

SECTION EDITOR: MICHAEL BIGBY, MD; ASSISTANT SECTION EDITORS: OLIVIER CHOSIDOW, MD, PhD;
ROBERT P. DELLAVALLE, MD, PhD, MSPH; DAIHUNG DO, MD; URBÁ GONZÁLEZ, MD, PhD;
CATALIN M. POPESCU, MD, PhD; HYWEL WILLIAMS, MSc, PhD, FRCP

ONLINE FIRST

Measurement Properties of Outcome Measures for Vitiligo

A Systematic Review

Charlotte Vrijman, MD; May W. Linthorst Homan, MD; Jacqueline Limpens, PhD; Wietze van der Veen, MD, PhD;
Albert Wolkerstorfer, MD, PhD; Caroline B. Terwee, PhD; Phyllis I. Spuls, MD, PhD

Objective: To summarize and critically appraise the evidence on the measurement properties of clinician-, patient-, and observer-reported outcomes, measuring any construct of interest in patients with all types of vitiligo.

Data Sources: Electronic databases including PubMed (1948 to July 2011), OVID EMBASE (1980 to July 2011), and CINAHL (EBSCOhost) (1982 to July 2011) were searched.

Study Selection: Two authors independently screened all records for eligibility. For inclusion, the study population had to include patients with vitiligo, for which outcome measures were developed or evaluated on their measurement properties. The initial search retrieved 1249 records, of which 14 articles met the inclusion criteria.

Data Extraction: Characteristics of the included instruments, study population, and results of the measurement properties were extracted. The Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) 4-point checklist, combined with quality criteria for measurement properties,

was used to calculate the overall level of evidence per measurement property of each instrument. Independent extraction and assessment was performed by 2 authors.

Data Synthesis: Eleven different measurement instruments were identified. Strong evidence was found for a positive internal consistency of the Dermatology Life Quality Index. For other instruments, the evidence of measurement properties was limited or unknown.

Conclusions: Recommendations on the use of specific outcome measures for vitiligo should be formulated with caution because current evidence is insufficient owing to a low number of studies with poor methodological quality and unclear clinical relevance. To recommend outcome measures for vitiligo, further research on measurement properties of clinical relevant outcome measures for vitiligo according to COSMIN quality criteria is needed.

Arch Dermatol. 2012;148(11):1302-1310.

Published online September 17, 2012.

doi:10.1001/archdermatol.2012.3065

Author Affiliations: Department of Dermatology and Netherlands Institute for Pigment Disorders (Drs Vrijman, Linthorst Homan, van der Veen, Wolkerstorfer, and Spuls), Medical Library and Dutch Cochrane Center (Dr Limpens), Academic Medical Center, University of Amsterdam; The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital (Dr van der Veen); Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Centre (Dr Terwee); and Department of Dermatology, Onze Lieve Vrouwe Gasthuis (Dr Vrijman); Amsterdam, the Netherlands.

GOOD OUTCOME MEASURES are a necessity in evidence-based medicine.^{1,2} It may be important to differ between outcomes used for clinical practice and outcomes used for research, since they have a different aim. For clinical practice, outcome measures especially need to have clinical and biological significance to evaluate disease severity or treatment effect within patients. For research, outcome measures need to be clearly defined and characterized by good measurement properties to objectify disease severity or efficacy of therapies within and between patients.¹ Besides, consensus on the use of outcome measures needs to be achieved to make comparisons between different studies feasible.^{3,4} It is crucial to know

the quality of the current existing measures, and improving the indicated imperfections in outcome measures will lead to better decision making in patient care.

Research of measurement properties in dermatology is in its infancy. To date, most research on outcome measures has been done for psoriasis and atopic dermatitis, although a validated tool for evaluating the

See Editor's Note at end of article

methodological quality was lacking.⁵⁻⁸ Recently, the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN), an instrument to evaluate the methodological quality of studies on measurement properties,

has been developed.^{2,9,10} With COSMIN it is now possible to critically appraise and compare the quality of these studies.

For vitiligo, many different outcome measures are used. In a systematic review of randomized controlled trials of interventions for vitiligo up to 2009, approximately 90 different outcome measures were identified.¹¹ These outcome measures are poorly standardized and lack consensus and uniformity. The consensus report from the Vitiligo European Task Force in 2007 showed that besides the current lack of consensus on definition, there is also a lack of uniformity in methods of assessment, which makes it impractical to compare the outcomes of different studies.¹²

The aim of this systematic review was to summarize and critically appraise the evidence on the measurement properties of clinician-, patient-, and observer-reported outcomes, developed or used in studies of patients with vitiligo.

METHODS

SEARCH STRATEGY

A medical librarian (J.L.) with experience in conducting searches for systematic reviews undertook a systematic search of the electronic databases PubMed 1948 to July 2011), OVID EMBASE (1980 to July 2011), and CINAHL (EBSCOhost) (1982 to July 2011) to identify studies on measurement properties for scoring vitiligo. PubMed was chosen rather than OVID MEDLINE because the search filter for finding studies on measurement properties of measurement instruments was developed and validated in this database.¹³

The search in PubMed and OVID EMBASE consisted of Subject Headings (MeSH, SH) and free-text words for vitiligo, a special search filter for finding studies on measurement properties of measurement instruments (combined with the Boolean operator "AND") was used. This validated, sensitive, special search filter was used for PubMed and a provisional translation of this special search filter was used for EMBASE. Animal studies were excluded by double negation, ie, by using NOT ("animals" [MeSH Terms] NOT "humans" [MeSH Terms]). In CINAHL, we searched only for vitiligo in titles and abstracts. Reference Manager software (version 12.0; Thomson Reuters) was used to manage and deduplicate all identified references.

ELIGIBILITY CRITERIA

Inclusion Criteria

For inclusion, the study population had to include patients with at least 1 type of vitiligo, ie, segmental or nonsegmental vitiligo. If only a part of the study population included patients with vitiligo, the study was considered eligible as well. The instruments under study needed to be all clinician-, patient-, or observer-reported outcome measures. The aim of the included study had to be the development of a measurement instrument or the evaluation of 1 or more of its measurement properties.

Studies that focused only on interpretability, eg, the determination of minimal important change, were also considered to be eligible. The study had to be published as a full-text original article, and no restrictions regarding languages were made.

Exclusion Criteria

Studies in which the measurement instruments were used as an end point without studying the measurement properties were not considered eligible. Studies in which 2 measurement instruments were compared with each other without a validation of 1 explicit instrument were also considered not eligible because it is often unclear which of the 2 compared instruments is validated in these studies. Studies in which the instrument of interest was used in the validation process of another instrument, randomized controlled trials, or other longitudinal studies in which indirect evidence for responsiveness might be found were excluded because it is often difficult to interpret the evidence for validity or responsiveness provided in these studies, since no hypotheses about the validity or responsiveness of the instrument of interest are formulated and tested in these studies.

SELECTION

Two authors (C.V. and M.L.H.) independently screened initially all records on titles and abstracts in which a study was excluded if it did not refer to a study population of patients with vitiligo or did not study measurement properties. Full-text copies of the selected articles were obtained, and the reference lists of these articles were also screened for eligibility by both authors. The final selection was based on the assessment of the full-text articles by 2 independent authors (C.V. and M.L.H.) as well. A study was included if the previously described inclusion criteria were met. Disagreements were solved by discussion. If necessary, a third independent author (P.I.S.) was consulted to reach consensus.

EVALUATION OF METHODOLOGICAL QUALITY OF THE INCLUDED STUDIES

The methodological quality analysis of the included studies was performed using the COSMIN 4-point checklist.⁹ The measurement properties were divided over the following 3 domains: reliability, validity, and responsiveness (eTable 1; <http://www.archdermatol.com>).^{10,14} The COSMIN checklist consists of 9 boxes with 5 to 18 items concerning methodological standards on how each measurement property should be assessed. Except sample size, each item was scored on a 4-point rating scale (ie, "poor," "fair," "good," "excellent"). An overall methodological quality score of a study was determined for each measurement property separately by taking the lowest rating of any item in a box ("worst score counts"). Methodological quality scores for different studies were rated separately and were not combined.

DATA EXTRACTION

The following information was extracted from the included articles by 2 authors independently (C.V. and M.L.H.): (1) characteristics of the included instruments, (2) characteristics of the included study population in which the measurement properties were assessed (number of patients, percentage of patients with vitiligo in the study population, mean age, sex, setting, country, and language) and (3) results of the measurement properties. Interpretability was extracted by information on floor or ceiling effects or minimal important change or minimal important difference.¹⁵⁻¹⁷ Because there is no consensus of core domains for outcomes measures for vitiligo, the US Food and Drug Administration (FDA) qualification of clinical outcome assessments was used to categorize the outcome measures.¹⁸ This FDA qualification makes a differentiation of clinician-reported outcomes (defined as outcomes assessed or reported by physician or other cli-

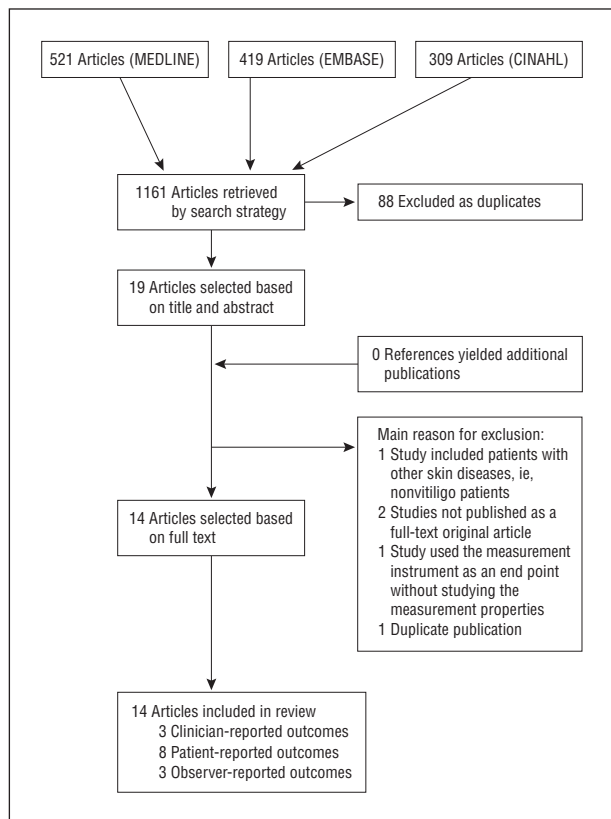


Figure. Flowchart of the search and selection process.

nician), patient-reported outcomes (defined as outcomes assessed or reported by patients), and observer-reported outcomes (defined as outcomes reported by other observers than clinicians or patients, for instance assessed by a computer).

DATA SYNTHESIS—QUALITY ASSESSMENT

Besides the methodological quality of the studies, the quality of the measurement properties of included instruments was evaluated as well. The results of the measurement properties were rated with the quality criteria for measurement properties (eTable 2).¹⁹ The overall evidence for separate measurement properties of the instruments was determined by combining the results of different studies adjusted for their methodological quality. The level of evidence was rated as “strong,” “moderate,” “limited,” “conflicting,” or “unknown,” as proposed by the Cochrane Back Review Group, with an additional weight for sample size because this was not used in the evaluation of the methodological quality of the studies (eTable 3).²⁰

RESULTS

ELECTRONIC LITERATURE SEARCH RESULTS

The **Figure** summarizes the selection process for all studies. The initial search retrieved 1249 records from the 3 databases searched. After exclusion of duplicates, 1161 articles remained. After screening for eligibility on title and abstract, 19 records were selected, among which no references yielded additional publications. Fifteen full-text articles that met the inclusion criteria were identified. One article was additionally excluded because this article²¹ was

a duplicate of an earlier published original article.²² Therefore, 14 articles remained for analysis. The main reasons for exclusion in the full-text selection are represented in the Figure. No systematic reviews about measurement properties of scorings systems for vitiligo were found.

DESCRIPTION OF THE INCLUDED STUDIES

The 14 articles evaluating the measurement properties of 11 different instruments were published between 1996 and 2010. Characteristics of the included instruments are given in **Table 1**, and characteristics of the included study populations are given in **Table 2**. In 4 studies the study population consisted of patients with vitiligo as well as patients with other skin diseases, whereas 10 studies had a study population consisting solely of patients with vitiligo.

The methodological quality according to the COSMIN checklist is summarized in **Table 3**. The results of the measurement properties are given in eTable 4, and **Table 4** summarizes the evidence on the measurement properties of the different instruments. The instruments that were evaluated are described in the following subsections.

EVALUATED INSTRUMENTS FOR VITILIGO

Clinician-Reported Outcomes

Vitiligo European Task Force (VETF) Assessment. The VETF assessment was evaluated on measurement error. The evidence was rated as unknown because there was only 1 study¹² with poor methodological quality. No information on the interpretability was found, since there was no information reported regarding floor and ceiling effects, minimal important change, or minimal important difference.

Vitiligo Area Scoring Index (VASI). Criterion validity is the only evaluated measurement property of the VASI. The evidence for the criterion validity was rated as unknown because only 1 study²³ with poor methodological quality was found. No information on the interpretability was found.

Point Counting. There is limited evidence for a positive reliability of the point-counting method because 1 study²⁴ with fair methodological quality showed a correlation coefficient greater than 0.70. No other measurement properties were evaluated. No information on the interpretability was reported.

Patient-Reported Outcomes

Skindex-29. For the Skindex-29, the internal consistency and criterion validity were evaluated. For both measurement properties the evidence was rated as unknown because for both properties 1 study was found with poor methodological quality.^{25,26} No information regarding floor or ceiling effects, minimal important change, or minimal important difference was found.

Skindex-16. The internal consistency, content validity, and construct validity were evaluated by 1 study.²⁷ For

Table 1. Characteristics of the Included Study Instruments

Instrument	Measure of Interest	Source	Description
Clinician-reported outcomes			
VETF assessment	Extent of disease; stage of disease; spreading of disease	Taieb et al ¹²	System that combines analysis of extent, stage of disease (staging), and disease progression (spreading). Extent is evaluated using the rule of 9. Staging is based on cutaneous and hair pigmentation in vitiligo patches, and disease is staged 0-4 on the largest macule in each body region except the hands and feet, which are assessed separately and globally as 1 unique area. Assessment of spreading is based on Wood's lamp examination of the same largest macule in each body area.
VASI	Surface of lesions	Hamzavi et al ²³	$VASI = \Sigma(\text{all body sites}) [\text{hand units}] \times [\text{residual depigmentation}]$ The body was divided into 5 separate and mutually exclusive regions: hands, upper extremities (excluding hands), trunk, lower extremities (excluding the feet), and feet. The face and neck areas were not included in the overall evaluation. One hand unit, which encompasses the palm plus the volar surface of all the digits, is approximately 1% of the total body surface area and was used as a guide to estimate the baseline percentage of vitiligo involvement of the body. For all affected hand units, the extent of the residual depigmentation was estimated to the nearest of the following percentages: 0%, 10%, 25%, 50%, 75%, 90%, or 100%.
Point counting	Surface of lesions	Aydin et al ²⁴	Borders of vitiligo lesions are marked on a transparent sheet. This transparent grid set was superimposed to cover the entire lesion projection area randomly. The number of intersection hitting the lesions was counted. The total area of the lesion was estimated by multiplying 0.1 cm ² by the number of points.
Patient-reported outcomes			
Skindex-29	HRQoL	Nijsten et al ²⁵ ; Zghal et al ²⁶	HRQoL questionnaire for skin disease: 29 questions in 3 domains (emotions, symptoms, and functioning) with a 5-point scale (total score range, 0-100, with 0 = best quality of life, and 100 = worst quality of life).
Skindex-16	HRQoL	AlGhamdi and AlShammari ²⁷	HRQoL questionnaire for skin disease: 16 questions in 3 domains (emotions, symptoms, and functioning) with a 5-point scale (total score range, 0-100, with 0 = best quality of life, and 100 = worst quality of life).
Skindex-Teen	HRQoL	Smidt et al ²⁸	HRQoL questionnaire for adolescents with skin disease. 22 questions in 2 domains (psychosocial functioning and physical symptoms) with a 5-point scale (total score not specified).
DLQI score	HRQoL	Aghaei et al ²² ; Kent and al-Abadie ²⁹	HRQoL questionnaire: 10 questions with a 4-point scale (total score range, 0-30, with 0 = best quality of life, and 30 = worst quality of life).
PBI	Benefit of therapy	Augustin et al ³⁰	Includes weighting of individual benefits by patients needs in the therapy of vitiligo. Needs prior to therapy (PNQ) and benefits achieved by therapy (PBQ) are converted to weighted index value, the PBI. $PBI = \Sigma (PNQ/\Sigma PNQ)PBQ$
PRISM	Representation of illness and perception of suffering	Rumpf et al ³¹	Self-administered measurement. Sheet of paper with a fixed disk at the bottom right-hand corner representing the subject self. Individuals were asked to imagine that the paper represents his or her life as it is at the moment. A second disk is used to represent the disease, and subjects are asked where to put the disk in his or her life at the moment. The outcome measure is the distance between the centers of the 2 disks, called self-illness separation.
Observer-reported outcomes			
DIAS	Surface of lesions	Van Geel et al ³²	DIAS for surface measurement of vitiligo lesions. Digital photographs copied on a transparent sheet. The surface measurement of the transparent sheet is done with a simple image processing program in Matlab (The Math Works Inc). This results in 3-dimensional lesions surface measurements.
Image analysis technique	Repigmentation of lesions	Nugroho et al ³³ ; Fadzil et al ³⁴	Objective digital skin image technique to monitor progression of repigmentation over a shorter period of 6 weeks. The technique is based on principal component analysis and independent component analysis, which convert red, green, and blue skin images into images that represent skin areas due to melanin and hemoglobin.

Abbreviations: DIAS, Digital Image Analysis System; DLQI, Dermatology Life Quality Index; HRQoL, health-related quality of life; PBI, Patient Benefit Index; PRISM, Pictorial Representation of Illness Measure; VASI, Vitiligo Area Scoring Index; VETF, Vitiligo European Task Force.

internal consistency and content validity, the evidence is unknown owing to poor methodological quality of the study. Limited evidence was rated for positive construct validity and positive results of the Arabic translation owing to fair methodological quality of the study. No information on interpretability was reported.

Skindex-Teen. One study²⁸ evaluated the Skindex-Teen on internal consistency, reliability, content validity, structural validity, and responsiveness. For the internal consistency, structural validity, reliability, and responsiveness, limited positive evidence was rated, while for content validity the evidence was rated as unknown owing to poor

Table 2. Characteristics of the Included Populations

Source	Instrument	No. of Patients	Vitiligo, %	Age, Mean (SD)	Sex, M/F	Setting	Country	Language
Clinician-reported outcomes								
Taieb et al, ¹² 2007	VETF assessment	13	100	NR	NR	Hospital	10 European countries	English
Hamzavi et al, ²³ 2004	VASI	22	100	47 (12.7) (range, 23-77)	9/13	Hospital	Canada	English
Aydin et al, ²⁴ 2007	Point counting	31	100	31.6 (17.7)	NR	Hospital	Turkey	English
Patient-reported outcomes								
Nijsten et al, ²⁵ 2009	Skindex-29	454	6	NR	NR	Hospital	Italy	Italian
Zghal et al, ²⁶ 2003	Skindex-29		33	38 (1)	31/29	Hospital	Tunis	French
AlGhamdi and AlShammari, ²⁷ 2007	Skindex-16	338	23	Cases, 28.53 (10.73); controls, 29.64 (11.03)	Cases, 172/166; controls, 177/163	Cases, hospital; controls, population + hospital	Saudi Arabia	Arabic
Smidt et al, ²⁸ 2010	Skindex-Teen	200	1.5	15.3 (1.4) (range, 12.9-18)	91/109	Hospital	USA	English
Aghaei et al, ²² 2004	DLQI	70	100	28.3 (11.09)	27/43	Hospital	Iran	Persian
Kent and al-Abadie, ²⁹ 1996	DLQI	614	100	46.6 (range, 16-81)	150/464	Vitiligo society	United Kingdom	English
Augustin et al, ³⁰ 2008	PBI	1023	100	NR	NR	Vitiligo society	Germany	German
Rumpf et al, ³¹ 2004	PRISM	343	100	41.8 (10.6)	93/250	Vitiligo society	Germany	German
Observer-reported outcomes								
Van Geel et al, ³² 2004, mean	DIAS	10	100	29.4 (range, 15-51)	4/6	Hospital	Belgium	Dutch
Nugroho et al, ³³ 2007	Image analysis technique	4	100	NR	NR	NR	Malaysia	Malay
Fadzil et al, ³⁴ 2009	Image analysis technique	20	100	NR	NR	Hospital	Malaysia	Malay

Abbreviations: DIAS, Digital Image Analysis System; DLQI, Dermatology Life Quality Index; NR, none reported; PBI, Patient Benefit Index; PRISM, Pictorial Representation of Illness Measure; USA, United States of America; VASI, Vitiligo Area Scoring Index; VETF, Vitiligo European Task Force.

Table 3. Methodological Quality of Each Study per Measurement Property and Instrument

Source	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Construct Validity	Translation	Criterion Validity	Responsiveness
Clinical-reported outcomes									
Taieb et al ¹² (VETF assessment)	Poor
Hamzavi et al ²³ (VASI)	Fair	...
Aydin et al ²⁴ (point counting)	...	Fair	Fair
Patient-reported outcomes									
Nijsten et al ²⁵ (Skindex-29)	Poor	...
Zghal et al ²⁶ (Skindex-29)	Poor
AlGhamdi and AlShammari ²⁷ (Skindex-16)	Poor	Poor	...	Fair	Fair
Smidt et al ²⁸ (Skindex-Teen)	Excellent	Fair	...	Poor	Excellent	Fair
Aghaei et al ²² (DLQI score)	Good	Good	Fair
Kent and al-Abadie ²⁹ (DLQI score)	Good	Poor
Augustin et al ³⁰ (PBI)	Poor	Poor
Rumpf et al ³¹ (PRISM)	Poor
Observer-reported outcomes									
Van Geel et al ³² (DIAS)	Fair	Poor	...
Nugroho et al ³³ (image analysis technique)	Poor
Fadzil et al ³⁴ (image analysis technique)	Poor

Abbreviations: DIAS, Digital Image Analysis System; DLQI, Dermatology Life Quality Index; PBI, Patient Benefit Index; PRISM, Pictorial Representation of Illness Measure; VASI, Vitiligo Area Scoring Index; VETF, Vitiligo European Task Force Assessment; ellipses, not evaluated and therefore not applicable.

Table 4. Quality of Measurement Properties per Instrument^a

Instrument	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Construct Validity	Translation	Criterion Validity	Responsiveness
Clinician-reported outcomes									
VETF assessment	NA	NA	?	NA	NA	NA	NA	NA	NA
VASI	NA	NA	NA	NA	NA	NA	NA	?	NA
Point counting	NA	+	NA	NA	NA	NA	NA	NA	NA
Patient-reported outcomes									
Skindex-29	?	NA	NA	NA	NA	NA	NA	?	NA
Skindex-16	?	NA	NA	?	NA	+	+	NA	NA
Skindex-Teen	+	+	NA	?	+	NA	NA	NA	+
DLQI score	+++	NA	NA	NA	NA	++	+	NA	NA
PBI	?	NA	NA	NA	NA	?	NA	NA	NA
PRISM	NA	NA	NA	NA	NA	?	NA	NA	NA
Observer-reported outcomes									
DIAS	NA	NA	?	NA	NA	NA	NA	?	NA
Image analysis technique	NA	NA	NA	NA	NA	?	NA	NA	NA

Abbreviations: DIAS, Digital Image Analysis System; DLQI, Dermatology Life Quality Index; NA, no information available; PBI, Patient Benefit Index; PRISM, Pictorial Representation of Illness Measure; VASI, Vitiligo Area Scoring Index; VETF, Vitiligo European Task Force Assessment.

^aRating: + + +, strong evidence positive result; + +, moderate evidence positive result; +, limited evidence positive result; ?, unknown due to poor methodological quality.

methodological quality of the study. No information on the interpretability was reported.

Dermatology Life Quality Index (DLQI) Score. The internal consistency and construct validity were evaluated for the DLQI.^{22,29} Strong evidence was found for positive internal consistency, whereas moderate evidence was rated for positive construct validity. The evidence of a positive Persian translation was rated as limited owing to fair methodological quality of the study. No information on interpretability was found.

Patient Benefit Index (PBI). There was 1 study³⁰ that evaluated both internal consistency and construct validity. For both measurement properties the evidence was rated as unknown because the methodological quality of the study was poor. Missing items were not clearly described, while distributional characteristics of the PBI were well described. It showed a right-skewed distribution with a mean of 1.03 and marked floor effects; particularly, only a few values above the threshold of 1 were observed (SD, 1.13; median, 0.61; interquartile range, 0.03-1.71; n = 711) with respect to all therapies. Consequently, only slight patient-defined benefit was attained for the majority of patients receiving therapy.

Pictorial Representation of Illness Measure (PRISM). The construct validity of the PRISM was evaluated in 1 study.³¹ Because the methodological quality of the study was poor, the evidence for construct validity was rated as unknown. No other measurement property was evaluated, and no information on interpretability was found.

Observer-Reported Outcomes

Digital Image Analysis System (DIAS). For the DIAS, the measurement error and criterion validity were evaluated by Van Geel et al.³² For the measurement error, fair

methodological quality was found but no minimal important change was defined. As a result, the evidence for the quality of the measurement error, as well as the evidence for the criterion validity, was rated as unknown owing to poor methodological quality. No information on interpretability was found.

Image Analysis Technique. Two studies evaluated the construct validity of the image analysis technique.^{33,34} However, the evidence of the construct validity was rated as unknown because both studies have poor methodological quality. No information on the interpretability was found.

COMMENT

In this systematic review, the current evidence is summarized on the measurement properties of outcome measures for vitiligo. Fourteen studies met the inclusion criteria, in which the following 11 measurement instruments were evaluated: VETF, VASI, point counting, Skindex-29, Skindex-16, Skindex-Teen, DLQI, PBI, PRISM, DIAS, and image analysis technique.

Because only 14 studies were included in this review, this indicates that little research was done on measurement properties of outcome measures for vitiligo. Moreover, the methodological quality of the included studies was mostly poor. Nevertheless, strong evidence was rated for a positive internal consistency of the DLQI. Most evidence was found for the Skindex-Teen, since 5 measurement properties were evaluated for this instrument. Limited positive evidence was rated for internal consistency, reliability, structural validity, and responsiveness. For content validity, the evidence was rated unknown owing to poor methodological quality. However, the study population of this study²⁸ consisted of just 3 patients with vitiligo (1.5%). Therefore, it is questionable whether this evidence is truly valid specifically for patients with vitiligo.

It is remarkable, that the majority of the evaluated instruments are patient-reported outcomes. An explanation could be that most of these patient-reported outcomes are health-related questionnaires, which more often are evaluated for their measurement properties than other outcome measures. The Cochrane review on interventions for vitiligo from the study by Whitton et al¹¹ found that the most frequently used measurement instrument was the percentage of repigmentation. However, no studies on measurement properties of this instrument were found for this review. Moreover, of the 11 included instruments of this review, only the DLQI and Skindex-29 were also found in the Cochrane review from Whitton et al.¹¹ Hence, it is evident that there is a need for studies on measurement properties in outcome measures used in current randomized controlled trials.

This systematic review on measurement properties in vitiligo uses the COSMIN checklist. The COSMIN checklist was developed for assessing the methodological quality of studies on measurement properties. It is based on consensus between experts in the field of health status questionnaires and has recently been adjusted with a 4-point rating system.^{9,14} However, the interrater reliability for many COSMIN items is poor when using the original dichotomous items, which is suggested to be because of interpretation of the checklist.³⁵ The interrater reliability of the 4-point scale and accompanying total quality scores as used in this review has not yet been evaluated. Although decisions were made in advance on how to score the different items, this could be a limitation in the review process.

The clinical and biological significance was noted by Bigby and Gadenne¹ to be important for good outcome measures. However, these significances are not included in the COSMIN checklist, and for that reason the included articles are not scored on their clinical and biological significance. It was not in the scope of the present review to define what is clinical and biological significant.

It is remarkable that in the title and abstract selection a large number (n = 1152) of noneligible records were excluded. This could be explained by the search filter used. A validated highly sensitive special search filter for finding studies on measurement properties was used to make sure no relevant studies were missed. This filter has a sensitivity of 97.4% and a precision of 4.4%. Using a more precise filter (sensitivity of 93.1% and precision of 9.4%) might lower the amount of noneligible records but raises the risk of missing relevant studies.¹³

It is good to note that for a good outcome measure different measurement properties are important. Hence, reliability is significant but not sufficient for a good outcome measure, that is, it must be reliable but it must also measure what it is intended to measure. This validity is essential and it is good to realize that there are different types of validity (eTable 1).

Research of outcome measures in some fields of medicine is still in its infancy. For rheumatology, the Outcome Measures in Rheumatology (OMERACT) was initiated in 1992, an international initiative to improve outcome measurement.^{4,36} To date, most research on outcome measures in dermatology has been done for psoriasis and atopic dermatitis.^{5,6,8} Besides, Harmonising Outcome Measures for Eczema (HOME) consensus meetings

have taken place to define a minimum set of core outcomes for future eczema (atopic dermatitis) research.³ For vitiligo there is no consensus yet for core domains to categorize outcome measures. To harmonize outcome measures for vitiligo in clinical practice and research, a Delphi study or consensus meeting is suggested to create consensus on terminology and to define core outcome domains. The next step will be to initiate further research on measurement properties of relevant outcome measures.

CONCLUSIONS

Current evidence on measurement properties of outcome measures for vitiligo is insufficient for 3 reasons. First, the number of included articles was very low. Second, the majority of the evaluated studies are of poor methodological quality. Third, the clinical relevance is unclear. Because of this insufficient evidence, recommendations on the use of specific measurement instruments should be formulated with caution. Strong evidence was found for a positive internal consistency of the DLQI. For other instruments the evidence of measurement properties was limited or unknown. To indicate and improve imperfections in outcome measures in vitiligo and to recommend these outcome measures, further research on measurement properties is needed. For future research it is recommended to evaluate relevant measurement instruments according to the COSMIN criteria to assure methodological quality.

Accepted for Publication: June 19, 2012.

Published Online: September 17, 2012. doi:10.1001/archdermatol.2012.3065

Correspondence: Charlotte Vrijman, MD, Netherlands Institute for Pigment Disorders, Department of Dermatology, Academic Medical Center, University of Amsterdam, Meibergdreef 35, 1105 AZ Amsterdam, the Netherlands (c.vrijman@amc.uva.nl).

Author Contributions: Drs Vrijman and Linthorst Homan had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Vrijman, Linthorst Homan, Wolkerstorfer, Terwee, and Spuls. *Acquisition of data:* Vrijman, Linthorst Homan, and Limpens. *Analysis and interpretation of data:* Vrijman, Linthorst Homan, van der Veen, Wolkerstorfer, Terwee, and Spuls. *Drafting of the manuscript:* Vrijman, Limpens, and Spuls. *Critical revision of the manuscript for important intellectual content:* Vrijman, Linthorst Homan, Limpens, van der Veen, Wolkerstorfer, Terwee, and Spuls. *Statistical analysis:* Vrijman and Spuls. *Administrative, technical, and material support:* Linthorst Homan and Limpens. *Study supervision:* van der Veen, Wolkerstorfer, Terwee, and Spuls. **Financial Disclosure:** None reported.

Online-Only Material: The eTables are available at <http://www.archdermatol.com>.

Additional Contributions: We are indebted to Arnold Leenders, Master of Library and Information Science (MLIS), clinical librarian, Academic Medical Center of Amsterdam.

REFERENCES

1. Bigby M, Gadenne AS. Understanding and evaluating clinical trials. *J Am Acad Dermatol*. 1996;34(4):555-590.
2. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.
3. Schmitt J, Williams H; HOME Development Group. Harmonising Outcome Measures for Eczema (HOME): report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol*. 2010;163(6):1166-1168.
4. Naredo E, Wakefield RJ, Iagnocco A, et al. The OMERACT ultrasound task force—status and perspectives. *J Rheumatol*. 2011;38(9):2063-2067.
5. Spuls PI, Lecluse LL, Poulsen ML, Bos JD, Stern RS, Nijsten T. How good are clinical severity and outcome measures for psoriasis? quantitative evaluation in a systematic review. *J Invest Dermatol*. 2010;130(4):933-943.
6. Schmitt J, Langan S, Williams HC; European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? a systematic review. *J Allergy Clin Immunol*. 2007;120(6):1389-1398.
7. Schmitt J, Langan S, Stamm T, Williams HC; Harmonizing Outcome Measurements in Eczema (HOME) Delphi panel. Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. *J Invest Dermatol*. 2011;131(3):623-630.
8. Schram ME, Spuls PI, Leeftang MM, Lindeboom R, Bos JD, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. *Allergy*. 2012;67(1):99-106.
9. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.
10. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.
11. Whitton ME, Pinart M, Batchelor J, Lushey C, Leonardi-Bee J, González U. Interventions for vitiligo. *Cochrane Database Syst Rev*. 2010;(1):CD003263.
12. Taieb A, Picardo M; VETF Members. The definition and assessment of vitiligo: a consensus report of the Vitiligo European Task Force. *Pigment Cell Res*. 2007;20(1):27-35.
13. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123.
14. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745.
15. de Vet HC, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied “minimally important change” values. *J Clin Epidemiol*. 2010;63(1):37-45.
16. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol*. 2010;63(7):804-805.
17. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63(5):524-534.
18. US Food and Drug Administration. Review and Qualification of Clinical Outcome Assessments; Public Workshop. <http://www.fda.gov/Drugs/NewsEvents/ucm276110.htm>. Accessed November 2011.
19. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
20. van Tulder M, Furlan A, Bombardier C, Bouter L; Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)*. 2003;28(12):1290-1299.
21. Jalel A, Soumaya GS, Hamdaoui MH. Dermatology life quality index scores in vitiligo: reliability and validity of the Tunisian version. *Indian J Dermatol*. 2009;54(4):330-333.
22. Aghaei S, Sodaifi M, Jafari P, Mazharinia N, Finlay AY. DLQI scores in vitiligo: reliability and validity of the Persian version. *BMC Dermatol*. 2004;4:8.
23. Hamzavi I, Jain H, McLean D, Shapiro J, Zeng H, Lui H. Parametric modeling of narrowband UV-B phototherapy for vitiligo using a novel quantitative tool: the Vitiligo Area Scoring Index. *Arch Dermatol*. 2004;140(6):677-683.
24. Aydin F, Senturk N, Sahin B, Bek Y, Yuksel EP, Turanli AY. A practical method for the estimation of vitiligo surface area: a comparison between the point counting and digital planimetry techniques. *Eur J Dermatol*. 2007;17(1):30-32.
25. Nijsten T, Sampogna F, Abeni D. Categorization of Skindex-29 scores using mixture analysis. *Dermatology*. 2009;218(2):151-154.
26. Zghal A, Zeglaoui F, Kallel L, et al. Quality of life in dermatology: Tunisian version of the Skindex-29. *Tunis Med*. 2003;81(1):34-37.
27. AlGhamdi KM, AlShammari SA. Arabic version of Skindex-16: translation and cultural adaptation, with assessment of reliability and validity. *Int J Dermatol*. 2007;46(3):247-252.
28. Smidt AC, Lai J-S, Cella D, Patel S, Mancini AJ, Chamlin SL. Development and validation of Skindex-Teen, a quality-of-life instrument for adolescents with skin disease. *Arch Dermatol*. 2010;146(8):865-869.
29. Kent G, al-Abadie M. Factors affecting responses on Dermatology Life Quality Index items among vitiligo sufferers. *Clin Exp Dermatol*. 1996;21(5):330-333.
30. Augustin M, Gajur AI, Reich C, Rustenbach SJ, Schaefer I. Benefit evaluation in vitiligo treatment: development and validation of a patient-defined outcome questionnaire. *Dermatology*. 2008;217(2):101-106.
31. Rumpf H-J, Löntz W, Uessler S. A self-administered version of a brief measure of suffering: first aspects of validity. *Psychother Psychosom*. 2004;73(1):53-56.
32. Van Geel N, Vander Haeghen Y, Ongenaë K, Naeyaert J-M. A new digital image analysis system useful for surface assessment of vitiligo lesions in transplantation studies. *Eur J Dermatol*. 2004;14(3):150-155.
33. Nugroho H, Fadzil MH, Yap VV, Norashikin S, Suraiya HH. Determination of skin repigmentation progression. *Conf Proc IEEE Eng Med Biol Soc*. 2007;2007:3442-3445.
34. Fadzil MH, Norashikin S, Suraiya HH, Nugroho H. Independent component analysis for assessing therapeutic response in vitiligo skin disorder. *J Med Eng Technol*. 2009;33(2):101-109.
35. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med Res Methodol*. 2010;10:82.
36. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials*. 2007;8:38.