Supplementary file 4: Methodological domains of the risk of bias criteria

Domains	Definitions	Grades and Criteria		
Validity Conceptual[1] Construct[2] Convergent[3]	Does the tool measure what it is supposed to measure? Are the relevant domains captured? Does tool confirm hypothesized difference (eg diagnosis, clinical disease severity, others) Does the tool relate to other tools measuring the same construct?	A1: well balanced, objective and subjective domains B1: more focused on objective or subjective domains C1: missing important HRQOL domains	A2:>75% of results are in accordance with specific hypotheses B2: <75 of results are in accordance with specific hypotheses C2: no information	A3: correlation> B3: correlation < C3: no information
Interpretability Norms Categorization MCID[4]	Are there standard comparative data from the general population and/or dermatology patients published and/or available? Are there categories of the obtained score available? Has the minimal change that is relevant to patients been reported?	A1: general and dermatology patients B1: general or dermatology patients C1: general nor dermatology patients	A2: using anchor or banding techniques B2: using distribution-based techniques C2: not reported	A3: MCID is known in heterogeneous sample B3: MCID is known in limited sample C3: not reported
Reliability[3,6] Internal consistency Retest-reliability	Does the tool provide a consistent answer? The extents to which items in a (sub) scale are intercorrelated, thus measuring the same construct (Cronbach's x)? Does a repeated administration of the tool within a reasonable period result in a similar outcome?	A1: 0.95>Cronbach's x>0.70 B1: Cronbach's x<0.7 or >0.95 C: Cronbach's x not reported	A2: x or ICC >0.7 B2: x or ICC <0.7 or correlation coefficients >0.7 C2: x or ICC not reported or correlation coefficient <0.7	
Structure	Have the domains and/or summary score of the tool been confirmed?	A: item response theory B: Factor analysis C: no factor analysis or item response theory		
Responsiveness	Is the tool sensitive to detect changes over time or due to therapy using patient centred and/or clinical criteria?	A: strong B: moderate or conflicting evidence C: absent, weak or solely based on statistical evidence		

Domains	**Definitions**	**Grades and Criteria**	
Item bias	Do the items of the tool function similar across external factors such as age, gender and diagnosis?	A: strong B: moderate or conflicting evidence C: absent or weak	
Cultural issues Translations Cultural equivalence	Has the tool been translated using guidelines? Has the tool been analysed in a cultural equivalence study?	A1: always B1: sometimes C1: never, not reported	A2: always B2: sometimes C2: never
Respondent burden	Is the length and content acceptable to the patients?	A: brief (<15min) B: long or problems of acceptability C: long and problems of acceptability	
Administrative burden	How easy is the tool to administer, score and interpret?	A: simple B: moderate C: complex	
Alternative forms	Is the tool available and tested for alternate forms of administration such as interviews in person or telephone, self-administration or computer-assisted interviews	A: strong evidence B: moderate or conflicting evidence C: absent or weak evidence	

Legend: ICC, intraclass correlation coefficient; [1] Adjusted from Lohr et al (1996); Andresen (2000) and Terwee et al (2007); [2] Objective and subjective domains are described by Muldoon et al (1998); [3] Criteria of construct validity and reliability were based on description by Terwee et al (2007); [4] MCID, minimal clinically important difference (ie the minimal difference, which is measured and is relevant to a patient and is not due to intrinsic variance of the instrument); [5] Refer to Table 2; [6] Reliability is concerned with the temporal stability of instrument scores (test-retest) and internal consistency, which is estimated by Cronbach's x, evaluates the relationship between all items (of a scale) and their ability to measure a single underlying domain. Test-retest reliability assess score consistency over two points in time assuming no change in health status and may provide a more rigorous of reliability due to the different sources of variance. Test-retest reliability should best be expressed in a x coefficient or ICC. Spearman's correlation coefficients are less optimal for retest reliability.