigenvalue=3.1), which is in agreement with the results found by Nijsten and colleagues.\textsuperscript{6} The first factor, consisting of 22 items (Psychosocial domain), explained 42.7\% of the variance. Factor loadings ranged from 0.43 to 0.85, except for item 2 (‘sleep’) with a factor loading of 0.23. Nevertheless, item 2 did clearly belong to the Psychosocial domain given its content. The second factor consisted of seven items, all from the original Skindex-29 Symptoms domain. Factor loadings ranged from 0.57 to 0.82, and explained another 10.6\% of the variation. This two-factor structure was replicated in our cross-validation sample, and similar results were found (data not shown but available on request).

Results of the OPLM analysis

Unidimensionality

A good fit to OPLM was found using the dichotomized response scale; goodness-of-fit statistic showed \(p=0.56\) for the Psychosocial domain and \(p=0.52\) for the Symptoms domain. The model also showed good fit when applied to the cross-validation sample; \(p=0.13\) for the Psychosocial domain and \(p=0.95\) for the Symptoms domain. The percentage of agreement between the observed and the OPLM’s expected item scores in the calibration sample ranged from 72.9 to 97.6\% (median 88\%). The percentage of agreement observed for the cross-validation sample ranged from 67.2 to 98.4\% (median 88\%).

Item weights and item difficulty estimates

Table 1 shows the item discrimination indices \(a_i\) and item difficulties \(\beta\) for the Psychosocial and Symptoms domains in the calibration sample, obtained after fitting the data to the OPLM model. Item discrimination indices may range from 1 (poor) to 5 (very high) and items
were ordered according to their level of difficulty. The item difficulty order was comparable for the calibration and cross-validation sample (data not shown but available on request). Our results show that six out of 12 items that were omitted from Skindex-29 to form the Skindex-17, had a good discriminative capacity ($a_i \geq 3$) (see the highlighted rows of Table 1).

### Table 1. Item discrimination ($a$) and item difficulty ($\beta$) for the Psychosocial and Symptoms domains in the calibration sample.

<table>
<thead>
<tr>
<th>Skindex-29 items(^1)</th>
<th>Psychosocial domain</th>
<th>$a$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. “deterioration”</td>
<td>2</td>
<td>-0.860</td>
<td></td>
</tr>
<tr>
<td>28. “annoyed”</td>
<td>2</td>
<td>-0.837</td>
<td></td>
</tr>
<tr>
<td>30. “tired”</td>
<td>3</td>
<td>-0.274</td>
<td></td>
</tr>
<tr>
<td>12. “ashamed”</td>
<td>3</td>
<td>-0.240</td>
<td></td>
</tr>
<tr>
<td>23. “frustrated”</td>
<td>3</td>
<td>-0.205</td>
<td></td>
</tr>
<tr>
<td>21. “embarrassed”</td>
<td>4</td>
<td>-0.177</td>
<td></td>
</tr>
<tr>
<td>15. “angry”</td>
<td>3</td>
<td>-0.169</td>
<td></td>
</tr>
<tr>
<td>5. “social life”</td>
<td>4</td>
<td>-0.121</td>
<td></td>
</tr>
<tr>
<td>3. “worried”</td>
<td>2</td>
<td>-0.118</td>
<td></td>
</tr>
<tr>
<td>4. “work or hobbies”</td>
<td>3</td>
<td>-0.069</td>
<td></td>
</tr>
<tr>
<td>11. “closeness with loved ones”</td>
<td>3</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>29. “sex life”</td>
<td>3</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>20. “interaction with others”</td>
<td>4</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>6. “depressed”</td>
<td>4</td>
<td>0.169</td>
<td></td>
</tr>
<tr>
<td>25. “desire to be with people”</td>
<td>5</td>
<td>0.198</td>
<td></td>
</tr>
<tr>
<td>9. “worried about scars”</td>
<td>2</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td>26. “humiliated”</td>
<td>4</td>
<td>0.242</td>
<td></td>
</tr>
<tr>
<td>17. “showing affection”</td>
<td>4</td>
<td>0.283</td>
<td></td>
</tr>
<tr>
<td>14. “do things alone”</td>
<td>4</td>
<td>0.326</td>
<td></td>
</tr>
<tr>
<td>2. “sleep”</td>
<td>1</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>8. “stay at home”</td>
<td>3</td>
<td>0.450</td>
<td></td>
</tr>
<tr>
<td>22. “problem for loved ones”</td>
<td>2</td>
<td>0.720</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Symptoms domain</th>
<th>$a$, $^1$</th>
<th>$\beta$, $^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>“itches”</td>
<td>1</td>
<td>-1.029</td>
</tr>
<tr>
<td>24.</td>
<td>“sensitive”</td>
<td>3</td>
<td>-0.653</td>
</tr>
<tr>
<td>19.</td>
<td>“irritated”</td>
<td>3</td>
<td>-0.612</td>
</tr>
<tr>
<td>7.</td>
<td>“burns or strings”</td>
<td>3</td>
<td>0.217</td>
</tr>
<tr>
<td>1.</td>
<td>“hurts”</td>
<td>3</td>
<td>0.456</td>
</tr>
<tr>
<td>16.</td>
<td>“bothered by water”</td>
<td>1</td>
<td>0.731</td>
</tr>
<tr>
<td>27.</td>
<td>“bleeding”</td>
<td>2</td>
<td>0.891</td>
</tr>
</tbody>
</table>

\(^1\) Items are ordered on their level of difficulty.
\(^2\) The grey lines indicate the 12 items that were discarded from the Skindex-29 in the development of the Skindex-17.
Skindex-29 scale cut-points that mark meaningful differences in HRQoL

Tables 2a and 2b show the scale cut-points for the revised Skindex-29 HRQoL scores. The revised Psychosocial domain sum scores, as obtained by weighting items by their discriminative capacity and with a dichotomized response scale, may range from 0-68 points. The revised sum scores of the Symptoms domain may range from 0-15 points.

There were 15 distinct scale cut-points for the Psychosocial domain (Table 2a), and six for the Symptoms domain (Table 2b). Skindex-29 items (i.e., complaints) are ordered from least difficult to most difficult to endorse, and each increase in a scale cut-point marks an extra complaint to be present (denoted with an ‘X’). For example, patients who have a revised sum score of 23 points on the Psychosocial domain (fourth scale cut-point), have more than 50% chance to endorse (i.e., answered ‘often’ or ‘all the time’) items 13 (‘deterioration’), 28 (‘annoyed’), 30 (‘tired’), 12 (‘ashamed’), and 23 (‘frustrated’). Patients in the lower adjacent

Table 2a. Skindex-29 Psychosocial domain scale cut-points marking a one-increment step of one or more items of the Skindex-29.

<table>
<thead>
<tr>
<th>Complaint present(\textsuperscript{a})</th>
<th>(\beta)</th>
<th>(\leq 5)</th>
<th>(\leq 19)</th>
<th>(\leq 21)</th>
<th>(\leq 23)</th>
<th>(\leq 25)</th>
<th>(\leq 28)</th>
<th>(\leq 32)</th>
<th>(\leq 34)</th>
<th>(\leq 40)</th>
<th>(\leq 43)</th>
<th>(\leq 46)</th>
<th>(\leq 48)</th>
<th>(\leq 53)</th>
<th>(\leq 61)</th>
<th>(\leq 68)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. “deterioration”</td>
<td>-0.860</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>28. “annoyed”</td>
<td>-0.837</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>30. “tired”</td>
<td>-0.274</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>12. “ashamed”</td>
<td>-0.240</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>23. “frustrated”</td>
<td>-0.205</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>21. “embarrassed”</td>
<td>-0.177</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>15. “angry”</td>
<td>-0.169</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5. “social life”</td>
<td>-0.121</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3. “worried”</td>
<td>-0.118</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4. “work or hobbies”</td>
<td>-0.069</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>11. “closeness with loved ones”</td>
<td>0.012</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>29. “sex life”</td>
<td>0.026</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>20. “interaction with others”</td>
<td>0.059</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>6. “depressed”</td>
<td>0.169</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>25. “desire to be with people”</td>
<td>0.198</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>9. “worried about scars”</td>
<td>0.232</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>26. “humiliated”</td>
<td>0.242</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>17. “showing affection”</td>
<td>0.283</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>14. “do things alone”</td>
<td>0.326</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2. “sleep”</td>
<td>0.356</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8. “stay at home”</td>
<td>0.450</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>22. “problem for loved ones”</td>
<td>0.720</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

\(\textsuperscript{a}\) Skindex-29 items ordered from least difficult to most difficult to endorse.

\(\textsuperscript{b}\) \(\beta\), item difficulty.

X, >50% probability for an item to be endorsed.
scale cut-point (≤21 points) have more than 50% chance to have endorsed the same items except item 23 (‘frustrated’).

The revised sum scores for each scale cut-point can easily be calculated from the original Skindex-29 scores. An example is provided in Table 3, illustrating a revised sum score of four for the Symptoms domain. First, the original response scale needed to be dichotomized into 0 (‘never’, ‘rarely’, and ‘sometimes’) and 1 (‘often’ and ‘all the time’). The endorsed items then needed to be multiplied with their discrimination indices (\(a_i\)), as listed in Table 1, and summed to obtain the weighted sum score. With a weighted Skindex-29 sum score of four, it is likely that this patient endorsed item 10 (‘itching’) and 24 (‘sensitive’), see also Table 2b. As the percentage of agreement between the observed and the OPLM’s expected item scores is high, the score pattern can be predicted from the scale score in terms of the complaints that are likely to be present.

**Results of the psychometric analysis of the revised Skindex-29 and Skindex-17**

We compared the reliability and validity of both Skindex versions in our calibration sample (N=291). We used the original 5-point response scale for the analysis. The reliability coefficients of the Psychosocial and Symptoms domains for the Skindex-29 were \( \alpha = 0.91 \) and \( \alpha = 0.80 \) respectively, and were comparable to those for the Skindex-17, namely: \( \alpha = 0.93 \) and \( \alpha = 0.77 \) respectively. Known groups validity of both Skindex versions were also comparable: no differences were found in the ROC area under the curve (AUC) statistics: AUC= 0.79 for the Skindex-29 and AUC=0.81 for the Skindex-17. This indicates a comparable overall classification accuracy of both instruments. In addition, comparable results were found for the Symptoms domain of both Skindex versions: AUC=0.71 and AUC=0.72 for the Skindex-29 and -17 respectively.

---

**Table 2b.** Skindex-29 Symptoms domain scale cut-points marking a one-increment step of one or more items of the Skindex-29.

<table>
<thead>
<tr>
<th>Complaint present*</th>
<th>( \beta^b )</th>
<th>Skindex-29 scale cut-points</th>
<th>( \leq 3 )</th>
<th>4</th>
<th>( \leq 9 )</th>
<th>10</th>
<th>( \leq 12 )</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. “itches”</td>
<td>-1.040</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>24. “sensitive”</td>
<td>-0.723</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>19. “irritated”</td>
<td>-0.681</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. “burns or strings”</td>
<td>0.256</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. “hurts”</td>
<td>0.480</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. “bothered by water”</td>
<td>0.812</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. “bleeding”</td>
<td>0.896</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Skindex-29 items ordered from least difficult to most difficult to endorse.

\( \beta \), item difficulty.

X, >50% probability for an item to be endorsed.
The results of the factor analysis and the subsequent OPLM analysis confirmed the two-factor structure proposed by Nijsten and colleagues. However, items clearly differed in their discriminative capacity and thus needed weighting before summation. After weighting items by their level of discrimination, and dichotomizing the response scale, a good fit to the unidimensional OPLM model for the revised version of the Skindex-29 was found in both the calibration and cross-validation sample. Of the 12 items that were omitted from the original Skindex-29 to form the 17-item version, six items had an above average level of discrimination (that is, an item weight of 3 or higher). Since items are weighted accordingly in calculating the sum score, these items represent a substantial part of HRQoL as measured with the Skindex-29. Both Skindex versions had similar internal consistency reliability and validity coefficients.

Fifteen scale cut-points for the Psychosocial domain and six scale cut-points for the Symptoms domain were identified that mark a one-increment step of one or more items, or complaints, of the Skindex-29. In addition, the response patterns of patients could be accurately predicted from their given Skindex-29 domain scores. The identified scale cut-points may be useful in monitoring patients’ HRQoL and also to judge the effects of therapy on patients HRQoL in clinical trials. For example, a treatment group with a mean sum score of 11 points versus a treatment group with a mean sum score of 17 points may be statistically significantly different but not clinically meaningful different, since the mean scores are associated with identical complaints (see Table 2a, second scale cut-point).

The Skindex-17 is a Rasch reduced version of the Skindex-29 and addresses the issue of the effect of external factors such as age, sex, diagnosis, or level of HRQoL. In the present study, we used the more flexible OPLM model, allowing items to differ in their discrimination capacity, which enabled us to retain all 29 Skindex items. The data presented showed that of the 12 items omitted from the Skindex-17, 11 items had a moderate to high discrimination capacity and, judging their content, measure important aspects of HRQoL in dermatology patients.

### Table 3. Calculation of a revised weighted sum score for the Symptoms domain from the original Skindex-29 scores.

<table>
<thead>
<tr>
<th>Skindex-29 items Observed answers of patient ‘X’</th>
<th>Original item score</th>
<th>Dichotomized item score</th>
<th>Item discrimination weight (a)</th>
<th>Weighted item score</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. “itches” Often 75</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>24. “sensitive” Often 75</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>19. “irritated” Sometimes 50</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7. “burns or strings” Sometimes 50</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. “hurts” Sometimes 50</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16. “bothered by water” Sometimes 50</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>27. “bleeding” Rarely 25</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Weighted sum score 4

*The original response scale is dichotomized into 0 (‘never’, ‘rarely’, and ‘sometimes’) and 1 (‘often’ and ‘all the time’). The endorsed items need to be multiplied with their discrimination indices (a) and summed to obtain the weighted sum score. With a weighted Skindex-29 sum score of four, it is likely that this patient endorsed item 10 (‘itching’) and 24 (‘sensitive’).
The revised Skindex-29 and the reduced Skindex-17 behave very similar in terms of reliability and validity. The choice in selecting an instrument consisting of 29 or 17 items thus remains arbitrary. For use in clinical practice, for example, the ideal situation would be that a HRQoL questionnaire is being completed by the patient prior to the consultation with the dermatologist, for example during waiting time. By doing so, the results can be directly discussed with the clinician during the consultation and can be integrated in patient management. The relatively small gain in completion time does, in our opinion, not outweigh the additional and valuable information that is being obtained when using all 29 items.

**Strengths and limitations**

Several methods have been used to obtain further insight in the interpretation of HRQoL scores. The strength of the analysis, to characterize score differences, is the invariance property of IRT modelling. Rather than the widely applied external anchors and distribution-based approaches, we employed the weighted sum score of the Skindex-29 as an internal anchor to estimate meaningful score differences. We found a high agreement between observed and expected item scores. This stems from the invariance property of IRT modelling and associated estimation methods (conditional maximum likelihood estimation) that make no assumptions on the distribution of HRQoL levels in the sample (i.e., a normal distribution). Therefore, the results of our study are likely to be generalizable to other samples. Further, as dichotomizing the response categories only applies to the calculation of the revised sum scores, there is no loss of information nor precision. However, the effect of external factors (such as age, sex, diagnosis, or level of HRQoL) on item responses, was not considered in our analysis. Second, even though we have found a good fit to the OPLM model in both samples, the cross-validation sample was relatively small for IRT analysis (N=183). Therefore, the fit of the data in the cross-validation sample to test the generalizability of the results, may be obtained by a lack of power to reject the OPLM model.

**Conclusion**

Items clearly differed in their discriminative capacity and thus needed weighting before summation. The reliability and validity of both the revised Skindex-29 and the reduced Skindex-17 were both comparable and considered to be adequate for its use in research and/or clinical practice. However, a considerable number of discriminating items that are of clinical importance were omitted from the Skindex-17. On a patient level, application of the Skindex-29 provides valuable information on item level that can be used in the patient management. In addition, with the results of this study we gained insight into the interpretation of Skindex-29 score differences and we provided an answer to the question what score difference represents a clinically meaningful difference. This may be also relevant in the judgment of treatment effects in randomized clinical trials.
REFERENCES


One-parameter logistic model analysis

We scaled the difficulty of the Skindex-29 domain items by using an extension of the item response theory (IRT) model, the one-parameter logistic model (OPLM). IRT assumes that a single “latent” health-related quality of life (HRQoL) variable determines the responses of individual patients on the test items. Latent means that the position of a respondent on the HRQoL scale can only be inferred indirectly from the item responses. A central concept in all IRT models is the item response function that specifies the probability to endorse a Skindex-29 item given the latent HRQoL level of the respondent. The assumed relation between the latent HRQoL variable and the item responses determines the choice for one model above the other. In the IRT model, the probability of endorsing an item statement depends on only one item characteristic, namely the difficulty level of an item. In the so-called two-parameter logistic models items are also characterised by a discrimination parameter that indicates the strength of the relation with the underlying HRQoL variable. OPLM combines the desirable properties of the IRT model - invariance of item and person measures and the raw sum score for a patient being a sufficient statistic for the unknown person parameter - with the greater flexibility of the two-parameter logistic model. With OPLM, items and patients can be calibrated on a common scale, and specific statistical tests are available to examine the fit of the OPLM model to the data. With dichotomous items it is assumed that the probability of endorsing an item \(X_i (X_i = 1)\) given the latent HRQoL level (\(\theta\)) is:

\[
P(X = 1|\theta) = \frac{\exp a_i (\theta - \beta_i)}{1 + \exp a_i (\theta - \beta_i)}
\]

The symbol ‘\(\theta\)’ denotes the latent HRQoL variable as measured with the Skindex-29. The item response function describes the probability of endorsing an item as a function of HRQoL level (\(\theta\)) and the difficulty (\(\beta_i\)) of the presented item. If the HRQoL level of a patient exceeds the difficulty of an item, the probability of endorsing the item is more than 50% and increases up to 100% as the difference between HRQoL level and item difficulty estimate increases. If the HRQoL level of a patient is lower than the difficulty of an item, the probability of endorsing an item will be less than 50% and will approach zero for patients with very low HRQoL scores (for Skindex-29, a lower score is indicative for a better HRQoL). When the HRQoL level estimate for a patient equals the estimated difficulty of an item, there is a fifty-fifty probability of endorsing the item in question. So, the item response function for an item links the probability of endorsing an item to the latent HRQoL level. The item difficulty (\(\beta\)) for an item corresponds to the position on the latent HRQoL continuum where the probability of endorsing the item is fifty-fifty. Item response functions of difficult items are located closer to one extreme of the latent scale than those of less difficult items.

The discrimination index (\(a\)) for an item indicates the steepness of the curve, with larger values for items having a steeper curve. An item with a higher index discriminates better in the ability region around the item parameters than an item with a lower index. The discrimination indices are imputed as integer values allowing conditional maximum likelihood estimation of the item difficulties, a prerequisite for “sample free” estimation and invariant estimates of
item and person measures. Fit measures, so-called M-tests (see below) are available and are informative with respect to the direction in which the discrimination indices have to be changed to obtain a better overall fit of the data to the OPLM model. The sum of the item scores, weighted by the discrimination indices, is used to calculate domain scores for the patients. Overall fit of the model can be assessed with the global fit-statistic R1c. Fit statistic p-values exceeding 0.05 indicate acceptation of the OPLM model. Fit of individual items are assessed by the Chi-square based S-test and three M-tests. The M-test statistics are sensitive to differences in discrimination capacity between items. To calculate M-tests, the patient scores are partitioned in a low, middle and high score group. For each score group, the expected number of patients endorsing an item is calculated using the fitted model. The differences with the observed numbers follow a Standard Normal distribution. Negative M-test-statistics for an item indicate upgrading of the discrimination index while positive M-tests indicate that downgrading of the discrimination index is needed to increase the item fit.
### Supplementary Material and Methods S2. Demographic and disease characteristics of the calibration sample and cross-validation sample

<table>
<thead>
<tr>
<th></th>
<th>Calibration sample</th>
<th>Cross-validation sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=291</td>
<td>N=183</td>
</tr>
<tr>
<td><strong>Female, n (%)</strong></td>
<td>158 (54.3)</td>
<td>113 (61.7)</td>
</tr>
<tr>
<td><strong>Mean age in years (SD, range)</strong></td>
<td>49.1 (17.0, 18.3-84.9)</td>
<td>53.1 (15.3, 18.0-83.0)</td>
</tr>
<tr>
<td><strong>Diagnoses, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Acne</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acne vulgaris</td>
<td>10 (3.4)</td>
<td>6 (3.3)</td>
</tr>
<tr>
<td>Hidradenitis suppurativa</td>
<td>4 (1.4)</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>Rosacea</td>
<td>2 (0.7)</td>
<td>8 (4.4)</td>
</tr>
<tr>
<td><strong>Autoimmune disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bullous pemphigoid</td>
<td>1 (0.3)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Lichen planus</td>
<td>5 (1.7)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Morfea</td>
<td>2 (0.7)</td>
<td>-</td>
</tr>
<tr>
<td>Parapsoriasis</td>
<td>2 (0.7)</td>
<td>-</td>
</tr>
<tr>
<td>Pemphigus vulgaris</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Systemic lupus erythematoses</td>
<td>1 (0.3)</td>
<td>8 (4.4)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td><strong>Benign skin and vascular tumors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keratosis actinica</td>
<td>6 (2.1)</td>
<td>10 (5.5)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>2 (0.7)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dermatitis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eczematous lesions</td>
<td>70 (24.1)</td>
<td>26 (14.2)</td>
</tr>
<tr>
<td>Prurigo</td>
<td>-</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>126 (43.3)</td>
<td>23 (12.6)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>-</td>
<td>5 (2.7)</td>
</tr>
<tr>
<td><strong>Folliculitis</strong></td>
<td>-</td>
<td>6 (3.3)</td>
</tr>
<tr>
<td><strong>Genetic disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ichthyosis</td>
<td>1 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td>Pemphigus benigus familiaris chronicus</td>
<td>1 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Genital skin disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condylomata acuminate</td>
<td>2 (0.7)</td>
<td>-</td>
</tr>
<tr>
<td>Lichen sclerosis</td>
<td>8 (2.7)</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td><strong>Hair and scalp disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alopecia androgenata</td>
<td>1 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td>Alopecia areata</td>
<td>1 (0.3)</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Frontal fibrosing alopecia</td>
<td>1 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td>Pseudopelade of brocq</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td><strong>Infectious diseases</strong></td>
<td>-</td>
<td>10 (5.5)</td>
</tr>
</tbody>
</table>
## Supplementary Material and Methods S2. Continued

<table>
<thead>
<tr>
<th>Malignancies</th>
<th>Calibration sample N=291</th>
<th>Cross-validation sample N=183</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal cell carcinoma</td>
<td>6 (2.1)</td>
<td>10 (5.5)</td>
</tr>
<tr>
<td>Melanoma</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Morbus Paget</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>1 (0.3)</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>3 (1.0)</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pigmentary disorders</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Naevus naevocellularis</td>
<td>-</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>Naevus of Ota</td>
<td>1 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td>Post-inflammatory hyper pigmentation</td>
<td>-</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Vitiligo</td>
<td>5 (1.7)</td>
<td>1 (0.5)</td>
</tr>
</tbody>
</table>

| Pruritus                          | 5 (1.7)                  | -                            |

| Reactive skin disorders and drug reactions | 5 (1.7) | - |

<table>
<thead>
<tr>
<th>Ulcers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ulcus cruris</td>
<td>-</td>
<td>6 (3.3)</td>
</tr>
<tr>
<td>Not otherwise specified</td>
<td>4 (1.4)</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Urticarial disorders</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naevus flammeus</td>
<td>-</td>
<td>7 (3.8)</td>
</tr>
<tr>
<td>Purpura pigmentosa progressiva</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Teleangiectasia nno</td>
<td>-</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Varices</td>
<td>1 (0.3)</td>
<td>8 (4.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Viral skin lesions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verrucae plantaris</td>
<td>3 (1.0)</td>
<td>-</td>
</tr>
<tr>
<td>Verrucae seborrhoica</td>
<td>-</td>
<td>2 (1.1)</td>
</tr>
</tbody>
</table>

| Other                             | 3 (1.0)                  | 13 (7.1)                     |

<table>
<thead>
<tr>
<th>Skindex-29* score (median, IQRb)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms</td>
<td>50.0 (28.6-64.3)</td>
<td>46.4 (28.6-60.6)</td>
</tr>
<tr>
<td>Emotions</td>
<td>35.0 (20.0-52.5)</td>
<td>35.0 (20.0-55.0)</td>
</tr>
<tr>
<td>Functioning</td>
<td>22.9 (8.3-39.6)</td>
<td>20.8 (6.3-39.6)</td>
</tr>
<tr>
<td>Overall</td>
<td>31.9 (21.6-50.0)</td>
<td>31.0 (19.0-49.1)</td>
</tr>
</tbody>
</table>

* The domain scores and the overall score are expressed on a 100-point scale, with higher scores indicating a lower level of HRQoL.

b IQR, interquartile range.
4

HEALTH-RELATED QUALITY OF LIFE APPLICATION IN CLINICAL PRACTICE
Health-related quality of life assessment in dermatologic practice: relevance and application

OD van Cranenburgh˚
CAC Prinsen˚
MAG Sprangers
Phl Spuls
J de Korte
˚In alphabetical order; both authors contributed equally to this paper

Dematol Clin 2012; 30(2):323-32
The following is a brief summary of important points and objectives for recall:

- Health-related quality of life (HRQoL) data of patients may be used for various purposes: (1) to increase a patient’s self-awareness and empowerment, (2) to increase patient-centeredness in health care, (3) to make an optimal choice for treatment, (4) to monitor treatment over time and determine treatment effectiveness, and (5) to improve treatment outcome.

- HRQoL assessment is particularly relevant for patients with chronic skin diseases that are known to have substantial and enduring adverse effects on HRQoL.

- Many HRQoL questionnaires are currently available. The selection of a HRQoL questionnaire will depend on several factors, such as the functions it has to fulfill in clinical practice, the specific patient population, the psychometric characteristics of a specific questionnaire, and the local policy and conditions.

- We have chosen the Skindex-29 as the questionnaire of first choice to be used in dermatology.

- An electronic assessment may facilitate the application of HRQoL in dermatologic practice.

- To use HRQoL data in clinical practice, scores should be interpreted promptly and accurately. Information on the interpretation of Skindex-29 scores is currently available.

- In discussing HRQoL scores, it is important not to focus on the overall score of the Skindex-29, but on the 3 domain scores.
INTRODUCTION

Patient-reported outcomes (PROs) are reports or assessments of any aspect of a patient’s health status or impact of treatment that come directly from the patient, without the interpretation of the responses by anyone else. Regulatory agencies in many countries take patient-relevant criteria into consideration in decisions on reimbursement of new therapies, resulting in an increased importance of PROs in clinical trials. The application of PROs in clinical practice is growing as well. Assessment of PROs, such as patients’ experienced disease severity, health-related quality of life (HRQoL), treatment adherence, and treatment satisfaction, appears to have added value for daily clinical practice.

In a systematic review of studies on the impact of PRO assessment in clinical practice, Valderas and colleagues stated that (1) PRO assessment can be time consuming, (2) both patients and physicians may perceive PRO questionnaires as burdensome, (3) the interpretation of PRO scores in a clinically meaningful manner requires additional resources, and (4) the implications for treatment are not apparent. On the other hand, PRO assessment can also have a positive impact on clinical practice, specifically by improving the diagnosis and recognition of problems, and in patient–physician communication. The investigators also pointed out that studies included in their review were heterogeneous and of an inferior methodological quality and that, as a result, no evident conclusion could be drawn with regard to the effect of PRO assessment in clinical practice.

An important PRO in health care is HRQoL. HRQoL reflects patients’ evaluation of the impact of disease and treatment on their physical, psychological, and social functioning and well-being. Chronic skin diseases, such as acne, eczema, hidradenitis suppurativa, psoriasis, and vitiligo, have been found to adversely affect patients’ HRQoL. In many patients, this impact is profound.

In such chronic skin diseases, dermatologic treatment can offer a temporary suppression and/or remission of severity and symptoms. As a result, many patients have to cope with the burden of their skin disease for years, or even throughout their entire lives. Patients often consider improvement of HRQoL as an important treatment goal; hence, dermatologic treatment should aim to decrease disease severity and to increase patients’ HRQoL.

HRQoL is gradually becoming a standard outcome parameter in clinical studies and health care management. Because the major goal of therapeutic interventions is to make patients feel better, HRQoL assessment is likely to become even more important in the future. Because of this development, the quality of HRQoL assessment itself, correct management and interpretation of HRQoL data, and the communication of such data with the patient, deserve attention.

HRQoL is generally measured with reliable and valid self-reported instruments (i.e., questionnaires). The application of such questionnaires in daily clinical practice may improve evidence-based practice, facilitate communication with the patient, and, herewith, the process of shared decision making between patients and physicians. In a randomized controlled trial (RCT) in the field of oncology, HRQoL assessment resulted in a significant increase of relevant information on and discussion of chronic symptoms; moreover, the explicit use of HRQoL information during patients’ consultation was associated with a significant improvement in patients’ well-being. Another RCT indicated that HRQoL assessment in
daily clinical oncology practice facilitates the discussion of HRQoL issues and heightens physicians’ awareness of their patient’s HRQoL.\textsuperscript{10} Nevertheless, the application of HRQoL assessment is not customary in dermatologic practice and there are several practical and attitudinal barriers. A deeper understanding of the benefits of HRQoL assessment for both dermatologists and patients may improve its application. Hence, members of a Dutch expertise center on HRQoL in dermatology took the initiative to start a working group, consisting of 10 dermatologists, a psychologist, and a clinical epidemiologist, on HRQoL assessment in clinical practice. This working group produced a guideline to support the application of HRQoL assessment in routine dermatologic practice.\textsuperscript{11} In this article, and following this guideline, we attempt to provide answers to the 3 questions: (1) What is the relevance of HRQoL assessment to dermatologic practice? (2) Which patients would benefit most from routine HRQoL assessment? (3) How can HRQoL assessment be applied in clinical practice? In answering these questions, we aim to contribute to the discussion on and the implementation of HRQoL assessment in routine dermatological practice.

**HRQoL ASSESSMENT IN DERMATOLOGY: WHY?**

**Patients’ self-awareness and empowerment**

By filling out an HRQoL questionnaire and communicating about the answers to the questions, patients may gain more insight into the impact of the skin disease on their own physical, psychological, and social functioning and well-being. Most likely, this insight will increase patients’ self-awareness; for instance, awareness of specific psychological problems and of specific health care needs. Such awareness, and the acknowledgement of needs by the dermatologist, may further empower patients to share and discuss their problems with significant others, such as a partner, relatives and friends.\textsuperscript{12-14}

**Patient-centered health care**

Clinical evaluations of the severity of a skin disease are not highly correlated with patients’ perceptions of HRQoL.\textsuperscript{15,16} Consequently, HRQoL assessment is of particular importance to enable the dermatologist to grasp the impact of the skin disease and/or its treatment on an individual patient.\textsuperscript{17} In addition, such data may highlight specific aspects of HRQoL that are affected the most, for instance shame or depression. In patients with a chronic skin disease, this information might be of relevance, particularly because of a relatively high prevalence of psychosocial problems, often hidden “under the skin”.\textsuperscript{7,14,18}

Furthermore, insight into these problems creates an opportunity to communicate in an empathic and responsive way, thereby supporting patients in coping with their problems more effectively. Communication about HRQoL may also be helpful in engaging patients in a discussion on treatment preferences to allow mutual or shared decision making.

**An optimal choice of treatment**

HRQoL data, in addition to clinical information, contribute to a more comprehensive insight into a patient’s situation after diagnosis and before to the choice for a specific treatment. By including HRQoL data into the decision-making process, the dermatologist and the patient can
make an optimal, shared choice for a specific treatment in terms of its setting (e.g., inpatient, outpatient, day care, specialty care), intensity or invasiveness, position in the conceivable order of treatments over time, and/or combinations with other treatments. For instance, if a patient experiences a high level of symptomatic burden, a more intensive or invasive treatment can be considered. A better-tailored treatment is expected to be better tolerated and adhered to by the patient. Additionally, feasible aims of a specific treatment can be discussed using HRQoL data; for instance, a reduction of itch, a decrease in or clearance of visible lesions, or a reduction in the degree of disease severity within a specific time frame.

Furthermore, patients’ needs for additional care, as a supplement to regular dermatologic care, can be identified and addressed. Some patients may experience low levels of HRQoL that cannot be explained by disease severity only. Other patients may have serious problems with respect to specific domains or aspects of HRQoL, such as suffering from depression, feeling socially isolated, or encountering problems at work. In such instances, referral to a social worker, a psychologist, or a psychiatrist might be indicated, and can result in a valuable adjuvant therapy.

Monitor treatment over time and determine treatment effectiveness

HRQoL scores of a patient before treatment may be compared with scores at follow-up visits. In this way, the treatment process can be monitored over time. HRQoL data obtained at follow-up visits may also be helpful in checking negative consequences or side effects of treatments; for instance, an increase of itch, pain, irritation, tiredness, sleep, or depression. Such HRQoL data alert the dermatologist to adjust the treatment whenever necessary (e.g., dose, switch treatment, combination with other treatment); moreover, this tailored treatment is expected to be better tolerated and adhered to by the patient.

In the end, after completion of treatment, HRQoL scores of a patient can be compared with scores before treatment. An improvement in HRQoL, which is a main treatment goal for many patients,5,6 can be monitored, and may indicate treatment effectiveness.

Improvement of treatment outcome

Although the aforementioned functions suggest that application of HRQoL primarily has a positive effect on the process of health care, a positive effect on the outcomes of dermatologic treatment itself is expected as well. Empowerment of the patient, patient-centered health care, an optimal choice for treatment, monitoring treatment over time, and the explicit attention to HRQoL and/or the patient’s point of view is likely to have a positive impact on the patient’s HRQoL, treatment satisfaction, and disease severity. As evidence suggests that clinical and psychological outcomes, such as adherence to treatment advice, are optimized when patients’ emotional concerns are addressed, it is critical to recognize and manage the psychological needs of patients.19

Because of lack of evidence in dermatology concerning the aforementioned functions, a randomized controlled trial (Dutch Trial Register, NTR1364) was started to assess the efficacy of HRQoL assessment and HRQoL communication in dermatologic practice. The study is ongoing and we expect to publish the results in 2012.
HRQoL assessment in dermatology: WHO?

HRQoL assessment is particularly relevant for patients with chronic skin diseases that exert a large, negative impact on HRQoL. Psoriasis and eczema have been found to induce substantial decreases in patients’ HRQoL.4,20-22 These skin diseases also have high incidence rates, a high degree of chronicity, and may require long-term treatment; however, many more skin diseases affect HRQoL adversely, including acne, alopecia areata, hand/foot eczema, hidradenitis suppurativa, lichen planus, lichen sclerosus, pruritus/prurigo, seborrheic eczema, ulcers, urticaria, and vitiligo.

In addition to a specific diagnosis, the degree of disease severity, social visibility of the condition, age, personal circumstances, and the presence or absence of social support may influence patients’ HRQoL. So, a patient with severe psoriasis may experience a relatively good HRQoL, whereas another patient with only a mild degree of eczema may experience a relatively poor HRQoL. Assessments are thus relevant whenever a negative impact on HRQoL is suspected and whenever treatment does not meet the patient’s expectations.

The treatment setting itself may also play a role in selecting patients for HRQoL assessment. The inclusion of patients may be influenced by local policy, local conditions, presence or absence of facilities, and availability of staff. For instance, some dermatologists prefer to integrate HRQoL assessment in inpatient care or in a day care center rather than in outpatient settings. This preference may arise from the availability of sufficient room, accommodation, and staff; longer duration of treatment; and/or feasibility of counseling.

Others prefer integration in specific outpatient consultation hours, for instance a biologic therapy consultation hour in a psoriasis treatment center.

Although HRQoL assessment can be applicable to all aforementioned patients and settings, we do not recommend assessments in all attending patients. In patients with a skin disease that hardly affects their HRQoL, such as in most patients with actinic keratoses, naevi, warts, and onychomycosis, or in skin diseases where a single consultation or short-term treatment is sufficient, it does not appear to be of relevance. Last, HRQoL assessment should not induce aversion or resistance, for instance in patients who consider questions on psychosocial functioning as unnecessary, intrusive, or inappropriate.

HRQoL assessment in dermatology: HOW?

HRQoL questionnaires

There are simple ways to ask patients about their HRQoL, for instance by asking “How does your skin disease affect your daily life?” In fact, many dermatologists do ask patients how they are doing, and many patients do inform their dermatologists spontaneously about the impact of their skin disease on aspects of HRQoL, for instance on their mood, work, or family life. To collect data in a more objective and systematic way, however, reliable and validated HRQoL questionnaires may be required.

HRQoL questionnaires consist of a number of items or questions, most often to be answered by ticking off a multiple-choice answer. Multiple-choice responses may refer to intensity (e.g., from “mild” to “severe”) or frequency (e.g., from “never” to “all the time”) or may invite an opinion
with respect to given statements (e.g., from “strong disagreement” to “strong agreement”). Because HRQoL is a multidimensional construct (e.g., consisting of a physical, psychological, and social functioning domain), responses may result in domain scores, as well as an overall score. Currently, many questionnaires are available. In general, these can be distinguished into generic and specific HRQoL questionnaires. Generic questionnaires can be used for the measurement of HRQoL in all kinds of diseases and in the general population, whereas specific questionnaires are designed for the measurement of HRQoL in a specific disease, subgroup of disease, group of diseases, or patient population. Within dermatology, we distinguish dermatology-specific questionnaires designed for all kinds of skin diseases, and disease-specific questionnaires designed for a specific skin disease, for instance eczema or rosacea. The selection of an HRQoL questionnaire will depend on many factors, such as the functions it has to fulfill in clinical practice, the specific populations of patients, local policy and local conditions, and the psychometric characteristics of a specific questionnaire.

One could start using a simple, practical question to screen HRQoL, for instance a “questionnaire” consisting of only one question with multiple-choice responses: “To what extent does your skin disease affect your quality of life?” A more comprehensive questionnaire is the Dermatology Life Quality Index (DLQI), originally developed for routine dermatologic practice. The DLQI consists of only 10 questions and mainly focuses on limitations, for instance limitations in daily and social activities.

De Korte and colleagues and Both and colleagues systematically reviewed the quality of generic and dermatology-specific HRQoL questionnaires that are used in dermatology. For research, they recommended the use of a generic questionnaire in combination with a dermatology-specific questionnaire. A dermatology-specific questionnaire was explicitly recommended, as it encompasses all relevant dermatologic aspects and domains that a generic questionnaire may not include. For dermatology, the dermatology-specific questionnaire Skindex-29 was recommended.

Based on these reviews, we have chosen the Skindex-29 as the questionnaire of first choice for dermatologic practice in the Netherlands. Therefore, in this article we illustrate application of HRQoL assessment with the Skindex-29; however, there could be many reasons for making a different choice in different situations.

The Skindex-29 is a multidimensional questionnaire, assessing HRQoL during the past week, and consisting of 29 items that form 3 domains: symptoms, emotions, and functioning. Box 1 provides an overview of all Skindex-29 items, categorized per domain. Items are answered on a 5-point scale: Never = 0, Rarely = 25, Sometimes = 50, Often = 75, and All the Time = 100. The overall and domain scores are expressed on a 100-point scale, where higher scores indicate lower levels of quality of life. One item (item 18) about possible side effects of medication and/or treatment has been added to the questionnaire, but is not included in one of the domains, nor in the calculation of the overall score and domain scores. Research on the psychometric characteristics of the Skindex-29 indicated its reliability and validity. The Skindex-29 is currently available and psychometrically tested in many languages, including English, Dutch, French, German, Italian, and Spanish.
**Box 1. An overview of Skindex-29 items categorized per domain.**

### Symptoms
1. My skin hurts
7. My skin condition burns or strings
10. My skin itches
16. Water bothers my skin condition (bathing, washing hands)
19. My skin is irritated
24. My skin is sensitive
27. My skin condition bleeds

### Emotions
3. I worry that my skin condition may be serious
6. My skin condition makes me feel depressed
9. I worry about getting scars from my skin condition
12. I am ashamed of my skin condition
13. I worry that my skin condition may get worse
15. I am angry about my skin condition
21. I am embarrassed by my skin condition
23. I am frustrated by my skin condition
26. I am humiliated by my skin condition
28. I am annoyed by my skin condition

### Functioning
2. My skin condition affects how well I sleep
4. My skin condition makes it hard to work or do hobbies
5. My skin condition affects my social life
8. I tend to stay at home because of my skin condition
11. My skin condition affects how close I can be with those I love
14. I tend to do things by myself because of my skin condition
17. My skin condition makes showing affection difficult
20. My skin condition affects my interactions with others
22. My skin condition is a problem for the people I love
25. My skin condition affects my desire to be with people
29. My skin condition interferes with my sex life
30. My skin condition makes me tired

### Side effects
18. I worry about side effects from skin medications/treatments

---

**Electronic HRQoL assessment**

Assessment of HRQoL with a paper questionnaire has the advantage of simplicity. Nevertheless, it has several disadvantages. Apart from all the paper work, and the integration of data into the
medical record, it also implies the calculation of domain and overall scores by hand. Previous research showed that pen-and-paper HRQoL questionnaires that have to be scored by hand take too much time and are costly in the long-term. Previous research showed that pen-and-paper HRQoL questionnaires that have to be scored by hand take too much time and are costly in the long-term. Electronic assessment, on the other hand, may lower the resource burden and thereby encourage a more widespread use in clinical practice.8 To facilitate the application of HRQoL assessment in dermatologic clinical practice in the Netherlands, an electronic version of the Skindex-29 was developed. This enables patients to complete the questionnaire on a computer by touching the answers of choice on the screen. To our knowledge, this is the only dermatology-specific HRQoL questionnaire that is currently available as an electronic application. Completion of the electronic Skindex-29 takes no longer than five minutes and, immediately after answering all questions, all data are available: an overview of questions and answers arranged per domain, and an overview of scores visualized in a bar chart. Both can be printed, and answers that are bothersome, that is, marked with “often” or “all the time,” are displayed in bold on the screen as well as in the printout. International research indicates that electronic HRQoL assessment is increasingly applied within other specialties (e.g., oncology and hepatology) and has many advantages for clinical practice.33-38 Of course, a main advantage of this electronic version, compared with the paper version, is that answers and scores are immediately available and, thereby, facilitate “real-time” discussion with the patient, directly after the assessment (Figure 1). For this discussion, answers are displayed on screen, but it is preferable that both the dermatologist and the patient have a printed overview of the results as well. Answers in bold can serve as guidance during the patient’s consultation with the dermatologist, and the patient may take home the print-out, enabling him or her to discuss the results with significant others. The dermatologist may easily include the printed overview of HRQoL results in the patient medical record or may link the results to an electronic patient record. Another advantage of the electronic version, compared with the paper version, is that patient’s answers and scores are saved in a Skindex file and are automatically transferred to an MS Excel file. Saved scores can be used to analyze data on a group level and to follow the course of HRQoL of an individual patient over time: repeated assessments are graphically

Figure 1. Discussion of domain scores with the patient, directly after the self-assessment on a computer.
presented in a single bar chart (Figure 2). In this way, the treatment process of a specific patient can be monitored and the effect of a treatment can be determined. Anonymous scores on a group level can be used for scientific research as well.

**Interpretation of scores**

To use HRQoL data in clinical practice, scores should be interpreted promptly and accurately; however, the interpretation of scores of any HRQoL questionnaire is not straightforward and has little of no direct meaning. A score of 18 or 63 on a scale running from 0 to 100 cannot be interpreted without additional resources. Two types of methods to establish a clinically meaningful interpretation of HRQoL scores exist: distribution-based and anchor-based methods.7 For interpretation of scores on the aforementioned DLQI we refer to Hongbo and colleagues.39 In the case of the Skindex-29, a first psychometric study on the distribution of Skindex-29 scores among 454 Italian patients with various skin diseases resulted in a categorization of levels of severity,40 namely “very little,” “mild,” “moderate,” “severe,” and “extremely severe.” This study, using a distribution-based method, was a first attempt to interpret Skindex-29 scores.

In a study performed in the Netherlands among 339 patients at 9 general dermatologic outpatient clinics, Prinsen and colleagues7,41 used an anchor-based method to identify patients with mild, moderate, and severe impairment of HRQoL. The resulting cutoff scores for the Skindex-29 are presented in Table 1. To illustrate this: a patient with a score of 39 or higher on the emotions domain is likely to have a severe impairment, and a patient with a score of 32 or higher on the functioning domain is likely to have moderate impairment of HRQoL.

At first glance, these cutoff scores are not easy to apply in clinical practice; however, by using different colors in the bar chart (green for mild, yellow for moderate, and red for severe impairment), one is able to make prompt interpretations at a glance. To facilitate the application of the cutoff scores, one may also, as a rule of thumb or memory aid, round off the cutoff scores for mild, moderate, and severe impairment of HRQoL to 20 or higher, 30 or higher, and 40 or higher, respectively, with the exception of the symptoms domain.41 Patients with scores equal to or above the presented cutoff scores for “severe” (see Table 1) in at least 1 of the 3 domains are significantly affected by their skin disease. Prinsen

![Figure 2. Repeated assessments of Skindex-29 over time: an overview of domain scores.](image)
Table 1. Identification of patients with mild, moderate, and severe impairment of HRQoL according to Prinsen and colleagues.

<table>
<thead>
<tr>
<th>Level of Severity</th>
<th>Skindex-29 domains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Symptoms</td>
</tr>
<tr>
<td>Mild</td>
<td>≥39</td>
</tr>
<tr>
<td>Moderate</td>
<td>≥42</td>
</tr>
<tr>
<td>Severe</td>
<td>≥52</td>
</tr>
</tbody>
</table>


and colleagues’ indicated that these scores may signal a need for (adjustment of current) treatment or for additional care or support, but that scores do not automatically indicate what kind of treatment, care, or support is appropriate, and therefore the specific needs of an individual patient should be explored in direct contact with the patient. Although the preceding paragraphs provide some guidance in the interpretation of scores and answers, it may be clear that discussion with the patient may yield important additional information and promote a patient-specific interpretation.

Discussion of HRQoL data

Once a patient has completed the questionnaire, the summation of the domain scores provides a specific profile of a patient’s HRQoL. This profile may indicate which domain of HRQoL was influenced most during the preceding week. Because large differences may exist among the 3 domain scores of a patient, it is important not to discuss the overall score solely, but to focus on the domain scores separately. In fact, 2 patients may have about the same overall score, but when taking the domain scores into account, it might appear that the impact of the skin disease focuses on different domains of HRQoL. This is illustrated in Figure 3.

To focus on HRQoL in greater detail, answers to single questions might also be of relevance. In fact, we recommended this, as single questions often provide relevant, additional information that can be used in clinical practice. For instance, a patient may have a low symptoms score but could have indicated “often” or “all the time” to the question about itch, or a low functioning score but with answers “often” or “all the time” to questions about tiredness or sleep. This may signal patients’ specific needs. In Box 2, a case is described to illustrate the relevance of HRQoL assessment in dermatologic practice, using the Skindex-29 as an example.

SUMMARY

The aim of this article was to contribute to the discussion on and the implementation of HRQoL assessment in routine dermatologic practice. With respect to the relevance of HRQoL assessment, we focused on self-awareness and empowerment, patient-centered health care, and optimal dermatologic treatment. With respect to patients, we focused on patients with a chronic skin disease in which substantial and enduring adverse effects on
HRQoL are suspected. With respect to implementation in routine practice, we described the application of the electronic version of the Skindex-29 in the Netherlands. We realize we have presented an optimistic view on HRQoL assessment in clinical practice. We would like to stress that a clinically meaningful interpretation of scores and the implications for dermatologic treatment and care are not self-evident. We already referred to a systematic review in which no unambiguous conclusion could be drawn with regard to the effect of PRO assessment in clinical practice;³ and even in our local, Dutch situation, there are some drawbacks to report. For instance, installation of the Skindex-29 software was sometimes complicated by technical limitations, such as information technology system requirements or safety boundaries that differ from organization to organization. Another problem in some of the clinics was lack of accommodation and staff, and budget problems with respect to computers, laptops, or pocket computers for the patients. Meanwhile, to overcome some of the technical limitations, a Web-based version of the Dutch Skindex-29 has been developed. This Web-based version enables patients to gain access to the Skindex-29 from any computer with an Internet connection. In the Netherlands, a considerable percentage of the citizens have a personal computer with an Internet connection at home. As Skindex data are saved on an external, protected server, potential technical problems during installation of software or constraints owing to safety requirements in a clinic are avoided. In the future, other HRQoL and PRO questionnaires can be added, thus resulting in a “Web portal” for PROs.
in dermatology. For instance, questionnaires measuring disease severity from a patient’s perspective, adherence, or treatment satisfaction.

Although the available evidence on the added value of HRQoL assessment for clinical practice, especially the application of PRO assessment in general, is ambiguous, we believe that initiatives to integrate HRQoL data into the management of patients are most welcome, and will help create a more solid body of evidence. If HRQoL is considered to be an important outcome of routine dermatologic treatment, would it not be a bit “careless” not to measure this patient-reported outcome?

**ACKNOWLEDGEMENTS**

The authors thank the members of the Dutch Working Group “Quality of Life Assessment in Dermatologic Practice” (in alphabetical order): MTW Gaastra, MD; DB de Geer, MD; AY Goedkoop, MD, PhD; CLM van Hees, MD; WJA de Kort, MD; CM Legierse, MD; TEC Nijsten, MD, PhD; MCG van Praag, MD, PhD; MLA Schutteelaar, MD; AME Visser–van Andel, MD.
REFERENCES


23. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for


The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: study protocol for a multicenter randomized controlled trial

CAC Prinsen
Phl Spuls
MAG Sprangers
MA de Rie
CM Legierse
J de Korte
ABSTRACT

Background: Interest in health-related quality of life (HRQoL) outcome research in dermatology is increasing, especially in the systemic treatment of psoriasis with biologic agents. In other specialties, such as oncology, the application of a HRQoL intervention is considered to be an aid for monitoring disease and treatment over time, for the communication with the patient, and for improving treatment outcome. However, in dermatology practice, the application of this intervention is relatively new. Moreover, evidence on the effectiveness of a HRQoL intervention in dermatology is missing. It is hypothesized that the application of a HRQoL intervention in dermatology practice will have a positive impact on patients’ HRQoL as well as on doctor-patient communication.

Methods/design: In a prospective multicenter cluster randomized controlled trial, patients diagnosed with moderate to severe psoriasis who receive biologic treatment, will be followed for 48 weeks. The study sites, and not the patients, will be randomly allocated via a computer-based randomization system to either the intervention (treatment with etanercept and standardized HRQoL assessment and communication) or the control group (treatment with etanercept alone). The HRQoL intervention will include 1) the electronic assessment of the Skindex-29, a well-studied dermatology-specific HRQoL questionnaire, and 2) the communication of the resulting Skindex-29 data with the patient. Prior to study start, dermatologists in the intervention group will be educated and trained in standardized HRQoL assessment and communication using the Skindex-29. At six consecutive visits, patients at study sites in the intervention group will be asked to complete the Skindex-29 on a desk-top pc at the clinic, just before their consultation with the dermatologist. A print-out of the completed questionnaire will be made and, guided by this print-out, feedback on the HRQoL scores will be given during the consultation. Primary outcome parameters are the impact of the HRQoL intervention on patients’ HRQoL, and the effect of the HRQoL intervention on doctor-patient communication. Secondary outcomes include health status and disease severity.

Trial registration: The Netherlands National Trial Register (NTR): NTR1364.
**INTRODUCTION**

Patient-reported outcomes (PROs) are any aspect of a patient’s health status that comes directly from the patient without the interpretation of anyone other than the patient. PROs provide information on the disease or treatment from a patient’s perspective. Examples of PROs are disease severity, general health status, adherence to treatment, satisfaction with treatment, and health-related quality of life (HRQoL). In accordance with the definition of the World Health Organization, HRQoL can be defined as a reflection of patients’ physical, psychological, and social functioning and well-being, and is generally considered as a key outcome parameter. Chronic skin diseases, such as psoriasis, are known to have a relatively high, negative impact on HRQoL. As a result, interest in HRQoL outcome research in dermatology is increasing, especially in the systemic treatment of psoriasis with biologic agents. Psoriasis is a systemic inflammatory skin disease with increased epidermal proliferation, affecting 2% of the population. Common physical complaints of this chronic skin disease are skin soreness, burning sensations, itching, and joint pain. Therapies include topical treatments (for example, topical corticosteroids), photo(chemo)therapies (for example, UV-B and PUVA), and the conventional systemic treatments (for example, methotrexate and cyclosporin). For patients who fail or who are contraindicated or intolerant to these conventional treatments, biologic response modifiers, or biologics, are available, such as adalimumab, etanercept, infliximab, and ustekinumab. However, since dermatological treatment can only offer a temporary suppression or remission of symptoms, treatment efforts may increasingly be directed towards both a decrease of disease severity and an increase of a patient’s HRQoL. Previous research, including studies in oncology, suggested that the application of a HRQoL intervention is considered to be an aid for monitoring disease and treatment over time, for the communication with the patient, and for improving treatment outcome. However, in dermatology practice, the application of a HRQoL intervention is relatively new. Moreover, evidence on the effectiveness of a HRQoL intervention in dermatology is still missing. The present study may rectify this deficiency. It is hypothesized that a HRQoL intervention will significantly improve patients’ HRQoL, and will facilitate doctor-patient communication. In addition, this study offers the opportunity to examine the course of HRQoL during treatment, the mid- and long-term effects of treatment on HRQoL, and the relationship between disease severity and HRQoL during treatment.

**METHODS/DESIGN**

**Design of the study**

This prospective multicenter randomized controlled trial (RCT) will include a 48-week follow-up period with six consecutive, predefined visits: V1 (week 0; baseline visit), V2 (week 6; 6 weeks after the first injection of etanercept), V3 (week 12; 12 weeks after the first injection), and continuing on in this manner for V4 (week 24), V5 (week 36), and V6 (week 48). For a comprehensive study flow chart, see Additional file 1 (for the intervention group) and Additional file 2 (for the control group). Via ALEA, a software package to support online patient registration and randomization in healthcare research, the participating study sites
will be cluster randomized (www.tenalea.com). Thus, it is not the patients, but the study sites that will be randomly allocated to either the intervention group (treatment with etanercept and standardized HRQoL assessment and communication) or the control group (treatment with etanercept alone). With this computer-based system, allocation of concealment will be assured. Random assignment of patients instead of study sites was considered, but rejected due to practical limitations, that may cause a contamination effect.

In order to create equal comparison groups, the randomization will be stratified by four clusters: academic centers versus non-academic centers, and centers naive to HRQoL assessment and communication versus centers not (completely) naive (Figure 1a). Study sites that have participated in a national Working Group Quality of Life Assessment in Dermatological Practice (2007 to 2008) will be considered not naive, since these sites do have a high interest in, and are involved in HRQoL assessment at their local clinical settings. This will result in a randomization scheme consisting of three blocks (Figure 1b). Herewith, the sizes of the intervention and control group will be similar, taking an equal distribution of type of study sites into account.

Since the intervention concerns both patients (for example, completion of the intervention questionnaire) and dermatologists (for example, providing feedback on the HRQoL scores of the intervention questionnaire), dermatologists in the intervention group will be educated and trained in standardized HRQoL assessment and communication prior to study start, during the ‘Intervention Training’ session on group level. The aim of this training is to enable dermatologists in the intervention group to adequately discuss HRQoL scores, as well as aspects of HRQoL, such as coping behavior and disease management. As a result, dermatologists will become more familiar and, thereby, comfortable with HRQoL assessment and communication in clinical practice. To this aim, the Intervention Training will include evidence-based background information on the construct HRQoL, its relevance to clinical practice, and the impact of psoriasis on HRQoL.5-10 Dermatologists will be trained in the electronic assessment of the intervention questionnaire, the Skindex-29, and the interpretation of its scores. In addition, information on coping behavior and disease management will be provided.

In this study design neither the patients nor the dermatologists can be blinded to the intervention.

Setting

The study will be conducted at 18 outpatient dermatology clinics in the Netherlands. Psoriasis patients will be consecutively invited to participate in the study after a medical decision is made to start biologic treatment with etanercept. Patients eligible for the study are diagnosed with a moderate to severe psoriasis (that is, a Psoriasis Area and Severity Index (PASI) of ≥8), and are 18 years or older at the time of informed consent. Patients who are mentally and/or physically unable to complete the study questionnaire(s), who speak the Dutch language insufficiently to fully understand and complete the intervention and/or study questionnaire(s), or who are not willing or not able to discuss HRQoL issues, will be excluded from participation. A patient information form will be handed to all potential participating patients. Written informed consent needs to be obtained from all participating patients prior to study start.
This study is registered in the Dutch Trial Register (NTR1364). The study protocol has been reviewed and approved by the Competent Authority of The Netherlands (Centrale Commissie Mensgebonden Onderzoek - CCMO) (NL24494.018.08 BI), and reviewed by the Medical Ethics Committee of the Academic Medical Center, University of Amsterdam (EC AMC) (MEC 08/302). The EC AMC exempted this study for ethical approval. A written confirmation of this statement was given by the EC AMC (Ref: MEC 08/302# 08.17.1716; MEC 08/302# 08.17.1933). The study will be conducted in accordance with the applicable laws and regulations following the Declaration of Helsinki protocols.
Participants
Psoriasis patients who will receive etanercept treatment by prescription from their dermatologists will be asked to participate in the study.\textsuperscript{11-14} Thus, all included patients of both the intervention and control group will receive etanercept, as a constant factor, in accordance with routine clinical practice and, ideally, following the Dutch treatment guidelines for psoriasis: a PASI ≥8; ineffective or contra-indications to PUVA treatment twice weekly for 10 weeks; ineffective or contra-indications to treatment with cyclosporine 3-5 mg/kg/day for 16 weeks; and/or ineffective or contra-indications to treatment with methotrexate 22.5 mg/day for 16 weeks.\textsuperscript{15} The per-protocol dose regimen of etanercept in week 0 to 12 (V1 to V3) is in accordance with the current Dutch summary of product characteristics (SmPC), namely: 2 x 50 mg/week. Depending on the PASI as measured at week 12 (V3), patients either discontinue treatment in case of PASI <50, or continue treatment during week 12 to 24 (V3 to V4). This treatment phase is called the induction phase. If PASI <75, patients will continue treatment with etanercept with 2 x 50 mg/week; if PASI ≥75, patients will continue with 2 x 25 or 1 x 50 mg/week. However, current clinical practice showed that a growing number of Dutch dermatologists continue treatment directly after the maximum treatment period of 24 weeks. Supporting evidence was found for an entire dosage regimen (week 0 to 48) in the British Association of Dermatologists Guidelines, in the German guidelines for systemic therapy, and in a study on the long-term safety and efficacy of 2 x 50 mg/week.\textsuperscript{16-19} Per study protocol, patients will continue treatment during week 24 to 48 (V4 to V6), with 2 x 25 or 1 x 50 mg/week, which is called the maintenance phase (see Figure 2). The dosing schedule as suggested per protocol will be used as a guide for dermatologists. However, the actual dosing schedule to be given to patients will depend on routine clinical practice and the dermatologists’ prescriptions. Previous treatment with any biologic is allowed, taking the applicable wash-out period into account prior to treatment with etanercept. Concomitant topical treatment will be permitted.

Intervention
The HRQoL intervention in this study consists of 1) the electronic assessment of the Skindex-29, a well-studied dermatology-specific HRQoL questionnaire, and 2) the communication of the resulting Skindex-29 data with the patient. The Skindex-29 consists of 29 questions, or items (for example items on pain, itch, embarrassment, frustration, social life, and interaction with others). Questions concern patients’ perception of the impact of the skin disease on aspects of HRQoL during the past week. These 29 items form three domains: symptoms, emotions, and functioning. Responses are given on a five-point response scale (ranging from ‘never’ to ‘all the time’), where scores are transformed to a 100-point scale, and domain and overall scores are being calculated by averaging responses to items in a given domain. Higher scores are indicating lower levels of HRQoL.\textsuperscript{20,21} Several reviews suggested that the Skindex-29 is considered to be the instrument of choice in dermatology.\textsuperscript{9,22} Before study start, dermatologists in the intervention group will be supported in the installation of the electronic version of the Skindex-29 on a desktop pc at the study sites, including a connection with a printer device. Prior to each consultation, patients in the intervention group will be asked to complete the Skindex-29 on a desk-top pc at the clinic. Patients will receive instructions from a dermatologist, or a designated representative, on how to complete this
Figure 2. Per-protocol dose regimen schedule for etanercept in the treatment of moderate to severe psoriasis.

electronic questionnaire, without providing any information on the content or answers of the items. After completion, both the patient and the dermatologist will receive a print-out of the questionnaire, including the answers given and a matrix visually displaying the domain and overall scores. Guided by this print-out, the dermatologist will be able to systematically discuss the Skindex-29 answers and scores with the patient. In addition, and during consecutive
consultations, the patient and the dermatologist will also be able to follow the HRQoL-scores over time as being displayed by this matrix showing repeated measurements. The HRQoL intervention will be supported by patient information on quality-of-life, coping behavior, and disease management. At the first consultation, dermatologists will be instructed to provide two brochures to the patient; one with general information on HRQoL, and one on coping with and the management of psoriasis. At the second consultation, patients will receive another brochure about the treatment of psoriasis, including chapter on HRQoL. In addition, a quality-of-life issue of a patient magazine from the Dutch Psoriasis Foundation will be provided to patients as well. At the third consultation, patients will receive a DVD on coping with psoriasis. At the fourth consultation, patients will receive a book entitled: ‘Psoriasis - Image, Experience, Treatment’, and patients will be instructed to read chapter 15: ‘The experience of psoriasis’. And, at the fifth consultation, patients will be asked to read chapter 16: ‘Coping with psoriasis’. Patients will be instructed to read and to watch the applicable patient information at home and prior to their subsequent consultation. The patient information can be stored in a study specific binder called ‘My Quality of Life’. Additionally, all patients, both in the intervention and control group, will be asked to complete a diary to record the etanercept injections. These diaries will be used for monitoring the drug compliance.

Comparison
Patients in the control group will not receive the HRQoL intervention, but will receive etanercept treatment only. The consultation will be conducted according to routine clinical practice, or usual care. Dermatologists in the control group will be asked to stay naive to HRQoL assessment and communication during the course of this study; that is, they are not to implement electronic HRQoL assessment with and communication about the Skindex-29 during their consultations with study participants.

Outcome measures
The primary outcome parameters in this study are 1) the impact of the HRQoL intervention on patients’ HRQoL and 2) the effect of the HRQoL intervention on doctor-patient communication. Since the Skindex-29 is used as the intervention questionnaire, the impact on HRQoL will be measured with another dermatology-specific HRQoL questionnaire, namely the Dermatology Life Quality Index (DLQI). The DLQI consists of ten items and each item is scored on a four-point scale ranging from ‘not at all/not relevant’ to ‘very much’. Scores of individual items (0 to 3) are added to yield a total score (0 to 30). Higher scores mean greater impairment of a patient’s HRQoL. A study-specific, two-dimensional communication questionnaire will be used to measure the effect on doctor-patient communication. This communication questionnaire consists of two parts, measuring: 1) the satisfaction with doctor-patient communication and 2) the quantity of HRQoL communication during consultations. With approval from the authors, the first part of the communication questionnaire is a slightly adapted version of the Patient Satisfaction Questionnaire and includes items on satisfaction with information, (emotional) support, communication, and fulfillment of needs. Answers are given on a visual analogue scale (VAS) ranging from
0 to 10, for example, ‘not at all satisfied’ to ‘very much satisfied’, or ‘not at all important’ to ‘very much important’. The cut-off point for satisfaction will be 6.0 (60%). The second part of the communication questionnaire includes items on symptoms, mood or emotions, social functioning, and overall quality-of-life. Patients will be asked if these aspects were discussed during the patient consultation or not, by ticking ‘yes’, ‘no’, or ‘I don’t know’. The communication questionnaire was successfully pilot tested in seven in-patients with psoriasis at the Academic Medical Center, Amsterdam.

The secondary efficacy endpoints include health status and disease severity. Health status will be measured with the Medical Outcomes Study 36-item Short-Form General Health Survey (SF-36), a well-established, eight-dimensional, generic HRQoL instrument.\textsuperscript{26,27} The items represent issues relevant to health status, and relate to the past four weeks. The SF-36 will be completed at week 0 (V1), week 24 (V4), and week 48 (V6) by patients in both the intervention and control group. Disease severity will be measured with the PASI, the most widely used instrument for the measurement of the severity of psoriasis,\textsuperscript{28} and with an investigator global assessment (IGA) and a patient global assessment (PGA). The IGA is asking dermatologists: ‘In your opinion, how severe is your patient’s current skin condition?’, and the PGA is asking patients: ‘How severe is your current skin condition?’. Answers can be given on a five-point scale, ranging from ‘not severe’ to ‘very severe’. All patients will be asked to complete this set of questionnaires at the end of each patient consultation.

In addition, two evaluation questionnaires will be introduced to measure 1) aspects on satisfaction and 2) aspects on the feasibility of HRQoL assessment and communication in dermatology practice. First, at week 24 (V4) and week 48 (V6), an overall evaluation questionnaire will be handed to dermatologists in the intervention group, and to patients in both the intervention and the control group. This questionnaire comprises a study-specific questionnaire about 1) satisfaction with the treatment process, including doctor-patient communication, and 2) satisfaction with treatment outcomes. For example, ‘How satisfied are you about the conversations you have had with your doctor/patient?’. Answers are given on a VAS, ranging from ‘not at all satisfied’ to ‘very much satisfied’. Second, after each patient consultation, dermatologists in the intervention group will be asked to answer questions on aspects of feasibility, such as completion time, duration of the consultations, and (the number of) aspects of HRQoL discussed during the consultations. Dermatologist in the control group will only be asked about the duration of the consultations.

Data collection
A Case Report Form (CRF) will be provided for each patient. All protocol-required information that needs to be collected during this study will be entered by the dermatologist, or a designated representative. Baseline characteristics will be obtained at visit 1 (week 0).

Safety
All patients will be monitored for safety during their participation in the study, including adverse events (AEs) and premature discontinuation from the study. AEs are defined as any undesirable experience occurring to a patient during the study, whether or not considered related to etanercept. Serious adverse events (SAEs) will be recorded by the responsible
dermatologist and reported within 24 hours of notification to the trade holder of etanercept. According to the Dutch Personal Data Protection Act, all data will be handled confidentially and anonymously, and will be stored for 15 years after study close out (that is, last patient last visit).

Analysis

Power calculation

The sample size calculation is based on the two primary outcome parameters and its corresponding outcomes measures, namely the DLQI and the communication questionnaire. Based on previous research, we will expect a HRQoL improvement, as measured with the DLQI, of 1.0 standard deviation (SD) in the intervention group and of 0.7 in the control group. We expect a communication effect size of 0.8 (that is, an improvement of 0.8 SD) in the intervention group, and of 0.4 in the control group. With 100 patients in each group, the power to detect a longitudinal effect would be larger than 99%, assuming a significance level of 5% and an autocorrelation (that is, the correlation between scores at baseline and end of induction phase) of 0.5.

The power to detect differences in longitudinal DLQI effects between the intervention and control group will be 56%. In case of one-tailed testing, the power to detect differences in longitudinal DLQI effects between the intervention and control group will be 68%. A mean DLQI score of 12.0 (SD 7.0), and a clinically meaningful improvement of 5 points is expected. In the intervention group, an additional improvement of 2 points effect size: 0.3) is expected. The power to detect differences in longitudinal communication effects between the intervention and control group is 80%. The power calculation is conservatively based on an estimate of 200 patients; that is, it is based on 100 patients per group and an expected withdrawal rate of approximately 20%.

Statistical analysis

Statistical analyses will comprise two components: 1) the induction phase analysis (week 0 to week 24) and 2) the maintenance phase analysis (week 24 to 48).

The induction phase analysis will concern all patients who receive any treatment with etanercept, who respond to the study questionnaires on HRQoL, communication, health status, and disease severity, and of whom PASI is assessed at baseline. The last visit of patients who may discontinue from treatment for any reason, including patients who will not meet the criteria to continue treatment with etanercept after week 12 (PASI <50), will be considered as End of Induction Phase (week 24).

The maintenance phase analysis will concern all patients who receive any treatment with etanercept after week 24, who respond to the questionnaires, and of whom PASI is assessed at week 24. The last visit of patients, who may discontinue from treatment for any reason, will be considered as End of Maintenance Phase as well as End of Treatment (week 48).

All analyses will be performed on an intention-to-treat basis. In any case where patients prematurely discontinue from treatment, they will be asked to complete an Early Termination visit. The content of this visit is similar to V6, and data will be handled based on the ‘last value carried forward’ principle.

Patient characteristics will be described by randomized groups. Categorical data will be summarized as frequencies and percentages. Continuous data will be summarized by
means and standard deviations. Linear mixed model analysis will be used to investigate the course of the outcomes (that is, DLQI scores for HRQoL; communication scores for doctor-patient communication; SF-36 scores for health status; and PASI, IGA, and PGA for disease severity), and to test the effects of possible explanatory variables on the outcome variables. Longitudinal models will have a random intercept to account for individual differences at baseline, fixed regression coefficients for each of the measurement occasions following baseline, for the intervention effect, for study sites (to account for differences between sites), and for possible other explanatory variables (such as age and sex). We will investigate which longitudinal structure is most appropriate, whether one or more of the fixed regression coefficients should be considered random, and whether there are interaction effects.

**DISCUSSION**

This study investigates the impact of HRQoL assessment and communication on patients’ HRQoL, and its effect on doctor-patient communication. To the best of our knowledge, this is the first RCT that examines the effectiveness of such HRQoL intervention in patients with moderate to severe psoriasis. Several studies in, for example, oncology showed a positive impact on patients’ well-being as well as a significant increase in the discussion of symptoms. Valderas and colleagues (2008) conducted a systematic review on the impact of PRO assessment in clinical practice in different clinical settings, such as internal medicine, oncology, and primary care. They concluded that, because the studies analyzed were heterogeneous in the types of setting, patients, intensity of intervention, and diversity of outcomes, no apparent conclusion could be drawn. In addition, the included studies were of limited methodological quality. Despite this, they found grounds for optimism, and they recommended well-designed and well-conducted future randomized studies.

A major strength of this study is that clinical staff, rather than research staff, is responsible for the implementation of the intervention. Herewith, the HRQoL intervention is implemented in a regular clinical setting. In addition, the HRQoL intervention will be evaluated in multiple clinical settings with diverse doctor and patient samples.

A few limitations of this study need to be addressed. First, our patient sample cannot be considered truly representative for psoriasis patients in general since only psoriasis patients who are being treated with etanercept will be asked to participate in the study. This may affect the external validity. Treatment with other biologics was considered, but rejected due to methodological limitations, such as route of administration, dosing schedule, and corresponding consultations to the dermatologists, that may affect the outcome. Second, the improvement of patients’ HRQoL over time may be due not only to the intervention, but also to their psoriasis treatment with a biologic. However, an additional improvement on top in HRQoL is expected because of the HRQoL intervention. Third, neither the dermatologists nor the patients can be blinded in this study design, which may affect the internal validity. And lastly, we used block randomization to generate equal comparison groups with respect to academic versus non-academic centers and centers being naive versus not-naive to HRQoL assessment. However, since only two academic centers agreed
to participate, with one center being naive and the other not naive to HRQoL assessment, the intervention and control conditions in the academic centers are not comparable with respect to this latter variable. The findings to be reported in this study are the first in dermatology. Therefore, some caution is needed when interpreting the results of this study, and confirmation through future study results is necessary. Nevertheless, the results of this study may be encouraging for further research and future use in dermatology practice.

ACKNOWLEDGEMENTS

This Investigated Initiated Research is financially supported by Pfizer bv, Rotterdam, the Netherlands, with an unrestricted research grant.

REFERENCES


Additional file 1. Study flow chart intervention group.

<table>
<thead>
<tr>
<th>VISIT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Wk 0</th>
<th>Wk 6</th>
<th>Wk 12</th>
<th>Wk 24</th>
<th>Wk 36</th>
<th>Wk 48</th>
<th>EV²</th>
<th>ET³</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEEK</td>
<td>Wk 0 Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In- and exclusion criteria</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informed Consent</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical history</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical status</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Medication</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adverse Events (AEs)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SkinEx-29</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Communication Questionnaire</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DLQI</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SF-36</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PASI</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Global Disease Severity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Overall Evaluation</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Experience with the intervention questionnaire</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Checklist</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

¹ End Of Study visit.
² Extra visit, if applicable.
³ Early Termination visit, if applicable.
## Additional file 2. Study flow chart control group.

<table>
<thead>
<tr>
<th>VISIT →</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEEK →</td>
<td>Wk 0</td>
<td>Wk 6</td>
<td>Wk 12</td>
<td>Wk 24</td>
<td>Wk 36</td>
<td>Wk 48 EOS¹</td>
</tr>
<tr>
<td>STUDY PART</td>
<td>Baseline</td>
<td>EV²</td>
<td>ET³</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STUDY PART</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In- en exclusion criteria</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informed Consent</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical history</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical status</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Medication</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adverse Events (AEs)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Communication Questionnaire</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DLQI</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SF-36</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASI</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Global Disease Severity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Overall Evaluation</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Checklist</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

¹ End Of Study visit.
² Extra visit, if applicable.
³ Early Termination visit, if applicable.
The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: results of a multicenter randomized controlled trial

CAC Prinsen
Phl Spuls
R Lindeboom
MAG Sprangers
MA de Rie
J de Korte

Submitted
ABSTRACT

Background: The routine use of health-related quality of life (HRQoL) data in dermatological practice is relatively new and evidence of the effectiveness of a HRQoL intervention in dermatology is missing.

Objective: The primary objective was to investigate the efficacy of a HRQoL intervention on patients’ HRQoL and doctor-patient communication. Secondary outcomes included health status and disease severity.

Methods: In a prospective, multicenter, cluster-randomized controlled trial, outpatients with moderate to severe psoriasis received either a biologic with the HRQoL intervention or biologic treatment alone, and were followed for 48 weeks. The intervention included the electronic assessment of the Skindex-29 and communication of the resulting data. Multilevel analysis was used to assess the course of outcomes over time.

Results: Eighteen study sites were randomized by a computer-based randomization system to assure allocation concealment. Of the 80 included patients, 29 received the HRQoL intervention. The HRQoL intervention showed no effect on the primary and secondary outcomes (effect size = 0.001). Nevertheless, the discussion of HRQoL aspects was enhanced, although at the expense of longer consultation times (31.6 vs. 18.2 minutes, p < 0.001), and a positive impact on patients’ and physicians’ satisfaction with the process of care was observed (p < 0.05).

Conclusions: No evidence was found of the efficacy of a HRQoL intervention in dermatological practice on the primary and secondary study outcomes. Nevertheless, the intervention appeared to have an, as yet unquantified, effect on process variables. Herewith, the HRQoL intervention may serve as a tool for improving patient care.

Trial registration: The Netherlands National Trial Register: NTR1364.
INTRODUCTION

Background and objectives

The impact of patient-reported outcome (PRO) measurement in clinical practice is under debate. Valderas and colleagues conducted a systematic review on the impact of PRO measurement in clinical practice in different clinical settings. The studies analyzed were heterogeneous in the types of setting, participants, intensity of intervention, and outcomes. Nonetheless, evidence of the impact of PROs in clinical practice included the detection of physical or psychological problems, and monitoring of disease progression, the facilitation of doctor-patient communication, and shared decision making. The need for additional studies concerning the efficacy of the routine use of health-related quality of life (HRQoL) data in clinical practice was highlighted.

In dermatology, it is well known that psoriasis has a relatively high, negative impact on HRQoL, especially in patients with a moderate to severe form. This systemic, inflammatory skin disease is affecting approximately 2% of the population. Impact on physical functioning may include skin soreness, burning sensations, itching and joint pain. Impact on psychological functioning may include shame, depression and frustration. Impact on social functioning may include work of hobbies, interaction with others, and social life. Therapies comprise of topical treatments, photo(chemo)therapies, and conventional systemic treatments. For patients who fail or who are contraindicated or intolerant to these treatments, biologic treatments are available, such as adalimumab, etanercept, infliximab, and ustekinumab. Since dermatological treatment can only offer a temporary suppression or remission of symptoms, treatment efforts are increasingly directed towards both a decrease of disease severity and an increase of a patient’s HRQoL.

The routine use of a HRQoL intervention in dermatological practice is relatively new, and evidence on the effectiveness of such intervention in dermatology is missing. The aim of this study was to investigate the efficacy of a HRQoL intervention in clinical practice during 48-weeks of systemic treatment of moderate to severe psoriasis. It was hypothesized that the HRQoL intervention has a positive impact on patients’ HRQoL, doctor-patient communication, health status and disease severity. In addition, the process of care was evaluated, and the feasibility of the HRQoL intervention was assessed.

MATERIALS AND METHODS

As details on methods and design have been published previously, materials and methods in this section are restricted to a summary.

Trial design

In a prospective, multicenter cluster randomized controlled trial (The Netherlands Trial Register, http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1364, protocol number NTR1364), patients diagnosed with moderate to severe psoriasis received either etanercept in combination with the HRQoL intervention (see below), or etanercept alone, and were followed for 48 weeks.

Participants

The study was initiated at 18 outpatient dermatology clinics in the Netherlands. Eligible patients were diagnosed with moderate to severe psoriasis (Psoriasis Area and Severity Index (PASI) $\geq 8$).
and were 18 years or older at the time of informed consent. Patients who were mentally and/or physically unable to complete the study questionnaire(s), who were insufficiently able to speak the Dutch language, or who were not willing or not able to discuss HRQoL issues, were excluded from participation. Patients who received etanercept treatment by prescription were asked for their participation in the study. All included patients, both in the intervention and control group received etanercept treatment. Treatment from baseline until week 24 was called the Induction Phase; treatment during week 24 to 48 was called the Maintenance Phase. Written informed consent was obtained from all patients prior to study start.

**Intervention**

The HRQoL intervention consisted of the electronic measurement of HRQoL with the multi-dimensional Skindex-29 and communication of the resulting Skindex-29 data with the patient. Several reviews suggested to use the Skindex-29 in clinical practice, and information on the interpretation of its scores is known. The Skindex-29 was selected as the intervention questionnaire; the Dermatology Life Quality Index (DLQI) was chosen to measure the effect on the primary outcome.

Prior to each consultation, patients in the intervention group were asked to complete the electronic Skindex-29 on a desk-top pc at the study site. Guided by a print-out, including the answers given and a matrix visually displaying the domain and overall scores, the dermatologist was instructed to systematically discuss the resulting Skindex-29 data with the patient during the consultation.

**Comparison**

The Skindex-29 was not presented to patients in the control group and dermatologists were instructed not to implement the Skindex-29 during their consultations with study patients.

**Outcomes**

The primary outcomes in this study were the effect of the HRQoL intervention on patients’ HRQoL and doctor-patient communication. Secondary outcomes included health status and disease severity.

**HRQoL** was measured with the DLQI. Scores ranged from 0 to 30 points, with higher scores meaning a greater impairment of patients’ HRQoL.

**Doctor-patient communication** was measured with six items measuring satisfaction with doctor-patient communication, such as items on satisfaction with information, support, communication, and fulfilment of needs. Answers ranged from 0 to 10, with higher scores indicating higher levels of satisfaction. As internal consistency was high (r ranged from 0.70 to 0.93), total scores were averaged to obtain a 0 to 10 test result. The second part inquires on the quantity of HRQoL communication during consultations. It consisted of 14 items with a binary response scale (yes/no) asking about whether HRQoL was discussed and, if so, which aspects on HRQoL had been discussed.

**Health status** was measured with the Medical Outcomes Study 36-item Short-Form General Health Survey (SF-36), with a 0–100 norm-based scoring algorithm, where higher scores indicate a better health status.
Disease severity was measured with the PASI, a Patient Global Assessment (PGA) score, and an Investigator Global Assessment (IGA) score. PASI scores may range from 0 to 72 points and involve the total skin surface area. PGA and IGA scores were obtained on a 5-point response scale, ranging from 'not severe' to 'very severe'.

In addition, two study-specific evaluation questionnaires were introduced at week 24 and 48; one to evaluate the process of care by patients and dermatologists, and one to evaluate the feasibility of the HRQoL intervention among patients and dermatologists in the intervention group.

Measurements took place at baseline, at 6 weeks after the first injection with etanercept, and at weeks 12, 24 and 48 weeks. The SF-36 was completed at baseline, week 24 and 48; both evaluation questionnaires only at week 24 and 48. To separate the effects of the electronic mode of the intervention questionnaire from the outcome questionnaires as much as possible, the outcome questionnaires were completed by paper-and-pencil.

**Sample size**

Sample size calculation was based on conservative power estimates and the two primary outcomes. One hundred patients in each group were needed to detect a clinically significant change in HRQoL (1.0 standard deviation in the intervention group and 0.7 in the control group) and communication (an effect size of 0.8 in the intervention group, and of 0.4 in the control group), with 99% power and a significance level of 5%, allowing for 20% attrition. Post hoc sample size calculations showed that with 29 intervention and 51 control patients, the power to detect statistically significant differences on the primary endpoints was 47%.

**Randomization**

The study sites were cluster-randomly allocated to either the intervention or the control group by a computer-based randomization system (www.tenalea.nl). Herewith, allocation of concealment was assured. The randomization was stratified by academic vs. non-academic study sites, and study sites naive to HRQoL measurement vs. study sites not (completely) naive.

**Blinding**

Random assignment of patients instead of study sites was considered, but rejected due to the impossibility of blinding, and the risk of contamination.

**Statistical methods**

Statistical analysis comprised of two sets: i) the induction phase analysis (week 0-24), and ii) the maintenance phase analysis (week 24-48). Statistical analyses were performed on an ‘intention-to-treat’ basis. In case of premature discontinuation from the study, data analysis (i.e., descriptive statistics) was also based on the ‘last observation carried forward’ principle. The last visit of patients who discontinued from the study was either considered as end of Induction Phase or end of Maintenance Phase. Missing data was addressed by the multilevel analysis, which assumes that missing data is not related to the intervention. Baseline data on patient characteristics were described by randomized groups, and examined for imbalances. Differences between groups on continuous variables were tested using Mann-Whitney U tests, and differences on categorical variables were tested using Fischer exact test. Multilevel linear models analysis with repeated measures was used to compare and test the course of each
study outcome and to correct for possible baseline differences using a 2-sided criterion of \( p < 0.05 \). All analyses were performed with SPSS version 19.0 for Windows (SPSS Inc, Chicago, IL).

**Ethical considerations**

This study has been registered in The Netherlands National Trial Register (NTR): NTR1364. The study protocol has been reviewed and approved by the Competent Authority of The Netherlands (Centrale Commissie Mensgebonden Onderzoek – CCMO) (NL24494.018.08 BI), and by the Medical Ethics Committee of the Academic Medical Center, University of Amsterdam (EC AMC) (MEC 08/302). The EC AMC exempted this study for ethical approval. A written confirmation of this statement was given by the EC AMC (Ref: MEC 08/302# 08.17.1716; MEC 08/302# 08.17.1933). The study has been conducted in accordance with the applicable laws and regulations following the Declaration of Helsinki protocols, and was reported according to the “CONSORT 2010 checklist of information to include when reporting a randomized trial”.

**RESULTS**

**Participant flow**

Of the 60 out-patient dermatologic clinics in The Netherlands that were initially invited to participate in the study, 18 study sites agreed. The main reasons for non-participation included workload, no experience with biologic treatment, and non-response to the invitation. The 18 study sites were cluster randomized. Patients were screened for eligibility by 15 study sites: six intervention and nine control sites. Three study sites, randomized to the intervention group, did not screen any patients due to high workload at the clinics. This may explain the lower number of patients in the intervention group as compared to the control group.

**Recruitment**

A flow diagram is displayed in Figure 1. Patient recruitment took place between March 2009 and July 2012. A total of 116 patients with moderate to severe psoriasis were consecutively screened for eligibility. Thirty-six patients (31.0%) were not eligible or refused participation, leaving 80 patients (69.0%) aged between 18.7 to 77.0 years (mean age 49.3 years, sd 13.1) of whom 29 received the HRQoL intervention.

A 2x50 mg/week etanercept treatment regimen was initiated for all included patients (for details on the per-protocol dose regimen we refer to the protocol article). Evidence on long term safety of etanercept supported our rationale for etanercept treatment in the study.

The study was terminated early on 31 July 2012 due to major difficulties in patient recruitment.

**Baseline data**

Baseline data (Table 1) showed a good balance between the intervention and control group for all demographic characteristics, except for the non-dermatological comorbidities: patients in the control group more often had comorbidities (43.1%) than patients in the intervention group (17.2%) \( (p = 0.02) \).

Patients in the intervention group had more favourable baseline DLQI scores (8.0 vs. 15.0 points; a moderate degree of effect on patients’ life vs. a very large degree of effect, \( p < 0.001 \)), SF-36 Mental Component Summary (MCS) scores (49.6 vs. 44.8 points, \( p = 0.03 \)),...
and PGA scores ($p=0.02$). As to the quantity of HRQoL communication, patients in the intervention group reported that (aspects of) HRQoL were more frequently discussed (79.3%) compared to patients in the control group (49.0%).

**Numbers analysed**

The induction phase analysis (week 0-24) included 29 patients in the intervention group, and 51 patients in the control group. The maintenance phase analysis (week 24-48) included 27 and 42 patients respectively; two patients in the intervention group and nine in the control group were lost to follow-up (Figure 1).

**Outcomes and estimation**

Tables 2a and 2b show the results of the multilevel analysis with week 24 and week 48 as follow-up time respectively. Corrected for the baseline difference in DLQI scores between intervention and control group, the HRQoL intervention had no significant effect on patients’ HRQoL. At week 24, the average effect on HRQoL was 0.17 DLQI points (95% CI: -1.93-2.26) ($p=0.88$) in favor of the intervention group. At week 48 the effect was 0.13 points (95% CI: -1.94-2.20) ($p=0.90$). The same goes for the other outcomes: there were no significant differences between intervention and control group over time in doctor-patient communication scores, health status scores and disease severity scores ($p>0.05$).

Figures 2a to 2d illustrate the course of the primary and secondary outcomes over time. All outcomes improved in both the intervention and control group. To indicate, at week 24, 69.0% of the patients in the intervention group improved with >75% in PASI score. Patients in the control group improved in 51.0% of the cases. In addition, clearance or almost clearance (as measured with the PGA) was reported by 72.4% of the patients in the intervention group and by 62.7% of the patients in the control group.

Table 3 shows the outcomes between the intervention and control group at week 24 and week 48. Of the significant differences between both groups in primary and secondary outcomes at baseline, DLQI scores were still significant at week 24 (1.0 vs. 3.0 points, $p=0.02$), after correcting for baseline differences.

At week 24, the median doctor-patient communication scores, measuring satisfaction with doctor-patient communication, were 8.7 points in the intervention group vs. 7.6 points in the control group ($p=0.003$). At week 48, comparable results were found (8.9 vs. 7.5 points, $p<0.001$). With regard to the quantity of HRQoL communication, at week 24, 82.8% of the patients in the intervention group reported that (aspects on) HRQoL were discussed during the consultations, compared to 39.2% in the control group. In addition, the number of HRQoL aspects discussed was higher: ‘itch’ ($p=0.02$), ‘pain’ ($p=0.01$), ‘fatigue’ ($p=0.004$), and ‘problems with sleep’ ($p=0.01$). At week 48, 82.8% of the patients in the intervention group reported that (aspects on) HRQoL were discussed, compared to 27.5% in the control group, however, the mean number of aspects on HRQoL discussed were comparable ($p>0.05$).

**Patients’ evaluation of the process of care**

A complete overview of the patients’ evaluation of the process of care is displayed in Table 4. Patients in the intervention group were significantly more satisfied with their conversations
## Table 1. Demographic and disease characteristics of the study population.

<table>
<thead>
<tr>
<th></th>
<th>Intervention group</th>
<th>Control group</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study sites, no. (%)</strong></td>
<td>9 (50.0)</td>
<td>9 (50.0)</td>
<td></td>
</tr>
<tr>
<td>Academic study sites</td>
<td>1 (11.1)</td>
<td>1 (11.1)</td>
<td></td>
</tr>
<tr>
<td>Non-academic study sites</td>
<td>8 (88.9)</td>
<td>8 (88.9)</td>
<td></td>
</tr>
<tr>
<td>Study sites naive to HRQoL</td>
<td>7 (77.8)</td>
<td>6 (66.7)</td>
<td></td>
</tr>
<tr>
<td>Study sites not naive to HRQoL</td>
<td>2 (22.2)</td>
<td>3 (33.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Male, no. (%)</strong></td>
<td>20 (69.0)</td>
<td>32 (62.7)</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Mean age in years (SD, range)</strong></td>
<td>49.2 (11.7, 27.7-72.5)</td>
<td>49.6 (14.0, 18.7-77.0)</td>
<td>0.90</td>
</tr>
<tr>
<td><strong>Diagnosis psoriasis vulgaris, no. (%)</strong></td>
<td>29 (100)</td>
<td>51 (100)</td>
<td></td>
</tr>
<tr>
<td><strong>Mean disease duration in years (SD, range)</strong></td>
<td>18.5 (11.7, 4.0-42.0)</td>
<td>18.4 (11.9, 1.0-56.6)</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Marital status, no. (%)</strong></td>
<td>20 (69.0)</td>
<td>35 (68.6)</td>
<td>0.17</td>
</tr>
<tr>
<td>Married or partnership</td>
<td>3 (10.3)</td>
<td>10 (19.6)</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>5 (17.2)</td>
<td>5 (9.8)</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (3.4)</td>
<td>1 (2.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Education, no. (%)</strong></td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (&lt;high school)</td>
<td>15 (51.7)</td>
<td>24 (47.1)</td>
<td></td>
</tr>
<tr>
<td>Middle (high school / college)</td>
<td>9 (31.0)</td>
<td>21 (41.2)</td>
<td></td>
</tr>
<tr>
<td>Advanced (university)</td>
<td>5 (17.2)</td>
<td>6 (11.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity, no. (%)</strong></td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>27 (93.1)</td>
<td>45 (88.2)</td>
<td></td>
</tr>
<tr>
<td>Other1</td>
<td>2 (6.9)</td>
<td>6 (11.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Dermatological comorbidities, no. (%)</strong></td>
<td>9 (31.0)</td>
<td>16 (31.4)</td>
<td>0.98</td>
</tr>
<tr>
<td>Non-dermatological comorbidities, no. (%)</td>
<td>5 (17.2)</td>
<td>22 (43.1)</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Health-related quality of life (DLQI), median (IQR)</strong></td>
<td>8.0 (4.5-13.5)</td>
<td>15.0 (11.0-21.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Doctor-patient communication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with communication, median (IQR)</td>
<td>8.2 (7.3-9.0)</td>
<td>7.7 (6.6-8.5)</td>
<td>0.06</td>
</tr>
<tr>
<td>Quantity of HRQoL communication (%)</td>
<td>79.3</td>
<td>49.0</td>
<td></td>
</tr>
<tr>
<td><strong>Health status (SF-36), median (IQR)</strong></td>
<td>49.4 (39.9-54.6)</td>
<td>45.8 (40.0-51.3)</td>
<td>0.12</td>
</tr>
<tr>
<td>Physical Component Summary</td>
<td>49.6 (43.3-54.3)</td>
<td>44.8 (30.6-53.4)</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Disease severity, median (IQR)</strong></td>
<td>15.2 (11.9-19.9)</td>
<td>13.4 (9.8-21.8)</td>
<td>0.27</td>
</tr>
<tr>
<td>Patient Global Assessment</td>
<td>4.00 (3.0-4.0)</td>
<td>4.00 (4.0-5.0)</td>
<td>0.02</td>
</tr>
<tr>
<td>Investigator Global Assessment</td>
<td>4.00 (3.0-4.0)</td>
<td>4.00 (3.0-4.0)</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Abbreviations: DLQI, Dermatology Life Quality Index; HRQoL, health-related quality of life; IQR, inter quartile range; NA, not applicable; PASI, Psoriasis Area and Severity Index; SD, standard deviation; SF-36, Medical Outcomes Study 36-item Short-Form General Health Survey.

1 Intervention group: Surinamese / Hindu (1) and Turkish (1); Control group: Chinese (1), Egyptian (1), Surinamese / Hindu (3) and Turkish (1).
2 Intervention group: alopecia totalis (1), lentigo maligna (1), psoriasis capitus (1), psoriasis unguium (1), psoriatic arthritis (9); Control group: acne vulgaris (1), arcosvesiculeus eczema (1), atopic dermatitis (1), atopic syndrome (1), dermatomyositis pedum (1), psoriasis palmpilantaris (1), psoriatic arthritis (11), rosacea (1).
3 A patient can suffer from more than one dermatological condition.
4 Data not shown but available on request.
5 Data are presented as the percentage of patients who indicated that (aspects of) HRQoL were discussed during the consultation.
with their dermatologists (item 1) at week 24 ($p=0.02$) and at week 48 ($p=0.001$). They also indicated feeling more supported by their dermatologists to improve their HRQoL (item 5) ($p=0.002$ and $p=0.004$ respectively), and they had significantly higher scores on “Attention given to HRQoL during the consultations” (item 3) ($p<0.001$).

In addition, patients were equally satisfied with their psoriasis treatment at both time points (item 7) ($p=0.08$ and $p=0.07$ respectively), and were equally likely to recommend etanercept
Table 2a. Outcomes of the multilevel analysis at the end of the induction phase (week 24).

<table>
<thead>
<tr>
<th>Variables in order entered into model</th>
<th>Estimate of effects</th>
<th>Standard error</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health-related quality of life (DLQI)</td>
<td>0.17</td>
<td>1.05</td>
<td>-1.93-2.26</td>
<td>0.88</td>
</tr>
<tr>
<td>Doctor-patient communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with communication</td>
<td>0.40</td>
<td>0.87</td>
<td>-1.55-2.34</td>
<td>0.66</td>
</tr>
<tr>
<td>Health status (SF-36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Component Summary</td>
<td>2.80</td>
<td>2.18</td>
<td>-2.62-8.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Mental Component Summary</td>
<td>0.72</td>
<td>1.02</td>
<td>-1.32-2.76</td>
<td>0.48</td>
</tr>
<tr>
<td>Disease severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASI</td>
<td>-0.10</td>
<td>1.24</td>
<td>-2.57-2.38</td>
<td>0.94</td>
</tr>
<tr>
<td>Patient Global Assessment</td>
<td>-0.23</td>
<td>0.20</td>
<td>-0.62-0.16</td>
<td>0.25</td>
</tr>
<tr>
<td>Investigator Global Assessment</td>
<td>-0.09</td>
<td>0.19</td>
<td>-0.54-0.36</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; DLQI, Dermatology Life Quality Index; PASI, Psoriasis Area and Severity Index; SF-36, Medical Outcomes Study 36-item Short-Form General Health Survey.

Table 2b. Outcomes of the multilevel analysis at the end of the maintenance phase (week 48).

<table>
<thead>
<tr>
<th>Variables in order entered into model</th>
<th>Estimate of effects</th>
<th>Standard error</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health-related quality of life (DLQI)</td>
<td>0.13</td>
<td>1.04</td>
<td>-1.94-2.20</td>
<td>0.90</td>
</tr>
<tr>
<td>Doctor-patient communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with communication</td>
<td>0.68</td>
<td>0.76</td>
<td>-1.03-2.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Health status (SF-36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Component Summary</td>
<td>2.88</td>
<td>2.02</td>
<td>-1.13-6.90</td>
<td>0.16</td>
</tr>
<tr>
<td>Mental Component Summary</td>
<td>0.92</td>
<td>1.46</td>
<td>-2.45-4.30</td>
<td>0.55</td>
</tr>
<tr>
<td>Disease severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASI</td>
<td>-0.30</td>
<td>1.20</td>
<td>-2.70-2.10</td>
<td>0.80</td>
</tr>
<tr>
<td>Patient Global Assessment</td>
<td>-0.21</td>
<td>0.19</td>
<td>-0.59-0.16</td>
<td>0.26</td>
</tr>
<tr>
<td>Investigator Global Assessment</td>
<td>-0.14</td>
<td>0.16</td>
<td>-0.46-0.18</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; DLQI, Dermatology Life Quality Index; PASI, Psoriasis Area and Severity Index; SF-36, Medical Outcomes Study 36-item Short-Form General Health Survey.

treatment to others (item 9) (p=0.16 and p=0.10 respectively). Patients in the intervention group, however, reported a higher burden of treatment compared to controls (item 8) (p=0.03 and p<0.001 respectively).

Dermatologists’ evaluation of the process of care
In general, dermatologists in the intervention group were satisfied with the process of care. They highly recognized the importance of the measurement of HRQoL in clinical practice: the median (inter quartile range, IQR) score at the end of week 24 was 8.6 (7.9-9.3), and 9.0 (7.8-9.5) at the end of week 48. They also indicated the importance of discussing HRQoL issues with their patients: 8.6 (7.5-9.3) and 8.8 (6.7-9.3) points respectively. In addition, they highly recommended the HRQoL intervention to other dermatologists: 9.2 (8.2-9.6) and 8.5 (7.0-9.7) points respectively.
Table 3. Primary and secondary outcomes at end of induction phase (week 24) and end of maintenance phase (week 48).

<table>
<thead>
<tr>
<th>Primary and secondary outcomes</th>
<th>End of induction (week 24)</th>
<th>End of maintenance (week 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention N=29</td>
<td>Control N=51</td>
</tr>
<tr>
<td>Health-related quality of life (DLQI)</td>
<td>1.0 (0-3.5)</td>
<td>3.0 (1.0-9.0)</td>
</tr>
<tr>
<td>Doctor-patient communication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with communication</td>
<td>8.7 (8.0-9.5)</td>
<td>7.6 (6.3-8.9)</td>
</tr>
<tr>
<td>Quantity of HRQoL communication (%)(^1)</td>
<td>82.8</td>
<td>39.2</td>
</tr>
<tr>
<td>Health status (SF-36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Component Summary</td>
<td>54.2 (46.7-57.3)</td>
<td>51.2 (39.6-57.3)</td>
</tr>
<tr>
<td>Mental Component Summary</td>
<td>53.5 (44.1-56.8)</td>
<td>46.9 (39.0-54.8)</td>
</tr>
<tr>
<td>Disease severity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASI</td>
<td>3.3 (1.3-4.9)</td>
<td>3.8 (2.2-6.9)</td>
</tr>
<tr>
<td>Patient Global Assessment</td>
<td>2.0 (1.0-2.8)</td>
<td>2.0 (1.0-3.0)</td>
</tr>
<tr>
<td>Investigator Global Assessment</td>
<td>2.0 (2.0-2.0)</td>
<td>2.0 (1.0-3.0)</td>
</tr>
</tbody>
</table>

Data are presented as median (IQR) and are based on the ‘last value carried forward’ principle.
Abbreviations: DLQI, Dermatology Life Quality Index; HRQoL, health-related quality of life; IQR, inter quartile range; PASI, Psoriasis Area and Severity Index; SF-36, Medical Outcomes Study 36-item Short-Form General Health Survey.
\(^1\) Data are presented as the percentage of patients who indicated that (aspects of) HRQoL were discussed during the consultation.
\(^*\) Different sample sizes are because of missing data.

Feasibility of the HRQoL intervention

Completion time of the electronic Skindex-29
At baseline, 51.7% of the patients completed the electronic Skindex-29 within 5-10 minutes. At the end of both the induction and maintenance phase, the Skindex-29 was completed within 1-5 minutes by most patients (57.7% and 59.3% respectively).

Patients indicated that they did not experience any major difficulties in completing the electronic Skindex-29 on a desktop pc (82.8%, 96.2%, and 96.3% at baseline week 24 and wk 48 respectively). Only a few difficulties were reported, such as printing issues (N=2), difficulties with working on a desktop pc (N=1), and difficulties working with a mouse (N=1).

Consultation time

The outpatient consultation time in the intervention group took significantly longer compared to the consultation time in the control group. Mean (sd) consultation times in the intervention group were 31.5’ (2.0) at baseline, 31.6’ (1.4) at week 24 and 30.3’ (4.2) at week 48. On average 5-10 minutes were used to discuss HRQoL issues with the patients as guided by the Skindex-29 data. The mean (sd) consultation times in the control group were 21.7’ (10.9) at baseline, 18.2’ (10.3) at week 24 and 15.4’ (9.8) at week 48 (p<0.001 for all comparisons with the intervention group).

DISCUSSION

This multicenter study, with repeated measures, investigated the effect of a HRQoL intervention on patients’ HRQoL, doctor-patient communication, health status and disease
Table 4. Patients’ evaluation of the process of care.

<table>
<thead>
<tr>
<th>End of induction (week 24)</th>
<th>End of maintenance (week 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
</tr>
<tr>
<td></td>
<td>N=27&quot;</td>
</tr>
<tr>
<td>1 Satisfaction about conversations with dermatologists (not at all satisfied – highly satisfied)</td>
<td>8.7 (7.8-9.5)</td>
</tr>
<tr>
<td>2 Importance of the assessment of HRQoL (not at all important – highly important)</td>
<td>8.3 (7.4-9.5)</td>
</tr>
<tr>
<td>3 Attention given to HRQoL during consultations (no attention at all – a lot of attention)</td>
<td>8.5 (7.5-9.5)</td>
</tr>
<tr>
<td>4 Importance of discussing HRQoL during consultations (not at all important – highly important)</td>
<td>8.5 (7.7-9.5)</td>
</tr>
<tr>
<td>5 Support given by dermatologists to improve HRQoL (no support at all – a lot of support)</td>
<td>8.5 (7.1-9.5)</td>
</tr>
<tr>
<td>6 Recommend HRQoL intervention to others (not recommended at all – highly recommended)</td>
<td>9.0 (8.2-9.7)</td>
</tr>
<tr>
<td>7 Satisfaction with psoriasis treatment (not at all satisfied – highly satisfied)</td>
<td>8.9 (7.7-9.6)</td>
</tr>
<tr>
<td>8 Burden of psoriasis treatment (highly burdensome – not burdensome at all)</td>
<td>8.6 (7.2-9.5)</td>
</tr>
<tr>
<td>9 Recommend treatment to others (not recommended at all – highly recommended)</td>
<td>9.2 (8.6-9.7)</td>
</tr>
</tbody>
</table>

Data are presented as median (IQR) and are based on the ‘last value carried forward’ principle. 
Abbreviations: HRQoL, health-related quality of life; IQR, inter quartile range; NA, not applicable.

* Different sample sizes are because of missing data.

# Patients in the control group did not complete the questions regarding the HRQoL intervention (items 2, 4, and 6).
severity in patients with moderate to severe psoriasis undergoing a long-term systemic treatment with a biologic. These patients form a relatively homogenous group, considered to be highly suitable to study the effects of HRQoL measurement in dermatological practice. To the best of our knowledge, this is the first RCT that examines the efficacy of a HRQoL intervention in routine dermatology practice.

The HRQoL intervention had no additional effect on patients’ HRQoL, doctor-patient communication, health status, and disease severity. These results are in contrast with a study performed by Velikova and colleagues, but consistent with other findings. All patients improved over time on all outcomes due to the treatment of psoriasis with a biologic. A significantly higher level of satisfaction with the process of care was found in the intervention group, which is in agreement with the results found by others.

The intended effect of the HRQoL intervention was reflected by the higher number of HRQoL aspects discussed during the consultations, which may provide valuable information that can be used in the patient management, although at the cost of lengthening the consultation time. The prolonged consultation time, however, differs from previous reports. Nonetheless, HRQoL aspects could be considered beneficial to care strategies, as individual patients’ needs could be early explored in direct contact with the patient. This may add to the efficiency of dermatological treatment and care.

Generalisability

The intervention appeared to be feasible, and the importance of the HRQoL intervention was recognised by all dermatologists in the intervention group, who highly recommended the HRQoL intervention to others. Although the sample of dermatologists was relatively small, there are no reasons to believe that these results would differ in a larger group of dermatologists.

Limitations

There are limitations that need to be acknowledged. First, the study was terminated early due to hampering patient recruitment because of novel biologic therapies, such as adalimumab and ustekinumab. Despite an extension of the inclusion period with 16 months and a number of intensive attempts to improve patient recruitment, recruitment was stopped at an inclusion rate of 80 patients. As a consequence, we had a lower power than we anticipated. However, the observed effect on the primary outcome (HRQoL) was very small (effect size=0.001): post hoc sample size calculation indicated that more than thousand patients would be needed to statistically detect this observed effect with 80% power. Secondly, patients in the intervention group had a significantly better HRQoL, mental health, patients’ perceived disease severity, and significantly less often a comorbidity than had the controls. These related baseline imbalances may be the result of the chance due to the cluster randomization instead of individual randomization. We therefore needed to correct for baseline differences. Strengths of this study were that the intervention was implemented in routine outpatient clinical practice, in multiple dermatology clinics with diverse doctor and patient samples, and clinical staff rather than research staff was responsible for the implementation of the HRQoL intervention.
CONCLUSION

We did not find evidence of the efficacy of a HRQoL intervention in dermatological practice on the primary and secondary study outcomes. Nevertheless, the intervention appeared to have an, as yet unquantified, effect on process variables: the discussion of HRQoL aspects was enhanced, and a positive effect on patients’ and physicians’ satisfaction with the process of care was found. Herewith, the HRQoL intervention may serve as a tool for the improvement of patient care, deserving further evaluation.

ACKNOWLEDGEMENTS

We thank all patients whose participation made this study possible. We thank all participating dermatologists: Dr MTW Gaastra; Dr AJ Oosting; Dr M van Steveninck-Wensing; Dr PM Ossenkoppele; Dr MB Crijns; Dr IJ Schornagel; Dr CW den Hengst; Dr ThW van den Akker; Dr CLM van Hees; Dr RLP Lijnen; Dr GAM Krekels; Dr JM Muche; Dr TJ Stoor; Dr B de Swaan. We also thank Ms A van Hasselaar who monitored the study data, Ms SF van Vugt for data entry, and the nursing and administrative staff who supported the run of this study.

REFERENCES


SUMMARY AND CONCLUSIONS
SAMENVATTING EN CONCLUSIES
Chapter 5

SUMMARY AND CONCLUSIONS

This thesis consists of three parts. The aim of part one, ‘Measurement of health-related quality of life in dermatology’, was to contribute to the quality of health-related quality of life (HRQoL) measurement in dermatological research and clinical practice. The aim of part two, ‘Interpretation of health-related quality of life scores’, was to facilitate the interpretation of Skindex-29 scores and score differences. The aim of part three, ‘Health-related quality of life application in clinical practice’, was to support the application of HRQoL measurement in dermatological practice and to investigate the efficacy of a HRQoL intervention on patients’ HRQoL, doctor-patient communication, health status, and disease severity. The overall aim of the thesis was to contribute to the improvement of HRQoL of patients with a chronic skin disease.
Chapter 1. General introduction

Chapter 1 is the general introduction and presents the background, aims and outline of the thesis. The concept HRQoL is illustrated, and its context in dermatology is described. Subsequently, a preface to different types of measurement instruments is provided and the Skindex-29 questionnaire is introduced. The importance of the psychometric quality of measurement instruments is illustrated, as well as the clinical meaningfulness of HRQoL data. Lastly, the added value of the application of HRQoL measurement and the use of HRQoL data in clinical practice are described.

Chapter 2. Measurement of health-related quality of life in dermatology

2.1 Measurement of health-related quality of life in dermatological research and practice: outcome of the EADV Taskforce on Quality of Life

Measurement of HRQoL is generally considered to be important in clinical trials, in the assessment of disease severity and in patient management. Therefore, a good understanding of the concept HRQoL and its measurement instruments is essential for both researchers and clinicians. The European Academy of Dermatology and Venereology (EADV) Taskforce on Quality of Life is encouraging the application of HRQoL measurement in dermatology. Our aim was to further contribute to the quality of HRQoL measurement in dermatological research and clinical practice. Chapter 2.1 presents a review article that is written on behalf of the EADV Taskforce and intended for clinicians and junior researchers who are relatively unfamiliar with the measurement of HRQoL. An introduction to the concept and methodology of HRQoL measurement is given, and the most commonly used generic and specific instruments in dermatology are summarized, such as the Medical Outcomes Study 36-item Short Form Health Survey (SF-36), Dermatology Life Quality Index (DLQI), Skindex-29 and Skindex-17. In order to use reliable and valid HRQoL instruments, information is given on the most important psychometric characteristics of measurement instruments (such as scale structure, reliability, validity, and responsiveness), their definitions, and the basic principles of the related statistical tests. Further, background information is given on the methods that are generally being used to establish a clinically meaningful interpretation of HRQoL scores: anchor- and distribution-based methods. In addition, guidance in the selection process of a HRQoL instrument is provided and supported by a flow chart. To date, there is no consensus as to which HRQoL instrument is to be preferred in dermatology. The EADV Taskforce suggested using a combination of a generic- and dermatology-specific instrument in research. For clinical practice purposes, a dermatology-specific instrument is suggested. Lastly, it is recommended that present and future instruments should preferably be tested according to modern test theory models.
Chapter 3. Interpretation of health-related quality of life scores

3.1 Health-related quality of life assessment in dermatology: interpretation of Skindex-29 scores using patient-based anchors

A correct interpretation of HRQoL scores is important in the application of HRQoL data in research and clinical practice, especially because a score in itself has little or no direct meaning. The Skindex-29 is a dermatology-specific HRQoL instrument. It combines 29 items that form three domains: Symptoms, Emotions, and Functioning. Responses are given on a five-point response scale, ranging from ‘never’ to ‘all the time’. The domain scores and overall score are expressed on a 100-point scale, with higher scores indicating lower levels of HRQoL. However, little is known about the interpretation of Skindex-29 scores. Our aim was to facilitate the interpretation of Skindex-29 domain and overall scores by identifying clinically meaningful cut-off scores. Chapter 3.1 presents the results of a multi center survey study. Four hundred and thirty-four consecutively included dermatology outpatients, from nine different dermatological clinics in the Netherlands, were asked to complete the Skindex-29 and a questionnaire consisting of four sets of anchor-based questions. This questionnaire included items on the impact of skin disease on HRQoL, on global disease severity, and on psychiatric morbidity. Of the 339 returned questionnaires (response rate: 78%), 322 questionnaires could be used for data analysis. Pearson’s correlations and Receiver Operating Characteristic curve analysis were used to determine the most optimal cut-off scores for the Skindex-29 domain and overall scores indicating patients with (very) severely impaired HRQoL. Cut-off scores associated with the patient-based anchor on the impact of skin disease on HRQoL showed the highest accuracy (Area Under the Curve ranged from 0.83-0.91). The identified cut-off scores for severely impaired HRQoL were: ≥52 points on the Symptoms domain, ≥39 on Emotions, ≥37 on Functioning, and ≥44 on the overall score. Patients with scores equal to or above the established cut-off scores are significantly affected by their skin disease on that particular domain. This may signal the need for additional care or support.

3.2 Interpretation of Skindex-29 scores: cutoffs for mild, moderate, and severe impairment of health-related quality of life

In a commentary to chapter 3.1, Professor M.M. Chren stressed the relevance of establishing Skindex-29 cut-off scores for mild and moderate degrees of effect in addition to the scores for a severe degree of effect (Addendum I). Chapter 3.2 presents our response to Professor M.M. Chren’s suggestion. We re-analyzed the data of our sample of 322 patients. Again, patient responses on the Skindex-29 were related to the four sets of anchor-based questions, and optimal cut-off scores corresponding to mildly and moderately impaired HRQoL could be identified. The following cut-off scores for mildly impaired HRQoL were found: ≥39 points on Symptoms, ≥24 on Emotions, ≥21 on Functioning, and ≥25 on the overall score. Cut-off scores for moderately impaired HRQoL were: ≥42, ≥35, ≥32 and ≥32 respectively. As a rule of thumb or memory aid in clinical practice, it was suggested to round off the cut-off scores to ≥20, ≥30, and ≥40 points respectively for the domain and overall scores, with the exception of the Symptoms domain.

With the studies on the interpretation of Skindex-29 scores, we were able to identify the clinically meaningfulness of Skindex-29 domain and overall scores. The presented cut-off
scores can be adequately used to identify patients with (very) severely, moderately and/or mildly impaired HRQoL.

3.3 Interpretation of Skindex-29 scores: response to Sampogna and Abeni

In another commentary to chapter 3.2, written by Dr F. Sampogna and Dr D.D. Abeni, two methods being used to categorize Skindex-29 scores were compared: an anchor- and a distribution-based method. The authors indicated that different methods used may result in different categories of Skindex-29 scores and, as a consequence, may have different clinical implications (Addendum II). Unfortunately, in this commentary a misinterpretation of the established Skindex-29 cut-off scores was made, which has led to an incorrect comparison of the categories. Chapter 3.3 presents our reflection on this commentary. A correct overview of the categorization of Skindex-29 cut-off scores is provided. The use of standardized anchors in a future study is recommended, preferably in an international collaboration, as well as a combination of an anchor- and distribution-based method. This would allow an objective comparison.

3.4 Determining the meaningfulness of differences in Skindex-29 scores using item response theory modelling

Modern test theory models, such as item response theory (IRT) or Rasch models, have advantages over classical test theory models. They provide in-depth information on various advanced psychometric characteristics of instruments, such as uni-dimensionality, item weights and item difficulties. In addition, modern test theory models provide further insight in the clinical meaningfulness of HRQoL scores. The Rasch model was used by Professor T.E.C. Nijsten and colleagues to test the Skindex-29. This resulted in a reduced, 17-item Skindex version. A limitation of the Rasch model is that the model does not permit items to differ in their level of discrimination. Only items with a similar level of discrimination are maintained in the questionnaire. The omitted items, however, may provide valuable information that can be used in the clinical management of dermatology patients. The aim of chapter 3.4 was threefold. Our first aim was to examine the discriminating capacity of the Skindex-29 items. Secondly, in the absence of evidence concerning Skindex-29 score differences, our aim was to determine the meaningfulness of differences in Skindex-29 scores. Furthermore, this study offered the possibility to obtain additional information on the psychometric characteristics of the 29-item and 17-item versions of the Skindex. Our third aim was to test the reliability and validity of both Skindex versions. Chapter 3.4 presents the results of an empirical study among 291 dermatology outpatients, consecutively included at nine dermatology outpatient clinics in the Netherlands, who were asked to complete the Skindex-29. The data of these patients were applied to a specific extension of the IRT model, the one-parameter logistic model (OPLM). OPLM was used to establish item weights and to investigate the clinically meaningfulness of Skindex-29 score differences. An independent sample of 183 patients was used for validation purposes.

Preliminary factor analysis indicated that the Skindex-29 consists of two domains: a Psychosocial (Eigenvalue=12.4) and a Symptoms domain (Eigenvalue=3.1). Secondly, the response options of the Skindex-29 were dichotomized (0 = ‘never’, ‘rarely’, and ‘sometimes’; 1 = ‘often’ and ‘all the time’) and an overall fit to the OPLM was demonstrated ($p=0.56$ for the Psychosocial
domain and $p=0.52$ for the Symptoms domain). Subsequently, item weights and item difficulty estimates were identified. Cross-validation on our sample of 183 patients confirmed these results. Fifteen scale cut-points (ranging from ≤5 to ≤68 points) could be distinguished for the Psychosocial domain, and six scale cut-points for the Symptoms domain (ranging from ≤0 to ≤15 points). As items were ordered based on their level of difficulty, each scale cut-point marked a one-increment step of one or more items (complaints) of the Skindex-29. Of the 12 items that were omitted from the original Skindex-29 to form the 17-item version, six items had an above average level of discrimination ($a_i\geq3$). These items represent a substantial part of HRQoL and provide valuable information on item level that can be used in the patient management. Lastly, the reliability and validity coefficients were found to be comparable for both Skindex versions, which is indicative for an overall accuracy of both instruments.

With the results of this study we gained insight into the interpretation of Skindex-29 score differences and we provided an answer to the question what score difference represents a clinically meaningful difference.

**Chapter 4. Health-related quality of life application in clinical practice**

4.1 Health-related quality of life assessment in dermatologic practice: relevance and application

Although HRQoL is generally considered to be an important outcome in clinical research and health care management, measurement of HRQoL and application of HRQoL data in dermatological practice are not yet customary. There may be practical and/or attitudinal barriers, such as completion time, management of HRQoL data, or difficulties in responding to the identified needs of patients, that need to be overcome. Our aim was to support the application of HRQoL measurement in dermatological practice. Chapter 4.1 presents i) why HRQoL measurement is relevant to dermatological practice, ii) which patients would benefit most from routine HRQoL measurement, and iii) how HRQoL assessment can be applied in clinical practice.

First, the measurement of HRQoL may increase patients’ self-awareness and empowerment; it may facilitate patient-centered health care; and it may contribute to the decision-making process for a specific treatment. Further, HRQoL measurement can be used to monitor treatment over time, to facilitate doctor-patient communication, and to improve treatment outcome. Secondly, HRQoL measurement is considered to be particularly beneficial for patients with substantial and enduring adverse effects on HRQoL, or whenever a negative impact on HRQoL is suspected. In patients whose skin disease hardly affects HRQoL, or in patients who consider HRQoL measurement as unnecessary, intrusive, or inappropriate, HRQoL measurement is not recommended. Thirdly, HRQoL data should be easily accessible for clinicians and integrated in the patient medical record. To facilitate this, an electronic version of the Dutch Skindex-29 with automatic scoring was developed by the National Skin Foundation. Answers to each item, as well as an overview of the domain and overall scores visually displayed in a bar chart, are immediately available in print and can be used directly during the patient consultation. This enhances doctor-patient communication and may help clarify priorities of care.
4.2 The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: study protocol for a multicenter randomized controlled trial

In the systemic treatment of psoriasis patients with immunomodulating therapies (i.e., biologic response modifiers or biologics), interest in HRQoL outcome research is rapidly increasing. Despite the positive outcomes of biologics, treatment can only offer a temporary suppression or remission of symptoms. Therefore, treatment efforts are also directed towards an increase of patients’ HRQoL. Previous research, including studies in oncology, suggested that the application of a HRQoL intervention is considered to be an aid for monitoring disease and treatment over time, for the communication with the patient, and for improving treatment outcome.

Routine use of a HRQoL intervention in dermatological practice is relatively new. Moreover, evidence of the efficacy of such an intervention in dermatology is missing. Our aim was to investigate the efficacy of a HRQoL intervention on, primarily, patients’ HRQoL and on doctor-patient communication, and secondary on health status and disease severity. Chapter 4.2 presents a study protocol that describes the rationale and design of a prospective, multicenter randomized controlled trial (RCT) to investigate the effect of a HRQoL intervention in patients with moderate to severe psoriasis who receive biologic treatment. Chapter 4.3 presents the results of this RCT.

4.3 The efficacy of a health-related quality of life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: results of a multicenter randomized controlled trial

Eighteen study sites were cluster-randomized to either the intervention or the control group. Patient recruitment took place between March 2009 and July 2012. A total of 80/116 (69%) patients with moderate to severe psoriasis appeared to be eligible for participation in the study and a 2x50 mg/week etanercept treatment regimen was initiated for all included patients. Of the 80 included patients, 29 patients received the HRQoL intervention, that is: the electronic assessment of the Skindex-29 and communication about the resulting data with their dermatologists. The Skindex-29 was not presented to patients in the control group. Patients were asked to come for six scheduled visits and were followed for 48 weeks. Multilevel analysis was used to investigate the course of the outcomes over time and to assess the efficacy of the HRQoL intervention.

The HRQoL intervention had no effect on the primary and secondary outcomes (effect size=0.001). Nevertheless, the HRQoL intervention facilitated the discussion of HRQoL aspects (in 82.8% of the cases in the intervention group compared to 39.2% in the control group), although at the expense of longer consultation times (31.6 vs. 18.2 minutes respectively, \( p < 0.001 \)). In addition, we found a positive effect on patients’ and physicians’ satisfaction with the process of care (\( p < 0.05 \)). Herewith, the HRQoL intervention may serve as a tool for the improvement of patient care, which should be evaluated by future research.
SAMENVATTING EN CONCLUSIES

Dit proefschrift bestaat uit drie delen. Het eerste deel, ‘Meten van gezondheidsgerelateerde kwaliteit van leven in de dermatologie’, had als doelstelling het leveren van een bijdrage aan de kwaliteit van het meten van kwaliteit van leven (KvL) in dermatologisch onderzoek en in de klinische praktijk. Het tweede deel, ‘Interpretatie van gezondheidsgerelateerde kwaliteit van leven scores’, had als doelstelling het faciliteren van de interpretatie van Skindex-29 scores en score verschillen. Het derde deel, ‘Toepassing van gezondheidsgerelateerde kwaliteit van leven in de klinische praktijk’, had als doelstelling het ondersteunen van de toepassing van het meten van KvL in de dermatologische praktijk en het onderzoeken van de effectiviteit van een KvL interventie op de KvL van patiënten, op de communicatie tussen de dokter en de patiënt, op gezondheidsstatus en op ziekte-ernst.

De algehele doelstelling van het onderhavige proefschrift was het leveren van een bijdrage tot verbetering van gezondheidsgerelateerde KvL van patiënten met een chronische huidaandoening.
Hoofdstuk 1. Algemene introductie

Hoofdstuk 1 is de algemene introductie en beschrijft de achtergrond en de doelstellingen van dit proefschrift en presenteert een overzicht van de verschillende hoofdstukken. Het concept KvL werd toegelicht en haar context in de dermatologie werd beschreven. Vervolgens werd een inleiding gegeven ten aanzien van verschillende typen meetinstrumenten en werd de Skindex-29 geïntroduceerd. Het belang van de psychometrische kwaliteit van meetinstrumenten werd geïllustreerd, als ook de meerwaarde van KvL gegevens voor de klinische praktijk. Tenslotte werd de meerwaarde van de toepassing van het meten van KvL en het gebruik van KvL gegevens in de klinische praktijk beschreven.

Hoofdstuk 2. Meten van gezondheidsgerelateerde kwaliteit van leven in de dermatologie

2.1 Meten van gezondheidsgerelateerde kwaliteit van leven in dermatologisch onderzoek en praktijk: uitkomst van de EADV Taskforce on Quality of Life

In het algemeen wordt KvL als belangrijke uitkomstmaat beschouwd in klinisch onderzoek, bij het vaststellen van ziekte-ernst en in de patiëntenzorg. Een goed begrip van het concept KvL en haar meetinstrumenten is hierbij essentieel voor zowel onderzoekers als clinici. De toepassing van het meten van KvL in de dermatologie wordt aangemoedigd door de European Academy of Dermatology and Venereology (EADV) Taskforce on Quality of Life. Onze doelstelling was het leveren van een verdere bijdrage aan de kwaliteit van het meten van KvL in dermatologisch onderzoek en klinische praktijk. Hoofdstuk 2.1 presenteert een overzichtsartikel dat werd geschreven namens de EADV Taskforce en bedoeld voor clinici en junior onderzoekers die relatief onbekend zijn met het meten van KvL. Het concept KvL en de methodologie omtrent het meten van KvL werden geïntroduceerd en de in de dermatologie meest gebruikte generieke (d.w.z. te gebruiken voor alle ziekten) en ziekte-specifieke (d.w.z. te gebruiken voor een specifieke (huid)ziekte) meetinstrumenten werden samengevat. Een voorbeeld van een generiek meetinstrument is de Medical Outcomes Study 36-item Short Form Health Survey (SF-36). Voorbeelden van ziekte-specificiteitsmeetinstrumenten zijn de Dermatology Life Quality Index (DLQI), de Skindex-29 en de Skindex-17. Om betrouwbare en valide KvL meetinstrumenten te gebruiken, worden de meest belangrijke psychometrische kenmerken (~meet eigenschappen) van meetinstrumenten beschreven, waaronder schaalstructuur, betrouwbaarheid, validiteit en responsiviteit. Definities van deze begrippen werden beschreven als ook de basisprincipes van de bijbehorende statistische toetsen. Verder werd achtergrondinformatie gegeven over de methoden die doorgaans worden toegepast bij het vaststellen van een klinisch relevante interpretatie van KvL scores, te weten: anker- en distributiemethoden. Daarnaast werd richting gegeven aan het selecteren van een KvL meetinstrument. Dit werd ondersteund door een schematische weergave van het selectieproces.

Tenslotte werd aanbevolen om de huidige en toekomstige meetinstrumenten bij voorkeur te testen volgens moderne testtheorie methoden.

**Hoofdstuk 3. Interpretatie van gezondheidgerelateerde kwaliteit van leven scores**

3.1 Meten van gezondheidgerelateerde kwaliteit van leven in de dermatologie: interpretatie van Skindex-29 scores door middel van een ankermethode

Een correcte interpretatie van KvL scores is belangrijk in de toepassing van KvL gegevens in onderzoek en klinische praktijk, vooral omdat een score zelf weinig tot geen directe betekenis heeft. De Skindex-29 is een dermatologie-specifiek KvL meetinstrument. Deze vragenlijst bestaat uit 29 vragen welke drie domeinen vormen, te weten: Symptomen, Emoties en Functioneren. Antwoorden kunnen worden gegeven op een vijfpunts-antwoordschaal, welke varieert van ‘nooit’ tot ‘altijd’. De domein scores en de totaal score worden uitgedrukt op een 100-puntsschaal, waarbij hogere scores een verminderde kwaliteit van leven betekenen. Echter, tot op heden is er weinig bekend over hoe de scores van de Skindex-29 kunnen worden geïnterpreteerd. Onze doelstelling was het faciliteren van de interpretatie van Skindex-29 domein en totaal scores door het identificeren van klinisch betekenisvolle afkapwaarden.

Hoofdstuk 3.1 presenteert de resultaten van een multicenter vragenlijstonderzoek. Vierhonderd vierendertig dermatologische patiënten, van negen verschillende dermatologische centra in Nederland, werden opeenvolgend geïncludeerd tijdens de algemene spreekuren op de verschillende poliklinieken. Aan deze patiënten werd gevraagd om de Skindex-29 in te vullen, als ook voor de invloed van de huidaandoening op KvL, over ziekte-ernst en over psychiatrische morbiditeit. Van de 339 geretourneerde vragenlijsten (respons: 78%) konden 322 vragenlijsten worden gebruikt voor de data analyse. Pearson’s correlaties en Receiver Operating Characteristic curve analyse werden gebruikt om de meest optimale afkapwaarden vast te stellen voor de Skindex-29 domein en totaal scores, om zo patiënten te kunnen identificeren met een (zeer) ernstige verminderde KvL. De afkapwaarden geassocieerd met de ankervraag over de invloed van de huidaandoening op KvL lieten de hoogste mate van accuratesse zien (Area Under the Curve varieerde van 0.83-0.91). De afkapwaarden die konden worden vastgesteld voor een ernstige vermindering van KvL waren: ≥52 punten voor het Symptomen domein, ≥39 voor Emoties, ≥37 voor Functioneren en ≥44 voor de totaalscore. Patiënten met scores die gelijk of hoger zijn dan de vastgestelde afkapwaarden, zijn door hun huidaandoening significant aangedaan in hun KvL op dat specifieke domein. Dit zou een signaalfunctie kunnen zijn voor de behoefte aan additionele zorg of ondersteuning.

3.2 Interpretatie van Skindex-29 scores: afkapwaarden voor milde, matige en ernstige vermindering van gezondheidgerelateerde kwaliteit van leven

Publicatie van de in hoofdstuk 3.1 beschreven studie was reden voor Professor M.M. Chren om een commentaar te schrijven waarin zij de relevantie benadrukte van het eveneens vaststellen van Skindex-29 afkapwaarden voor milde en matige vermindering van KvL, in aanvulling op de afkapwaarden voor ernstige vermindering van KvL (Bijlage I). Hoofdstuk
3.2 presenteert onze reactie op het voorstel van Professor M.M. Chren. De data van onze studiepopulatie van 322 patiënten werden opnieuw geanalyseerd. De antwoorden op de Skindex-29 vragen werden opnieuw gerelateerd aan de vier sets met ankervragen en de optimale afkapwaarden voor een milde en matige vermindering van KvL konden worden vastgesteld. De volgende afkapwaarden voor een milde vermindering van KvL werden gevonden: ≥39 punten voor Symptomen, ≥24 voor Emoties, ≥21 voor Functioneren en ≥25 voor de totaal score. Afkapwaarden voor een matige vermindering van KvL waren respectievelijk ≥42, ≥35, ≥32 and ≥32. Als vuistregel of geheugensteun voor het gebruik van deze afkapwaarden in de klinische praktijk werd voorgesteld om deze afkapwaarden af te ronden naar respectievelijk ≥20, ≥30, en ≥40 punten voor de domein en totaal scores, met uitzondering van het Symptomen domein.

Met deze studies naar de interpretatie van Skindex-29 scores kon een klinisch relevante betekenis worden geven aan Skindex-29 domein en totaal scores. De gepresenteerde afkapwaarden kunnen adequaat worden gebruikt om patiënten te identificeren met een (zeer) ernstige, matige en/of milde vermindering van KvL.

3.3 Interpretatie van Skindex-29 scores: reactie op Sampogna en Abeni

Publicatie van de in hoofdstuk 3.2 beschreven resultaten was reden voor Dr. F. Sampogna en Dr. D.D. Abeni om een commentaar te schrijven waarin twee methoden, een ankermethode en een distributiemethode, met elkaar werden vergeleken. De auteurs gaven aan dat het gebruik van verschillende methoden kan resulteren in verschillende categorieën van Skindex-29 scores en dat dit, als consequentie, verschillende implicaties kan hebben voor de klinische praktijk (Bijlage II). Helaas is uit dit commentaar gebleken dat de vastgestelde afkapwaarden voor Skindex-29 scores foutief werden geïnterpreteerd. Dit heeft geleid tot een incorrecte vergelijking van de verschillende categorieën. Hoofdstuk 3.3 presenteert onze reactie op dit commentaar. Een correct overzicht van de categorisatie van Skindex-29 scores werd gepresenteerd. Het gebruik van gestandaardiseerde ankers in toekomstig onderzoek werd aanbevolen, bij voorkeur in een internationaal samenwerkingsverband, als ook een combinatie van ankermethode kan worden gemaakt.

3.4 Het vaststellen van de betekenis van Skindex-29 scoreverschillen door middel van item response theory modellering

in de dermatologische patiëntenzorg. De doelstelling van hoofdstuk 3.4 was drieledig. Onze eerste doelstelling was het onderzoeken van het discriminerend vermogen van de vragen uit de Skindex-29. Ten tweede, in afwezigheid van evidentie ten aanzien van Skindex-29 scoreverschillen, was onze doelstelling om de betekenis van Skindex-29 scoreverschillen vast te stellen. Daarnaast bood dit onderzoek de mogelijkheid om additionele informatie te vergaren over de psychometrische kenmerken van de Skindex-29 en de Skindex-17. Onze derde doelstelling was het testen van de betrouwbaarheid en de validiteit van beide Skindex versies. Hoofdstuk 3.4 presenteert de resultaten van een empirische studie uitgevoerd onder 291 dermatologische patiënten. Deze patiënten werden opeenvolgend geïncludeerd op de poliklinieken van negen verschillende dermatologische centra in Nederland en aan hen werd gevraagd om de Skindex-29 in te vullen. De data van deze patiënten werden gebruikt voor een specifieke extensie van het IRT model, namelijk het een-parameter logistisch model (OPLM). Dit model werd gebruikt ten behoeve van het vaststellen van item gewichten en om de klinisch relevante betekenis van Skindex-29 scoreverschillen te onderzoeken. Een onafhankelijke studiepopulatie van 183 patiënten werd gebruikt voor validatie doeleinden. Voorbereidende factor analyse liet zien dat de Skindex-29 uit twee domeinen bestaat: een Psychosociaal domein (Eigenvalue=12.4) en een Symptomen domein (Eigenvalue=3.1). Vervolgens werden de antwoordcategorieën van de Skindex-29 gedichotomiseerd (0 = ‘nooit’, ‘zelden’ en ‘soms’; 1 = ‘vaak’ en ‘altijd’) en de algehele fit van de data op het OPLM kon worden aangetoond (p=0.56 voor het Psychosociale domein en p=0.52 voor het Symptomen domein). Hierna konden de item gewichten en de item moeilijkheid worden vastgesteld. Validatie in de studiepopulatie van 183 patiënten bevestigde deze resultaten. Vijftien afkappunten konden worden onderscheiden voor het Psychosociale domein ( variërend van ≤5 tot ≤68 punten) en zes afkappunten voor het Symptomen domein ( variërend van ≤3 tot ≤15 punten). Omdat de vragen werden geordend naar moeilijkheidsgraad, markeert ieder volgend afkappunt een toename van één of meer vragen (klachten) van de Skindex-29. Van de 12 vragen die niet werden behouden in de Skindex-17 hadden zes vragen een bovengemiddeld discriminerend vermogen (ai≥3). Deze vragen representeren een substantieel deel van KvL en leveren op item-niveau waardevolle informatie op dat kan worden gebruikt in de patiëntenzorg. Tenslotte, de betrouwbaarheid en validiteit van beide Skindex versies zijn vergelijkbaar, wat betekent dat beide meetinstrumenten accuraat zijn. Met de resultaten van deze studie hebben we meer inzicht verkregen in de interpretatie van Skindex-29 scoreverschillen en konden we antwoord geven op de vraag welk score verschil een klinisch relevant verschil representeert.

**Hoofdstuk 4. Toepassing van gezondheidsgerelateerde kwaliteit van leven in de klinische praktijk**

4.1 Meten van gezondheidsgerelateerde kwaliteit van leven in de dermatologische praktijk

Ondanks het feit dat KvL in het algemeen als belangrijke uitkomstmaat wordt beschouwd in zowel klinisch onderzoek als in de gezondheidszorg, is het meten van KvL en de toepassing van KvL gegevens in de dermatologische praktijk nog niet gangbaar. Redenen hiervoor kunnen
praktisch van aard zijn en/of er kan sprake zijn van belemmering door attitude ten aanzien van het meten van KvL, zoals de tijd die nodig is om een vragenlijst in te vullen, het omgaan met de KvL gegevens, of het ondervinden van moeilijkheden in het reageren op specifieke behoeften van de patiënt die hieruit naar voren kunnen komen. Deze obstakels dienen te worden weggenomen. Onze doelstelling was om de toepassing van het meten van KvL in de dermatologische praktijk te ondersteunen. Hoofdstuk 4.1 presenteert i) waarom het meten van KvL relevant is voor de dermatologische praktijk, ii) welke patiënten het meeste baat hebben bij het routinematig meten van KvL, en iii) hoe het meten van KvL kan worden toegepast in de klinische praktijk.

Ten eerste, het meten van KvL kan het zelfbewustzijn en ziekte-inzicht van patiënten bevorderen; het kan patiëntgerichte gezondheidszorg faciliteren; en het kan relevante informatie opleveren voor de dermatologische behandeling. Daarnaast kan het meten van KvL worden ingezet om de behandeling over de tijd te monitoren, het kan de communicatie tussen de dokter en de patiënt bevorderen en het kan het resultaat van de behandeling verbeteren. Ten tweede, het meten van KvL heeft in het bijzonder meerwaarde voor patiënten die substantiële en langdurig negatieve effecten ondervinden van hun huidaandoening op hun KvL, of waarbij het vermoeden bestaat van een aanzienlijk negatief effect op KvL. Bij patiënten waarbij de huidaandoening nauwelijks van invloed is op KvL, of bij patiënten die het meten van KvL onnodig, opdringerig of ongepast vinden, wordt het meten van KvL niet aanbevolen. Ten derde, clinici dienen gemakkelijk toegang te hebben tot KvL gegevens en deze gegevens dienen aan het patiëntendossier te worden opgenomen. Om dit te faciliteren werd een elektronische versie van de Skindex-29, met een automatisch score systeem, ontwikkeld door het Huidfonds. De antwoorden op de vragen, als ook een visueel overzicht van de domein scores en de totaal score in de vorm van een grafiek zijn, direct beschikbaar in printbare vorm en kan direct worden gebruikt tijdens het spreekuur. Dit bevordert de communicatie tussen de arts en de patiënt en het kan de prioriteiten van de zorg verhelderen.

4.2 De effectiviteit van een gezondheidsgerelateerde kwaliteit van leven interventie gedurende 48 weken behandeling met een biologic van patiënten met een matige tot ernstige psoriasis: een studieprotocol voor een multicenter, gerandomiseerde, gecontroleerde studie

In de systemische behandeling van psoriasis patiënten met immunomodulerende therapieën (i.e., biologics) neemt de belangstelling voor KvL uitkomstonderzoek snel toe. Ondanks de positieve resultaten met biologics biedt de behandeling enkel een tijdelijke onderdrukking of remissie van de symptomen. Hierdoor richt de behandeling zich eveneens op een verbetering van de KvL van patiënten. Eerder onderzoeken, waaronder studies binnen de oncologie, suggereerden dat de toepassing van een KvL interventie kan worden beschouwd als een middel om ziekte en behandeling over de tijd te kunnen monitoren, voor de communicatie tussen de dokter en de patiënt en voor het verbeteren van het resultaat van de behandeling. Routinematig gebruik van een KvL interventie in de dermatologische praktijk is relatief nieuw. Bovendien ontbreekt het in de dermatologie momenteel aan evidentie over de effectiviteit van een dergelijke interventie. Onze doelstelling was het onderzoeken van de effectiviteit van een KvL interventie op, primair, de KvL van patiënten en de communicatie tussen arts en patiënt en, secundair, op gezondheidsstatus en ziekte-ernst. Hoofdstuk 4.2
4.3 De effectiviteit van een gezondheidsgerelateerde kwaliteit van leven interventie gedurende 48 weken behandeling met een biologic van patiënten met een matige tot ernstige psoriasis: resultaten van een multicenter, gerandomiseerde, gecontroleerde studie

Achtteien studiecentra werden gerandomiseerd in clusters naar ofwel de interventiegroep ofwel de controlegroep. Tussen maart 2009 en juli 2012 werden patiënten geworven voor de studie. Een totaal van 80/116 (69%) patiënten met matige tot ernstige psoriasis bleken te voldoen aan de inclusiecriteria voor deelname aan de studie en startten een behandelingsschema van 2x50mg/week etanercept. Van de 80 geïncludeerde patiënten ontvingen 29 patiënten de KvL interventie. Deze bestond uit de elektronische afname van de Skindex-29 en terugkoppeling van de resultaten tijdens het spreekuur met de dermatoloog. Patiënten in de controle groep ontvingen de KvL interventie niet. Aan patiënten werd gevraagd om gedurende een periode van 48 weken, zes maal te verschijnen op het spreekuur. Multilevel analyses werden gebruikt om de uitkomsten over de tijd te onderzoeken en om het effect van de KvL interventie op deze uitkomsten te kunnen vaststellen.

De KvL interventie bleek geen effect te hebben op de primaire en secundaire uitkomsten (grootte van het effect = 0.001). Desondanks bleek de KvL interventie de discussie van KvL aspecten tijdens het spreekuur te faciliteren (bij 82.8% van de patiënten in de interventiegroep ten opzichte van 39.2% in de controlegroep). Echter, dit ging ten koste van de duur van het spreekuur (respectievelijk 31.6 minuten versus 18.2 minuten, p<0.001). Daarnaast werd een positief effect gevonden op de tevredenheid met het zorgproces, door zowel patiënten als artsen (p<0.05). De KvL interventie kan hiermee als middel worden ingezet ter verbetering van de patiëntenzorg. Dit dient door middel van vervolgonderzoek te worden geëvalueerd.
INTRODUCTION

In the last decades, a lot has changed in the management of patients with chronic skin diseases. The introduction of health-related quality of life (HRQoL) in dermatology by Professor A.Y. Finlay and Dr G.K. Khan in 1994 has played an important role in this development. Putting the patient in the center of health care management and, thereby, aiming to improve patient care and shared decision making, requires reliable, valid, and interpretable patient-reported outcome measures (PROMs).

The overall aim of this thesis was to contribute to the improvement of HRQoL of patients with a chronic skin disease. The focus of the thesis was on (i) the measurement of HRQoL in dermatology, (ii) the interpretation of HRQoL data, and (iii) the application of a HRQoL intervention in daily clinical practice. First, in order to contribute to the quality of the measurement of HRQoL in dermatologic research and clinical practice, understanding of the concept HRQoL and the related methodology are prerequisites (chapter 2.1). Second, to effectively use HRQoL data in clinical practice, insight into the interpretation of HRQoL scores, or differences in scores, is required (chapters 3.1 through 3.4). Third, to support the application of a HRQoL intervention in clinical practice, it is essential that clinicians recognize and acknowledge the relevance of HRQoL measurement and the application of HRQoL data in their patient consultations (chapter 4.1). Finally, to enhance the implementation of a HRQoL intervention during patient consultations, the effectiveness of such an intervention in a clinical setting should be established (chapter 4.2 and 4.3).

The chapters of the thesis are summarized (chapter 5). This final chapter (chapter 6) reflects on the main findings and limitations, and concludes with a discussion on future perspectives.

WHERE ARE WE?

Measurement of health-related quality of life in dermatology

To contribute to the quality of HRQoL measurement in dermatology, in chapter 2.1 we presented a review article on the measurement of HRQoL in dermatological research and clinical practice. An introduction to the concept and methodology of HRQoL was provided, an overview of the most commonly used HRQoL instruments in dermatology was presented, background information was provided on the psychometric quality of these instruments, and a suggestion for the selection of a HRQoL instrument was given.

This review article, which was written on behalf of the European Academy of Dermatology and Venereology (EADV) Taskforce on Quality of Life, was intended for clinicians and junior researchers who are relatively unfamiliar with HRQoL measurement in dermatology. Therefore, less attention was given to modern test theory models. However, as modern test theory models are today generally acknowledged as a step forward in HRQoL instrument development and testing, it was highly recommended that instruments based on classical test theory models, as well as future instruments, are being (re-)analyzed according to modern test theory models. It was suggested that the EADV Taskforce should pay specific attention to modern test theory models in a subsequent article.

In addition, it was discussed that to date there is no univocal consensus among researchers and clinicians as to which particular HRQoL instrument should be used in dermatology. Hence,
similar outcomes will continuously be measured with different instruments. Agreement on a preferred HRQoL instrument, however, would enable adequate comparison of scores and syntheses of data across studies.

**Interpretation of health-related quality of life scores**

To effectively use HRQoL data in clinical practice, insight into the interpretation of HRQoL scores, or differences herein, are important prerequisites. In chapters 3.1 and 3.2, we presented clinically meaningful cut-off scores for the Skindex-29, using an anchor-based method. The established Skindex-29 domain and overall cut-off scores are indicative for patients with a mildly, moderately, or severely impaired HRQoL. With these cut-off scores, we were able to facilitate the interpretation of Skindex-29 scores. This may aid in understanding what these scores mean to dermatological patients whose HRQoL might be affected by their skin diseases. Relatively similar cut-off scores were found for mild and moderate impairment on the Symptoms domain. This resulted from the lower correlation of the particular anchor question with the corresponding Skindex-29 domain score and the lower discriminating capacity between patients who experienced a mildly or moderately impaired HRQoL. In addition, the established cut-off scores were generally higher compared to the ranges of scores that were established by a distribution-based method, presented by Professor T.E.C. Nijsten and colleagues. This may be the result of different methods used, differences in the distribution of diagnosis, and/or HRQoL scores of the samples.

In a commentary by Dr F. Sampogna and Dr D.D. Abeni, our cut-off scores were compared to ranges of scores of Professor T.E.C. Nijsten and colleagues. They showed that the use of different methods in different study populations does result in different categorizations of scores and, as a consequence, may have different clinical implications. In chapter 3.3 we reflected on this commentary. We suggested performing a future study to investigate the interpretability of Skindex-29 scores by combining anchor-based and distribution-based methods in a single study population. This would allow an adequate comparison of the categorization of scores.

In addition to the categorization of Skindex-29 scores, in chapter 3.4 we presented clinically meaningful differences in scores, using an extension of an item response theory (IRT) model, the one-parameter logistic model (OPLM). With the results of this study, we provided an answer to the question which score difference represents a clinically meaningful difference on item level in terms of complaints.

By applying Skindex-29 data to the OPLM, not only the clinical meaningfulness of differences in Skindex-29 scores could be determined, it also provided additional insight into the psychometric characteristics of the Skindex-29. It became very clear that the Skindex-29 consists of two instead of three domains, namely a Psychosocial and a Symptoms domain. In addition, it was confirmed that the original scale is not truly uni-dimensional, which is in agreement with the results found by Professor T.E.C. Nijsten and colleagues. In contrast with this study on the development of the Skindex-17, which is a Rasch reduced version of the Skindex-29, we found both clinical and psychometric arguments for maintaining all 29 items in the questionnaire. In addition, our results also suggested that Skindex-29 items clearly differed in their discriminating capacity and, thus, needed weighting before calculating domain scores.
The Skindex-29 in its current format does not meet the current high standards of instrument development. It was therefore suggested to perform confirmative and conclusive research on the original scoring system of the Skindex-29.

Health-related quality of life application in clinical practice

To support the use of HRQoL measurement in clinical practice, in chapter 4.1 we presented an overview of the relevance and application of the measurement of HRQoL in dermatological practice. We described why the measurement of HRQoL is important, which patients might benefit from it, and how HRQoL measurement can be applied in clinical practice. However, to successfully implement HRQoL measurement in clinical practice, valid conclusions with regard to the added value of a HRQoL intervention are essential.

Previous randomized controlled trials (RCTs) measuring the efficacy of patient reported outcome (PRO) interventions in other medical specialties have been highly heterogeneous in setting, patients, intensity and content of the PRO intervention. A diversity of outcomes was reported that makes it very difficult to interpret the evidence on the effectiveness of such interventions in clinical practice. Although Dr J.M. Valderas and colleagues could not draw valid conclusions regarding the effectiveness of such interventions, the possible impact of PRO measurement in clinical practice included the detection of physical or psychological problems that might otherwise be overlooked, the monitoring of disease progression, the facilitation of doctor-patient communication and shared decision making, and the potential of enhancing patient-centered care.

In dermatology, empirical data on the effectiveness of a HRQoL intervention in clinical practice is missing. In chapter 4.2 we described the rationale and design of a RCT investigating the efficacy of a HRQoL intervention on, primarily, patients’ HRQoL and doctor-patient communication and, secondary, on health status and disease severity in patients with moderate to severe psoriasis. In chapter 4.3 we presented the results of this RCT. We did not find evidence of efficacy of a HRQoL intervention in dermatological practice on the primary and secondary study outcomes. These results are in contrast with a study performed by Professor G. Velikova and colleagues, but consistent with other findings. Nevertheless, we observed a positive effect on patients’ and physicians’ satisfaction with the process of care, and the discussion of HRQoL aspects during the patient consultations was enhanced. The intended effect of the HRQoL intervention was reflected by the number of HRQoL aspects discussed, but at the expense of a longer consultation time. These results are in agreement with other studies, with the exception of the prolonged consultation. The discussion of HRQoL aspects may provide valuable information that can be used in patient management. Herewith, the HRQoL intervention may serve as a tool for improving patient care, which should be evaluated by future research.

WHERE SHOULD WE BE GOING?

The application of HRQoL measurement in dermatological practice and the use of HRQoL data may still interfere with current routine clinical practice, and outcomes research in dermatology remains challenging. In addition, insight into and knowledge of instrument development and testing has evolved.
Consensus on the preferred outcome measure

In the absence of consensus on a preferred HRQoL instrument in dermatology, the selection of an instrument remains a trade-off between, among others, the quality of an instrument, the research question, and target population. However, uniformity in PRO measurement, data collection and reporting will overcome limitations such as inconsistencies in the outcomes reported and difficulties in the interpretation, comparability, and synthesis of these outcomes.22 The importance of uniformity is being recognized by several international project groups. The International Society for Quality of Life Research (ISOQOL) developed standards for the design and selection of PROMs to be used in patient-centered outcomes research and comparative effectiveness research.23,24 The CONsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) Initiative aims to improve the selection of instruments, and developed a checklist containing standards for evaluating the methodological quality of studies on the psychometric characteristics of instruments (www.cosmin.nl/). Further, the Harmonising Outcome Measures for Eczema (HOME) Initiative in dermatology (www.homeforeczema.org/) and the Outcome Measures in Rheumatology (OMERACT) Initiative (www.omeract.org/) are developing consensus-based sets of outcomes to be measured in clinical research. In addition, the Core Outcome Measures in Effectiveness Trials (COMET) project group undertakes different initiatives to support the development and application of agreed standardized sets of outcomes (www.comet-initiative.org/). These project groups are closely connected and are currently investigating which outcomes should be measured and reported as a minimum in all clinical trials of a specific condition. A subsequent next step would be how to measure these outcomes; i.e., which instruments should be used.

In the absence of consensus on a preferred instrument, consensus-based methods, such as (international) Delphi studies, are useful methods to establish consensus on the instrument of choice. These methods are used to synthesize data from experts, based on expertise and scientific background.24,25 With an international Delphi study it would be possible to achieve consensus on the (type of) HRQoL instrument to be used in dermatology.

Modern test theory models

The preferred instrument to measure HRQoL should be of high methodological quality, that is, developed and tested according to modern test theory models. On either clinical or methodological grounds, the present dermatology-specific HRQoL instruments in their current formats do not meet these higher standards of instrument development and testing.5,26,27 Together with the absence of consensus on the preferred HRQoL instrument, we are currently standing on a crossroad. Outcomes researchers and clinicians should decide which direction to take: further refinement of existing HRQoL instruments and addressing the suggestions for future research, or developing new instruments according to innovative and promising methods. An ultimate next step in the evolution of PRO measurement in dermatology would be the development and/or validation of (existing) item banks for (domains of) HRQoL. Item banks are based on IRT models and are sets of items with their associated calibrations (i.e., hierarchy) that measure a certain construct. Item banks enable computer-adaptive testing (CAT). CAT is a method of administering an item bank by computer: the questions that are asked, adapt to the patient’s response to the previous question (i.e., ability level). With this,
item banks and CAT provide tailored measurements that reduce the test burden of both patients and clinicians, and provide efficient, precise, valid, and responsive instruments that can be adapted to different health conditions. Because of the hierarchy of the items, scores can be compared and data can be synthesized. For example, the Patient-Reported Outcome Measurement Information System (PROMIS) Network, funded by the National Institutes of Health, is developing series of item banks and provide clinicians and researchers access to these item banks (www.nihpromis.org). The development of item banks and CAT is innovative and considered to be the next step in the development of PROMs.28

Future perspectives
In dermatology, future research should focus on the development and validation of item banks for (domains of) HRQoL to meet the current high quality standards for PRO measurement. This may ultimately add to the application of standardized and uniform PROMs in dermatological research and clinical practice.

CONCLUSIONS
In the changing management of patients with chronic skin diseases, the culture of outcomes research and clinical practice in dermatology need to alter as well. It is well known that the severity of a skin disease is not fully reflected by the characteristics of the skin disease only; it is also illustrated and reflected by aspects that can only be determined by PROs.29 A good understanding of these aspects, as well as a positive attitude towards HRQoL measurement, good communication skills, skills in the interpretation of scores, and the willingness to apply such an important outcome parameter, are prerequisites to successfully implement HRQoL assessment in clinical practice.30,31 What we need are dedicated outcomes researchers, methodologists, and psychometricians who continue working on this important patient-centered outcome in research, as well as clinicians willing to apply PROMs in clinical practice.

REFERENCES
9. Lohr KN. Applications of health status assessment measures in clinical practice. Overview of the


Interpretation of quality-of-life scores

Patients’ reports of their experience with illness are a key health outcome, but scores that measure these reports can be difficult to interpret. Skindex-29 measures skin-related quality of life reliably and validly. In this issue, Prinsen and colleagues compare patients’ responses to “anchor” questions with their Skindex scores to derive clinically meaningful scores for the subscales of Skindex-29. The cutoff scores identify patients whose skin diseases severely affect their quality of life.
Clinicians routinely interpret clinical measurements such as blood pressure or hematocrit: almost without thinking, we “know” what the values mean for a patient’s health. This knowledge is based on our experience caring for many patients and on what we have learned from teachers and reading. For many measurements in clinical studies we lack this intuition, however. One of the original studies of finasteride for benign prostatic hyperplasia graphically illustrates this point.\(^1,2\)

Compared with placebo, the drug improved urine flow by an average of 3 ml/second – a rather bland finding, to be sure – until a subsequently published epidemiological study found that for men aged 40–74 typical urine flow rates decline approximately 0.2–0.3 ml/second per year of life.\(^3\) We can now interpret the real-world effect of finasteride: on average, it restores a man’s urinary flow to what it was 10–15 years earlier.

Interpreting measurements such as urine flow is relatively straightforward compared with interpreting more abstract health outcomes such as scores of patients’ reports concerning their experience with illness. The actual measurement of psychometric constructs is a highly advanced and rigorous science, and substantial progress has been made in applying this science to clinical medicine.\(^4\) Dermatology has lagged somewhat in rigorous studies to interpret the meaning of psychometric scores, however, and these scores remain unfamiliar to most researchers and clinicians. It is highly fitting that in what the Journal has designated the Year of the Patient,\(^5\) this issue contains a good example of a study to facilitate interpretation of one measurement of patients’ experience of illness.\(^6\) In this Commentary, I describe briefly where we are with respect to measuring patients’ reports, summarize the major findings of that article, and project where we might go next to advance this important aspect of clinical research.

Patients’ reports of their experience with illness are a key health outcome. This observation is especially true for skin diseases, which do not typically affect survival, laboratory values, or easily measured clinical changes. In fact, patients’ reports are arguably an essential health outcome for dermatology because skin diseases (unlike most “internal” diseases) can change appearance, and they may have psychological and functional effects that cannot be assessed except through patients’ reports.

Most scientific work on the assessment of patients’ experience with cutaneous illness has focused on instruments that measure skin-related quality of life. Generic and disease-specific quality-of-life instruments have been developed for dermatology and found to have reliability (i.e., they give the same result when quality of life is the same), validity (i.e., they measure quality of life), and responsiveness (i.e., they change when quality of life changes).\(^4\) But fewer data exist on the interpretability of scores with these instruments. What does a given score mean? Does the score indicate severe effects of the disease or mild effects? What do changes in scores mean? Have the effects of the disease changed substantially or only by a small amount?

By simply examining the content of questions and patients’ responses, one can begin to interpret a score, especially for a single question. For example, a typical item in Skindex-29 is “I am embarrassed by my skin condition.” Response choices and corresponding scores are “never” (0), “rarely” (25), “sometimes” (50), “often” (75), or “all the time” (100). If a patient’s item score is 25, we understand that he or she is only rarely embarrassed by the skin problem.
This “content-based” interpretation is less straightforward for scales that are derived from multiple items, however. Skindex is a multiscale index for which subscores are reported for Symptoms, Emotions, and Functioning. What can we do to put scale scores into context so that their meaning can be understood by clinicians?

A useful framework categorizes interpretation methods as either distribution-based or anchor-based. Distribution-based interpretations are based on the statistical distributions of scores in a given population. For example, I can begin to understand the magnitude of the effect my patients report by comparing their scores with those of a “normative” sample of unaffected persons (or of persons known to be severely affected). A recent paper using a distribution-based method reported that the distribution of responses to Skindex-29 could be clustered into statistically distinct categories based on the degree of reported quality-of-life effect. For the Symptoms subscale, for example, the categorization permitted cutoff values that corresponded to “very little” effect (≤3), “mild” effect (4–10), “moderate” effect (11–25), “severe” effect (26–49), and “extremely severe” effect (≥50).

Anchor-based interpretations, on the other hand, are made when scores are compared, or anchored, to other clinical results. A commonly used anchor is the response of patients to global rating questions that are themselves easily interpreted; in the current study, Prinsen et al. (2010) use this strategy to help interpret the meaning of Skindex-29 scores. The investigators administered Skindex-29 and a variety of anchor questions to a large sample of dermatology outpatients. The analyses compared Skindex subscale scores to patients’ responses to three major types of anchors: global questions about aspects of health-related quality of life, a question about the patient’s estimate of the clinical severity of his or her skin disease, and results on a standardized measure of psychiatric morbidity. For each of the anchors, the investigators predefined scores that indicate “severe” effects. They then determined the minimum Skindex scores (cutoff scores) that were most accurate in distinguishing patients who did or did not report severe effects. Skindex cutoff scores for severely impaired skin-related quality of life were ≥37 for Functioning, ≥39 for Emotions, and ≥52 for Symptoms. Prinsen et al. (2010) used receiver-operating characteristic (ROC) methodology to determine the accuracy of the cutoff score. ROC curves are commonly used to display the ability of a diagnostic test to distinguish between people with or without the condition of interest by describing the performance of the test as the relation between the true-positive rate and the false-positive (1-sensitivity) rate. Different cutoffs of scores have different sensitivities and specificities in relation to the criterion in question (e.g., global health-related quality of life). To determine cutoffs for Skindex scores, the authors selected the cutoff that maximized the sum of sensitivity and specificity. This decision does not ipso facto have clinical meaning but requires a judgment about the relative benefits and liabilities of accurately detecting and not missing severe quality-of-life effects. With justification, the authors could also have chosen different levels of sensitivity and specificity (e.g., to maximize true positives at the expense of also increasing the numbers of false positives). Their strategy seems reasonable, however. Lowering the Skindex Symptom cut-off score for severe quality-of-life effects from 52 to, say, 45 would have detected more patients with severe Symptom quality-of-life effects as measured by the global item, but also would have labeled some patients as severely affected who in fact did not have severe effects as measured by the global item.
I look forward to seeing more results from this important and careful study, particularly the cutoff scores for mild and moderate degrees of effect, as determined by the anchors. Such results would permit us to interpret changes in scores if, for example, a group of patients changed from “moderate” quality-of-life effects to “mild” effects over time or after an intervention. They would also permit a more in-depth comparison to the Skindex-29 cutoffs derived from the distribution method described above. Although the results for the interpretation of “severe” quality-of-life effects are similar in the two studies for Symptoms and Functioning, the cut-off for severe Emotional effects in the current study (≥39) would be classified as indicating only moderate effects in the distributional study. Using different anchors can inform interpretation even more. Prinsen et al. (2010) determined Skindex cutoff scores for patients’ judgments of severity of their disease and for their responses to a measure of psychiatric morbidity. In other work, changes in generic health-related quality of life have been correlated with the impact of stressful life events, with being diagnosed with chronic diseases, with resource utilization, and with survival.

Clinically meaningful interpretations of quality-of-life scores are important to patients, clinicians, researchers, public health personnel, and policy makers. These individuals will be comfortable with these scores only when they become familiar, which will require their routine use in clinical research and possibly in practice. But routine use alone is not sufficient. Even if widely interpretable scores are obtained and reported in clinical trials, the results may not be used to modify the conclusions. Ultimately, to improve clinical decision making, an explicit commitment to including patients’ perspectives is necessary in clinical research.

ACKNOWLEDGMENTS
This work was supported by National Institutes of Health grant K24 AR052667.

CLINICAL IMPLICATIONS

- Tools to measure patients’ reports of their experiences with illness must measure accurately, and the results must be interpretable.
- Survey instrument scores are often not meaningful because they are unfamiliar and not empirically derived or clinically based.
- Information about interpretation of quality-of-life scores provides a basis for their more widespread use in clinical medicine and research.
REFERENCES


Interpretation of Skindex-29 scores
ABSTRACT
In this Commentary, we compare two categorizations of a dermatological health–related quality-of-life (HRQoL) instrument, the Skindex-29. One was created on the basis of an anchor-based method, the other on a distribution-based method. Differences between the two classifications are discussed, emphasizing the importance of the interpretability of HRQoL measures.
Commentary

One essential property of a measurement instrument is the interpretability of its data. An understanding of the meaningfulness of a result is necessary for physicians in choosing treatments; for patients in understanding their conditions and their changes over time; and for policy makers in evaluating relationships among benefits, adverse effects, and cost. In the field of health-related quality of life (HRQoL), the problem of establishing meaningfulness is especially challenging. Although we easily understand, for example, the significance of an increase or a decrease of one degree Fahrenheit or Celsius in body temperature, it is not clear how to interpret a five-point change in a HRQoL scale measurement. In this issue, Prinsen et al. report scores for mildly, moderately, and severely impaired HRQoL for the three subscales—emotions, symptoms, and functioning—of the Skindex-29. This study completes a previous study, in which the authors identified cutoff points only for severely impaired HRQoL. To identify the categories, the authors used an anchor-based method that consists of comparing measures of HRQoL with other measures or phenomena that have relevance to patients. In particular, they used a cross-sectional method based on patients’ ratings of their HRQoL.

The categorization of the Skindex-29 was previously proposed by Nijsten et al. (2009), using mixture analysis (a distribution-based method) to obtain cutoff scores. This analysis assesses whether a distribution of a variable consists of different overlapping but independent “subdistributions” and categorizes the observations in different mixture components using posterior probabilities.

The categories obtained with the two approaches differ in several respects. In Table 1 we compare the categorization of data from studies performed by our group (and not previously used in the mixture analysis) using both sets of cutoffs. The first of each pair of columns, labeled “Psoriasis IMPROVE,” refers to data from the IMPROVE study of in-patients with psoriasis; the second of each pair, labeled “Survey 2010,” refers to data from a survey of more than 2,500 outpatients with various dermatological conditions. Mild or very mild HRQoL impairment, although reported in separate groups by Nijsten et al. (2009), are presented together by Prinsen et al. (2011, this issue). The discrepancies in the results are quite evident when dealing with a sample of all dermatological conditions (“Survey 2010”), several of which seldom highly impair patients’ HRQoL. Thus, in the Prinsen distribution the first category includes most cases. In addition, the range of some of the Prinsen classes is very narrow, so only a small percentage of patients fall into such categories. On the other hand, some Nijsten categories include a wide range of scores (the “mild” class of the emotions scale and the “moderate” class of the functioning scale, for example).

Our aim here is not to judge the correctness of the proposed classifications but to show how different methods may lead to different results. Both categorizations are valid and built according to well-established methods, but they may not represent the final say. In fact, to focus on a single example, it is difficult to reconcile results as discrepant as those seen for the symptoms scale in a sample of patients hospitalized for psoriasis. In this example, the upper limit of the “mild” category in the Prinsen classification is well above the lower limit of the “severe” category in the Nijsten classification; therefore, different decisions would be made if these categories were used to decide a patient’s eligibility for systemic treatment.
Table 1. Percentages of a sample of 936 patients hospitalized with psoriasis and a sample of 2,576 consecutive outpatients with various dermatological conditions according to the categorization of Skindex-29 by Prinsen et al. and Nijsten et al.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
<th>Symptoms</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
<th>Functioning</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff-Prinsen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>16.4</td>
<td>53.3</td>
<td>Mild</td>
<td>31.5</td>
<td>65.8</td>
<td>Mild</td>
<td>31.3</td>
<td>73.2</td>
</tr>
<tr>
<td>0–23.9</td>
<td>15.2</td>
<td>13.6</td>
<td>Moderate</td>
<td>4.2</td>
<td>4.0</td>
<td>Moderate</td>
<td>15.5</td>
<td>9.7</td>
</tr>
<tr>
<td>24–34.9</td>
<td>7.8</td>
<td>5.8</td>
<td>Severe</td>
<td>17.1</td>
<td>11.4</td>
<td>Severe</td>
<td>4.7</td>
<td>2.6</td>
</tr>
<tr>
<td>35–38.9</td>
<td>60.6</td>
<td>27.3</td>
<td>Very severe</td>
<td>47.2</td>
<td>18.8</td>
<td>Very severe</td>
<td>48.5</td>
<td>14.5</td>
</tr>
<tr>
<td>Cutoff-Nijsten</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very little</td>
<td>1.8</td>
<td>19.3</td>
<td>Very little</td>
<td>2.2</td>
<td>25.3</td>
<td>Very little</td>
<td>6.5</td>
<td>37.3</td>
</tr>
<tr>
<td>0–5.9</td>
<td>14.6</td>
<td>34.0</td>
<td>Mild</td>
<td>4.5</td>
<td>9.7</td>
<td>Mild</td>
<td>11.1</td>
<td>21.5</td>
</tr>
<tr>
<td>6–24.9</td>
<td>37.0</td>
<td>30.1</td>
<td>Moderate</td>
<td>10.6</td>
<td>18.4</td>
<td>Moderate</td>
<td>29.2</td>
<td>24.1</td>
</tr>
<tr>
<td>25–49.9</td>
<td>46.6</td>
<td>16.6</td>
<td>Severe</td>
<td>29.3</td>
<td>23.5</td>
<td>Severe</td>
<td>53.2</td>
<td>17.1</td>
</tr>
<tr>
<td>Severe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely severe</td>
<td>53.4</td>
<td>23.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretation of Skindex-29 scores
If “moderate to severe symptoms” was the criterion for systemic treatment in patients with psoriasis, 31.5% of patients hospitalized and followed in the IMPROVE study would not have been eligible according to the Prinsen categorization, compared with 6.7% according to Nijsten. Looking at a specific disease such as alopecia areata in the “Survey 2010” study, large discrepancies were observed (data not shown). For example, patients were classified with a mild HRQoL impairment in the functioning scale in 75.5% (Prinsen) versus 50.9% (Nijsten) of cases; similarly, patients were classified with a severe impact on emotions in 32.1% (Prinsen) versus 13.2% (Nijsten) of cases.

There is considerable debate over the advantages and disadvantages of anchor-based and distribution-based methods. Distribution-based approaches are based on the statistical characteristics of the sample and thus are sensitive to the homogeneity of the distribution of the sample from which they are derived. Anchor-based methods, which are based on patients’ ratings, are thought to provide the best measure of the significance of change from the individual’s perspective. However, anchor-based methods rely on global ratings, may vary on the basis of whether those anchors are collected prospectively or retrospectively, and may account for some of the variance in HRQoL scores. In addition, these methods rely heavily on the representativeness of the normative sample; this is even more true of anchor-based methods because the anchor points are often also dependent on subjective choices and sociocultural environments. In other words, how useful are such results when applied to sets of patients with a different case mix, in terms of both relative proportions of different skin conditions and clinical severity of disease within each condition?

Norman et al. (2001) conducted a simulation comparing the two methods and reported equivalent information, whereas Koloktin et al. (2002) reported comparable values for anchor-based and distribution-based methods in obesity-specific quality of life at moderate levels of impairment but markedly different values for those with severe and mild impairments. The debate is far from its conclusion. As suggested by Crosby et al. (2003), it would be useful to integrate information from the two approaches as some studies have attempted to do. In dermatology, it would be desirable to conduct analyses using both methods of categorization, provided that representative samples of the population with skin conditions are available. In any case, it is important to consider that a measurement instrument will be most useful when it is possible to interpret its results. And judgments about the usefulness of interpretive tools such as a given set of “categories of quality of life” will ultimately be based on the performance of such tools in the field, in daily clinical routines, and for different patient populations.

**CLINICAL IMPLICATIONS**

- “Categories of quality-of-life impairment” are necessary for applying quality-of-life data in clinical settings.
- Differences in methods of categorization and normative samples may yield different results.
- The usefulness of a given set of categories of quality-of-life impairment must be judged by the performance of such categories in the field.
REFERENCES


Skindex-29 (source questionnaire)
Skindex-29 (Nederlandse versie)
SKINDEX-29

English version
This survey concerns the skin condition which has bothered you the most during the past week.

Instruction
These questions concern your feelings over the past week about the skin condition that has bothered you the most.
Check the answer that comes closest to the way you have been feeling.
<table>
<thead>
<tr>
<th>Statement</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>All the time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My skin hurts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. My skin condition affects how well I sleep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I worry that my skin condition may be serious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. My skin condition makes it hard to work or do hobbies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. My skin condition affects my social life</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. My skin condition makes me feel depressed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. My skin condition burns or stings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I tend to stay at home because of my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I worry about getting scars from my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. My skin itches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. My skin condition affects how close I can be with those I love</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I am ashamed of my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I worry that my skin condition may get worse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. I tend to do things by myself because of my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I am angry about my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Water bothers my skin condition (bathing, washing hands)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. My skin condition makes showing affection difficult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. I worry about side-effects from skin medications / treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. My skin is irritated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. My skin condition affects my interactions with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. I am embarrassed by my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. My skin condition is a problem for the people I love</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. I am frustrated by my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. My skin is sensitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. My skin condition affects my desire to be with people</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. I am humiliated by my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. My skin condition bleeds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. I am annoyed by my skin condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. My skin condition interferes with my sex life</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. My skin condition makes me tired</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Skindex-29 (source questionnaire)
SKINDEX-29

Nederlandse versie

Deze vragenlijst gaat over de huidaandoening waarvan u de afgelopen week (het meeste) last heeft gehad.

Instructie

Deze vragen gaan over uw ervaringen in de afgelopen week.

Kruis per vraag één antwoordhokje aan dat het beste overeenkomt met hoe u zich hebt gevoeld. Als u twijfelt over het antwoord, geef dan het best mogelijke antwoord.

Slaat u alstublieft geen enkele vraag over.
Skindex-29 (Nederlandse versie)

Hoe vaak waren deze omschrijvingen in de afgelopen week op u van toepassing? Nooit Zelden Soms Vaak Altijd

1. Mijn huid doet pijn
2. Mijn huidaandoening beïnvloedt hoe ik slaap
3. Ik maak me zorgen dat mijn huidaandoening ernstig is
4. Door mijn huidaandoening is het moeilijk mijn werk of hobby’s te doen
5. Mijn huidaandoening beïnvloedt mijn sociale leven
6. Mijn huidaandoening maakt me depressief
7. Mijn huidaandoening is branderig of steekt
8. Ik ben geneigd om thuis te blijven door mijn huidaandoening
9. Ik maak me zorgen dat ik van mijn huidaandoening littekens kan krijgen
10. Mijn huid jeukt
11. Mijn huidaandoening belemmert mij intiem om te gaan met de mensen van wie ik hou
12. Ik schaam me voor mijn huidaandoening
13. Ik maak me zorgen dat mijn huidaandoening kan verergeren
14. Ik ben geneigd om dingen in mijn eentje te doen vanwege mijn huidaandoening
15. Mijn huidaandoening maakt mij boos
16. Water irriteert mijn huidaandoening (baden, douchen, handen wassen)
17. Door mijn huidaandoening is het moeilijk geneogenheid of affectie te tonen
18. Ik maak me zorgen over bijwerkingen van medicijnen en/of de behandeling die ik voor mijn huid krijg
19. Mijn huid is geïrriteerd
20. Mijn huidaandoening beïnvloedt mijn contacten met anderen
21. Ik voel me opgelaten en ongemakkelijk door mijn huidaandoening
22. Mijn huidaandoening is een probleem voor de mensen van wie ik houd
23. Ik voel me gefrustreerd door mijn huidaandoening
24. Mijn huid is gevoelig
25. Mijn huidaandoening beïnvloedt mijn verlangen om samen met anderen te zijn
26. Ik voel me vernederd door mijn huidaandoening
27. Mijn huidaandoening bloedt
28. Mijn huidaandoening ergert me
29. Mijn huidaandoening belemmert mijn seksuele leven
30. Mijn huidaandoening maakt mij moe
List of contributing authors
List of contributing authors

Prof. dr. Matthias Augustin  
Department of Dermatology, University Clinics of Hamburg, German Center for Health Services Research in Dermatology, Hamburg, Germany

Dr. Mohammad K.A. Basra  
Department of Dermatology and Wound Healing, Cardiff University School of Medicine, Cardiff, UK

Ms. Oda D. van Cranenburgh  
National Skin Foundation, Utrecht, The Netherlands

Dr. Elisabeth A. Holm  
Dermatology Clinic Ballerup, Healthhouse, Ballerup, Denmark

Dr. John de Korte  
Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Ms. Catharina M. (Wendy) Legierse  
Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Dr. Robert Lindeboom  
Division of Clinical Methods and Public Health, Master Evidence Based Practice, University of Amsterdam, Amsterdam, The Netherlands

Prof. dr. Tamar E.C. Nijsten  
Department of Dermatology, Erasmus University Medical Center, Rotterdam, The Netherlands

Prof. dr. Menno A. de Rie  
Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Prof. dr. M.S. (Sam) Salek  
Centre for Socioeconomic Research, Welsh School of Pharmacy, Cardiff University, Cardiff, UK

Dr. Francesca Sampogna  
Health Services Research Unit, Istituto Dermopatico dell’Immacolata (IDI) IRCCS, Rome, Italy

Prof. dr. Mirjam A.G. Sprangers  
Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Dr. Phyllis I. Spuls  
Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
Acknowledgements / Dankwoord
Dit proefschrift was niet tot stand gekomen zonder de inzet van vele dermatologen, (onderzoeks)verpleegkundigen en patiënten die bereid waren mee te werken aan de verschillende onderzoeken. Ook de samenwerking met en de support van vele personen in de totstandkoming van dit proefschrift waren onmisbaar. Aan hen ben ik mijn dank verschuldigd. Een aantal personen wil ik graag in het bijzonder bedanken:

Prof.dr. M.A. de Rie, mijn promotor. Beste professor De Rie, hartelijk dank voor uw bereidheid om zitting te nemen in de promotiebegeleidingscommissie toen de trein al rijdende was. Het was prettig om te weten dat uw deur altijd open stond zodat ik met vragen bij u terecht kon. Daarnaast wil ik u bedanken voor de mogelijkheid die u heeft geboden om ook de laatste maanden van mijn promotieonderzoek te kunnen afronden op de afdeling.

Prof.dr. M.A.G. Sprangers, mijn promotor. Beste Mirjam, dank je wel voor je begeleiding en betrokkenheid in de totstandkoming van dit proefschrift. Je immer snelle en kritische feedback als ook je onmisbare kennis op het gebied quality of life research zorgde ervoor dat de manuscripten konden worden aangescherpt. Het is fijn om te kunnen leren van iemand met zo veel kennis op dat gebied.

Dr. J. de Korte, mijn co-promotor. Beste John, dank je wel voor de leerzame jaren als onderzoeker-in-opleiding. Jouw grote mate van zorgvuldigheid en kritische commentaren hielden mij scherp en zorgden ervoor dat we er altijd kwalitatief goede ‘chocola van konden maken’. Dank ook voor de kansen die je me hebt geboden waardoor ik mijzelf verder heb kunnen ontwikkelen als wetenschapper binnen de (psycho)dermatologie, zowel op nationaal als internationaal niveau. Dat is bijzonder waardevol voor me geweest.

Dr. Ph.I. Spuls, mijn co-promotor. Beste Phyllis, dank voor het vertrouwen en de plezierige wijze waarop je me hebt begeleid. Jouw inzichten hielpen mij om de balans te vinden tussen wetenschap en klinische praktijk. Daarnaast zijn jouw ambitie en gedrevenheid een inspiratie voor me. De gezellige trialteam-etentjes bij jou thuis en je gastvrijheid zullen me zeker blijven. We delen de interesse voor dermato-epidemiologie, evidence based medicine en outcomes research –heel leuk dat we onze samenwerking kunnen voort zetten binnen de HOME en COMET Initiatives.

Dr. R. Lindeboom, mijn co-promotor. Beste Robert, jouw brede kennis van klinimetrie en statistiek, je creativiteit en ideeën, je inhoudelijk zeer waardevolle feedback en de prettige samenwerking waren onmisbaar. Je leerde mij de ins en outs kennen van item response theory en multilevel analyses tijdens de vele uren die we doorbrachten achter jouw pc. Tevens leerde je mij het overbrengen van een korte heldere boodschap: minder is beter. Heel veel dank voor je begeleiding.

Mijn paranimfen. Lieve Marian, wat ben ik blij dat jij naast me staat tijdens de verdediging. Als studiegenootjes leerden we elkaar kennen en al snel bleek dat we vaak aan één woord al genoeg hadden om elkaar te begrijpen. Hieruit groeide een bijzondere vriendschap. Je nimmer aflatende interesse in de voortgang van mijn promotieonderzoek, je fijngevoeligheid voor mijn persoonlijke kwaliteit-van-leven en je luisterend oor, hulp en advies op momenten dat ik dat nodig had. Dank dank dank! Lieve Frank, mijn woordje tot jou bewaar ik voor het laatst.
De leden van de promotiecommissie. Prof.dr. N.K. Aaronson, Prof.dr. A.W.M. Evers, Prof.dr. R.J. de Haan, Prof.dr. T.E.C. Nijsten en Dr. A. Wolkerstorfer. U allen wil ik hartelijk danken voor de beoordeling van het manuscript en uw bereidheid zitting te nemen in de promotiecommissie.

Prof.dr. J.D. Bos. U heeft mij de mogelijkheid gegeven om promotieonderzoek uit te voeren op de afdeling Dermatologie van het AMC. Hiervoor ben ik u zeer erkentelijk.

Alle dermatologen en (onderzoeks)verpleegkundigen uit de verschillende dermatologische centra met wie ik heb mogen samenwerken gedurende de afgelopen vier jaar. Heel hartelijk dank voor het includeren van patiënten tijdens de vaak toch al drukke spreekuren, voor de tijd die u heeft vrijgemaakt voor het aanleveren van studie data, de gastvrijheid op uw afdeling en natuurlijk voor de enorm prettige samenwerking. Zonder uw medewerking zou het proefschrift er niet in deze vorm zijn geweest. Dank!

Alle patiënten die vrijwillig aan de verschillende onderzoeken hebben meegewerkt, ontzettend bedankt!

Collegae van het Huidfonds. Anne-Marie, veel dank voor het begeleiden van de Q-ACT studiecentra in de uitvoering van het onderzoek en het monitoren van de studiedata. Met 18 studiecentra was ik blij dat we het werk samen konden doen. Oda, dank voor het stand-by staan als wij om logistieke redenen onze studiepatiënten in het AMC niet zelf konden zien. En Sabrina, dank je wel voor het invoeren van een grote hoeveelheid studiedata in de database. Het was een enorme klus en ik was blij dat ik daar niet alleen voor stond.

Trialteam collegae van de afdeling Dermatologie. Anna-Christa, Dominique, Evelien, Gabriëlle, Jascha, Job, Lidian, Mandy, Marijke, Marleen en Stef. De lunches bij de psych of buiten op het bankje, de vele wandelingetjes rond het AMC, de trialteam etentjes en congresreisjes naar het buitenland; dank jullie wel voor de ontzettend leuke tijd in het trialteam.

Collegae van de afdeling Dermatologie. Dank jullie wel voor de samenwerking en de leuke tijd tijdens mijn promotieonderzoek. De wintersportreis naar Jochberg in 2012 was uniek. Mijn dank gaat ook uit naar alle arts-assistenten die hebben bijgedragen aan de inclusie van patiënten voor de verschillende studies op de algemene spreekuren en op het biologic spreekuur.

Wendy, onze eerste werkdag samen begon in één van de kleinste kamertjes op de kliniek. Na een jaar ging jij in opleiding en niet veel later startte ik met mijn promotieonderzoek. Dank voor al je interesse, al die tijd.

May, Bas en Marjan, jullie wil ik danken voor alle ‘tips and tricks’ in de laatste fase van het promotietraject. Van het aanvragen van offertes van drukkerijen tot aan de organisatie rondom de promotie zelf; jullie advies hielp me op weg.

Robert, Marja en Mariska, dank jullie wel voor alle andere zaken buiten het onderzoek om dat ervoor zorgde dat ik mijn werk altijd goed kon uitvoeren.

Collegae van de afdeling Medische Psychologie. Hartelijk dank voor de samenwerking. De feedback op alle leerzame ‘research issues’ en de proefpromotie waren zeer waardevol.

Collegae van de Clinical Research Unit. Miranda en Irmgard, hartelijk dank voor de prettige samenwerking in de ontwikkeling van de Q-ACT database en het overzetten van het Access databestand naar werkbare SPSS bestanden zodat ik de data kon analyseren. Vaak had ik mijn vraag nog niet gesteld of ik had al een antwoord in mijn mailbox. Super.

Mijn nieuwe collegae van het VUmc. Caroline en Riekie, hartelijk dank voor de flexibiliteit die jullie mij de afgelopen maanden hebben geboden zodat ik naar het AMC kon pendelen om te werken aan de afronding van mijn proefschrift. Dit heb ik bijzonder gewaardeerd. Ik vind het daarnaast een eer dat ik deel uit maak van jullie team; jullie enthousiasme voor klinimetrie werkt aanstekelijk en is voor mij een bron van inspiratie. Carlijn, wat leuk dat we kamergenootjes zijn! Beiden hebben we ons ingewerkt in een nieuwe functie en beiden waren we daarnaast druk met de afronding van ons promotieonderzoek. Onder het genot van een stuk chocola en een kop thee konden we alles vaak wel weer relativeren.

Mevrouw Musaph, Monique Broekman en Jan van Hees. Veel dank voor de toestemming die u heeft gegeven voor het opnemen van de afbeelding van het kunstwerk in mijn proefschrift. De symboliek erachter betekent veel voor me.

Grietje, dank je wel voor de plezierige samenwerking de afgelopen jaren. En natuurlijk heel veel dank voor jullie support!

Offpage. Aleksandra, Dorothea and Ewa, thank you very much for your help and support in the lay-out and printing of my PhD thesis.

Alle sponsors. Hartelijk dank voor de financiële ondersteuning wat het drukken van dit proefschrift mogelijk maakte.

Ten slotte wil ik nog een aantal dierbare mensen in mijn leven bedanken voor hun support gedurende de afgelopen jaren vanaf de zijlijn.

Mijn vriendinnen. Dank voor alle gezelligheid tussen werk en promotieonderzoek doen in, hierdoor kon ik mijn onderzoeksstress even laten voor wat het was. Maar ook excuses voor de keren dat ik verstek liet gaan en onze afspraak soms op het laatste moment moest afzeggen omdat mijn werk nog niet af was. Dank jullie wel voor alle begrip hiervoor. Het is fijn om zoveel lieve vriendinnen om me heen te hebben; jullie zijn me enorm dierbaar. Lieve Brigitte, jou wil ik in het bijzonder bedanken. Jij kent me als geen ander en weet wat er in me omgaat. Dankjewel voor al je support! De afgelopen jaren zijn in meerdere opzichten speciaal en bijzonder geweest. Fijn dat we op die momenten zo dichtbij elkaar konden staan.

Mijn familie en schoonfamilie. Lieve pap & mam, 100 keer dank-je-wel zeggen is lang niet genoeg om mijn grote dank aan jullie te betuigen. Dankzij jullie never-ending support en betrokkenheid konden de dingen die tegelijk kwamen naast elkaar plaatsvinden. Ik hou van jullie. Lieve Jorn, Merijn & Iris, ik geniet van het samen zijn en de leukle dingen die we doen. Die frequentie kan wat mij betreft nu omhoog! Lieve Cor & Marion, Melcher, Arthur & Marijke en Marlous & Patrick, fijn om zo’n lieve schoonfamilie te hebben.
Frank, mijn liefste lief. Jou wil ik danken om wie je bent. Jouw rust en geduld, je eindeloze vertrouwen in mij, je goede zorgen en je praktische hulp tijdens de allerlaatste loodjes die echt het zwaarste wogen, vormden voor mij de stevige basis waarop ik altijd kon terugvallen. Promoveren is heel mooi, maar uiteindelijk zijn het andere dingen die het leven bijzonder maken. Ik kijk dan ook enorm uit naar de (vrije) tijd die voor ons ligt, samen met onze prachtige dochter Jade. Ik hou van je.
Bibliography
International publications


National publications


Published abstracts


Submitted articles
CAC Prinsen, J de Korte, Phl Spuls, MAG Sprangers, MA de Rie, R Lindeboom. Determining the meaningfulness of differences in Skindex-29 scores using item response theory modeling. Submitted

CAC Prinsen, Phl Spuls, R Lindeboom, MAG Sprangers, MA de Rie, J de Korte. The efficacy of a health-related quality-of-life intervention during 48 weeks of biologic treatment of patients with moderate to severe psoriasis: results of a multicenter randomized controlled trial. Submitted

International conferences
6th International Congress on Dermato-Epidemiology Association Congress (IDEA), Malmö, Sweden (2012) CAC Prinsen. Interpretation of Skindex-29 scores using Rasch analysis.

14th Congress of the European Society for Dermatology and Psychiatry (ESDaP), Zaragoza, Spain (2011) CAC Prinsen. Severe impairment of health-related quality of life: patients at risk.


National conferences

Lustrum Congress Master Study Evidence Based Practice, University of Amsterdam, Amsterdam, The Netherlands (2012) CAC Prinsen. Een onderhuids gevoel bij diagnostische criteria en uitkomstparameters.


International poster presentations
14th Congress of the European Society for Dermatology and Psychiatry (ESDaP), Zaragoza, Spain (2011) CAC Prinsen. Identification of patients with mild, moderate, and severe impairment of health-related quality of life.
Portfolio
# AMC GRADUATE SCHOOL FOR MEDICAL SCIENCES PHD PORTFOLIO
## Summary of PhD training, teaching and parameters of esteem

<table>
<thead>
<tr>
<th>Name:</th>
<th>C.A.C. (Sanna) Prinsen</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhD period:</td>
<td>March 2009 – December 2013</td>
</tr>
<tr>
<td>Promotores:</td>
<td>Prof. dr. M.A. de Rie, Prof. dr. M.A.G. Sprangers</td>
</tr>
<tr>
<td>Co-promotores:</td>
<td>Dr. J. de Korte, Dr. Ph.I. Spuls, Dr. R. Lindeboom</td>
</tr>
<tr>
<td>Institution:</td>
<td>Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands</td>
</tr>
</tbody>
</table>

## PhD training

<table>
<thead>
<tr>
<th>Year</th>
<th>Workload (ECTS)</th>
<th>Course Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>General courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Career Development</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Scientific Writing in English for Publication</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Oral Presentation in English</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Systematic Reviews</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Clinical Data Management (CRF design, database design and data entry screens, data entry and data validation, data transformation, merging and data export)</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Practical Biostatistics</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Expert Management of Medical Literature</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Qualitative Health Research</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Basic Course Legislation and Organization for Clinical Researchers (BROK) (Good Clinical Practice (GCP), project organization and data management, medical ethics, the role of the Medical Ethical Committee, methodological aspects of medical research, laboratory, pharmacy, legal aspects in medical research)</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>The AMC World of Science: Fundamental Knowledge and Skills for Scientific Research in Preparing the PhD Thesis</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>Specific courses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Advanced Biostatistics (the R statistical package, main statistical concepts, visual display, regression models, prediction, missing values, survival analysis)</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>Advanced Topics in Clinical Epidemiology (evaluation of medical tests, issues in clinical trials, prognostic and predictive models, screening, causality)</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>Seminars, workshops and master classes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Master Class by Professor G.H. Guyatt</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>Master Class by Dr. D. Feldman-Stewart</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>Master Class by Professor J.A. Sloan</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>(Inter)national oral presentations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dutch Society of Experimental Dermatology (NVED), Ede, The Netherlands. Interpretation of Skindex-29 score differences using item response theory modelling.</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>6th International Congress on Dermato-Epidemiology Association Congress (IDEA), Malmö, Sweden. Interpretation of Skindex-29 scores using Rasch analysis.</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>Lustrum Congress Master Study Evidence Based Practice, University of Amsterdam, Amsterdam, The Netherlands. Een onderhuids gevoel bij diagnostische criteria en uitkomstparameters.</td>
<td>2012</td>
</tr>
</tbody>
</table>
14th Congress of the European Society for Dermatology and Psychiatry (ESDaP), Zaragoza, Spain. Severe impairment of health-related quality of life: patients at risk. 2011 0.5

18th Congress of the European Academy of Dermatology and Venereology (EADV), Berlin, Germany. Health-related quality of life assessment in dermatology: interpretation of Skindex-29 scores using patient-based anchors. 2009 0.5

10e dermatologisch verpleegkundig congres, V&VN, Ede, The Netherlands. Kwaliteit van Leven in de Dermatologische Praktijk. 2009 0.5

International poster presentations
14th Congress of the European Society for Dermatology and Psychiatry (ESDaP), Zaragoza, Spain. Identification of patients with mild, moderate, and severe impairment of health-related quality of life. 2011 0.5

(International) national conferences
Dutch Society of Experimental Dermatology (NVED), Ede, The Netherlands 2013 0.25
Dutch Society of Quality of Life Research (ISOQOL-NL), Utrecht (2008), Amsterdam (2010), Utrecht (2012), Tilburg (2013), The Netherlands 2008-2013 1.0
International Dermato-Epidemiology Association Congress - European Dermato-Epidemiology Network (IDEA-EDEN), Nottingham (UK), Malmö (Sweden) 2008, 2012 0.5
European Society for Dermatology and Psychiatry (ESDaP), Zaragoza (Spain) 2011 0.25
International Society for Quality of Life Research (ISOQOL), London (UK) 2010 0.25
European Academy of Dermatology and Venereology (EADV), Paris (France, 2008), Berlin (Germany, 2009), Göteborg (Sweden, 2010) 2008-2010 0.75
Dutch Society of Psychodermatology (NVPD), Utrecht (2008), Nijmegen (2009), Utrecht (2010), The Netherlands 2008-2010 0.75

Other
Postdoc Career Development Initiative (PCDI) Retreat 2012 1.0

II Teaching
Year Workload (ECTS)

Lecturing
University Medical Center Utrecht, Utrecht, The Netherlands. Kwaliteit van Leven in de Dermatologische Praktijk. 2008-2012 5.0
Master study Evidence Based Practice, University of Amsterdam, Amsterdam, The Netherlands. Master Class Health-related quality of life assessment in dermatology: interpretation of Skindex-29 scores using patient-based anchors. 2010 1.0

Supervising
Sandra Nijland, master thesis Evidence Based Practice 2012-2013 1.0
Hélène Fortuin, bachelor thesis Huidtherapie 2012 1.0
Leonore Westerhoud, bachelor thesis Huidtherapie 2012 1.0
Josine de Bes, master thesis Medicine 2010-2011 1.0

Other
Hogeschool Utrecht, Huidtherapie, Utrecht, The Netherlands. Guest lecture Kwalitatief onderzoek in de gezondheidszorg. 2012 0.5
Isala Clinics, Zwolle, The Netherlands. Guest lecture Kwaliteit van Leven Assessment in de Dermatologische Praktijk. 2009 0.5
Psoriasis Vereniging Nederland, Rotterdam, The Netherlands. Guest lecture Leven met een Chronische Huidziekte – Kwaliteit van Leven bij Psoriasis. 2008 0.5
## III Parameters of Esteem

<table>
<thead>
<tr>
<th>Grants</th>
<th>Year</th>
<th>Workload (ECTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant for Health-related quality of life assessment and communication during 48 weeks of treatment of moderate to severe psoriasis – a randomized controlled clinical trial. (Principal investigator: J de Korte, PhD)</td>
<td>2009-2012</td>
<td>-</td>
</tr>
<tr>
<td>Grant for Vragenlijstonderzoek behandeltevredenheid en kwaliteit van leven bij lichen planus en lichen sclerosus. (Principal investigator: J de Korte, PhD)</td>
<td>2011-2012</td>
<td>-</td>
</tr>
<tr>
<td>Awards and prizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The paper “Health-related quality of life assessment in dermatology: interpretation of Skindex-29 scores using patient-based anchors” received the Herman Musaph Literature Prize 2010 by the Dutch Society of Psychodermatology (NVPD).</td>
<td>2010</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Member of the Harmonising Outcome Measures for Eczema (HOME) Initiative</td>
<td>2013</td>
<td>1.0</td>
</tr>
<tr>
<td>Member of the European Academy of Dermatology and Venereology (EADV) Taskforce on Quality of Life</td>
<td>2009-</td>
<td>1.0</td>
</tr>
<tr>
<td>Peer reviewer for several international scientific medical journals</td>
<td>2010-2017</td>
<td>1.0</td>
</tr>
<tr>
<td>Member of the ISOQOL New Investigators Special Interest Group</td>
<td>2011</td>
<td>1.0</td>
</tr>
<tr>
<td>Co-chair at the European Society for Dermatology and Psychiatry (ESDaP) congress, Zaragoza, Spain. Symposium 4, Plenary Room, Long Term Issues.</td>
<td>2011</td>
<td>1.0</td>
</tr>
<tr>
<td>Symposium organized: Kwaliteit-van-leven assessment in de dermatologische praktijk. Stichting Aquamarin, Zeist, The Netherlands</td>
<td>2008</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Curriculum Vitae
Cecilia Anna Catharina (Sanna) Prinsen was born on July 29th, 1979 in Amersfoort, the Netherlands. After graduating from high school in 1997, she completed her Bachelor’s degree in Nursing (2001). Subsequently, she worked as a clinical research associate at the Medical Department of a pharmaceutical company (2002-2005), and as a clinical project manager at a Clinical Research Organization (2005-2007). During her career, Sanna expanded her knowledge and skills in clinical research; she graduated from the University of Amsterdam and acquired her Master’s degree in Clinical Epidemiology (2009). She wrote her Master’s thesis at the National Skin Foundation, Department of Dermatology, at the Academic Medical Center, University of Amsterdam (AMC-UvA) where she investigated the interpretation of quality-of-life scores as measured with the Skindex-29. This publication received the Herman Musaph Literature Prize 2010 by the Dutch Society of Psychodermatology and founded the basis of her PhD thesis. As of March 2009, Sanna started working as a PhD student at the Department of Dermatology at the AMC-UvA. The research performed during this period is presented in this thesis. From May through December 2012, Sanna also worked as a part-time employee of the Dutch Society of Dermatology and Venereology (NVDV) where she was involved in developing and updating treatment guidelines for several dermatological conditions. Sanna is a member of the European Academy of Dermatology and Venereology (EADV) Taskforce on Quality of Life, and the Harmonising Outcome Measures for Eczema (HOME) Quality of Life Research Group. She acted as a peer reviewer for several international scientific medical journals. In addition, Sanna supervised bachelor and master students with their theses and she taught the course “Quality of Life Assessment in Dermatology” to nursing students specializing in dermatology at the University Medical Center Utrecht. As of January 2013, Sanna works as a Post Doctoral researcher at the Department of Epidemiology and Biostatistics at the EMGO+ Institute, VU University Medical Center, Amsterdam, The Netherlands, where she is involved in the Core Outcome Measures in Effectiveness Trials (COMET) Initiative.
The Herman Musaph Literature Prize object is created by Monique Broekman (www.puckworks.nl) and Jan van Hees (www.atelierjanvanhees.nl) – The portrait of Herman Musaph and a woman’s back with vitiligo symbolizes the relationship between the skin and the brain.