**Online Supplementary Material
Online supplementary methods 1**

*2.1. Search strategy*

No date or language restrictions were applied. The bibliographic records were imported and deduplicated using ENDNOTE X7 (Thomson Reuters, New York, NY, USA).

*2.2. Exclusion and inclusion criteria*

No minimum percentage of PWS patients in studies with mixed populations was applied. Studies had to be published as a full-text article. Technical instruments (e.g., colorimetry) were excluded. Conference abstracts were excluded. No further limitations on the types of papers were applied. Studies that presented new OMIs without any elaboration on how it was developed or any assessment of its measurement properties were not included [1, 2]. Studies that only correlated a clinical assessment with an objective instrument (e.g., colorimetry) as an adjunct to a clinical trial were also excluded because these generally provide very little usable information [3].

*2.3. Data extraction*

When possible, non-English articles were translated. Two reviewers (MIvR and SC) independently assessed titles and abstracts and selected full-text articles and their reference lists using Rayyan [4]. In case of disagreement between the two reviewers, a third reviewer (CvdH) was consulted until consensus was reached. Characteristics of the included instruments, study population, and data on the results of the OMI’s measurement properties, interpretability, and feasibility were extracted independently by two authors (MIvR and SC) and discrepancies were discussed until consensus was reached. In case of missing data, attempts were made to contact study authors.

*2.4. PROM readability*

The readability of included PROMs was determined using the Flesch-Kincaid grade level test incorporated into Microsoft Word 2013 (Microsoft Corp, Redmond, WA) and presented as the equivalent US grade level required to understand the questionnaire.

*2.5. Modified COSMIN standards*

The 5th standard for content validity, “recall period,” and “time interval,” the 2nd standard for reliability and measurement error, were ignored for the assessment of clinical photographs. The reliability criteria were used for both intrarater and inter-rater reliability. Professionals instead of patients were recognized as the subjects for a cognitive interview or pilot testing in the development of clinical OMIs.

*2.6. Additional sources used to evaluate the development and content validity of outcome measurement instruments*

**CBCL questionnaires**

Achenbach, T. M., Rescorla, L. A. (2000). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T.M. (2009). *The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

**PSQ**

Lawrence JW, Fauerbach JA, Heinberg LJ, Doctor M, Thombs BD. The reliability and validity of the Perceived Stigmatization Questionnaire (PSQ) and the Social Comfort Questionnaire (SCQ) among an adult burn survivor sample. *Psychol Assess. 2006;18*(1):106–11.

Masnari O, Landolt MA, Roessler J, Weingaertner SK, Neuhaus K, Meuli M, et al. Self- and parent-perceived stigmatisation in children and adolescents with congenital or acquired facial differences. *J Plast Reconstr Aesthetic Surg. 2012;65*(12):1664–70.

**TAPQOL**

M. Fekkes, M., Bruil, J., Vogels, T. (2004). *TAPQOL-Manual.* Leiden, the Netherlands: TNO Prevention and Health

Fekkes M, Theunissen NCM, Brugman E, Veen S, Verrips EGH, Koopman HM. Development and psychometric evaluation of the TAPQOL: A health-related quality of life instrument for 1-5-year-old children. *Qual Life Res. 2000;9*(8):961–72.

**KIDSCREEN-27**

The KIDSCREEN Group Europe. (2006). *The KIDSCREEN Questionnaires – Quality of life questionnaires for children and adolescents. Handbook.* Lengerich: Pabst Science Publishers.

Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W, et al. KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Rev Pharmacoeconomics Outcomes Res. 2005;5*(3):353–64.

Detmar SB, Bruil J, Ravens-Sieberer U, Gosch A, Bisegger C. The use of focus groups in the development of the KIDSCREEN HRQL questionnaire. *Qual Life Res. 2006;15*(8):1345–53.

Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M, et al. The KIDSCREEN-27 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Value Heal. 2008;11*(4):645–58.

Ravens-Sieberer U, Gosch A, Abel T, Auquier P, Bellach BM, Bruil J, et al. Quality of life in children and adolescents: a European public health perspective. *Soz Praventivmed. 2001;46*(5):294–302.

**DLQI**

Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol. 1994;19*(3):210–6.

Wang XL, Zhao TE, Zhang XQ, et al. [Assessment on the reliability and validity of the Dermatology Life Quality Index in Chinese version]. *Zhonghua Liu Xing Bing Xue Za Zhi 2004;25*(9):791–3.

**Supplemental references**

1. Quaba AA (1989) Results of argon laser treatment of port-wine stains: a method of assessment. Br J Plast Surg 42:125–132

2. Ginsbach G (1991) A Tool for the Evaluation of Colour in Port Wine Stains. Lasers Med Sci 6:49–52

3. Rah DK, Kim SC, Lee KH, Park BY, Kim DW (2001) Objective evaluation of treatment effects on port-wine stains using L\*a\*b\* color coordinates. Plast Reconstr Surg 108:842–847

4. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. Syst Rev 5:210

**Online supplementary Table 1.** Searches performed in MEDLINE and Embase.

Database(s): **Ovid MEDLINE and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily**1946 to March 29, 2019
Search Strategy: **2019-04-01**

|  |  |  |
| --- | --- | --- |
| **#** | **Searches** | **Results** |
| 1 | port-wine stain/ | 934 |
| 2 | ((port win\* or portwin\*) adj6 l?esion\*).tw,ot,kf. | 132 |
| 3 | (port-win\* or portwin\*).tw,ot,kf. and (exp hemangioma/ or exp angiomatosis/ or vascular malformations/ or capillaries/ or (PWS or stain\* or birthmark\* or mark or marks or n?evus or n?evi or h?emangiom\* or angiom\* or malform\* or anomal\* or SWS or Sturge or Weber or facial or capillar\*).tw,ot,kf.) | 1483 |
| 4 | (PWS\* adj3 (birthmark\* or birth-mark\* or n?evus or n?evi or h?emangiom\* or angiom\* or malformat\* or capillar\*)).tw,kf. | 70 |
| 5 | ((n?evus or n?evi) and (flamm?eus or vinos\*)).tw,ot,kf. | 308 |
| 6 | (vascular adj (ne?vus or n?evi)).tw,ot,kf. | 119 |
| 7 | (birthmarks or vascular birthmark\*).ti. | 133 |
| 8 | ((capillar\* adj4 malformat\*).tw,kf. or (capillaries/ and vascular malformations/)) and (laser\* or video\* or PDL or photo\*).mp. | 145 |
| 9 | ((facial or face or head or neck or lip or lips or trunk or arm or arms or leg or legs or skin or derm\* or cutaneous\*) adj3 (capillar\* adj2 malformat\*)).tw,kf. | 111 |
| **10** | **or/1-9 [PWS]** | **2263** |
| 11 | exp animals/ not humans/ | 4564068 |
| **12** | **10 not 11 [ human PWS ]** | **2223** |
| 13 | ((exp Health Status Indicators/ or exp "Outcome Assessment (Health Care)"/ or Comparative Study/ or (instrumentation or methods).fs.) not case reports.pt.) or validation studies.pt. or observer variation/ or discriminant analysis/ or Psychometrics/ or "Reproducibility of Results"/ or factor analysis, statistical/ or evaluation studies/ or (audit or audits or psychometr\* or clin?metr\* or ((outcome\* or clinical or observer\* or utility or satisfaction or QoL or quality of life or score or scores or method or methods or physicians or dermatologists or modelling or objective) adj3 assessm\*) or clinical asses\* or outcome measure\* or observer variation\* or reproducib\* or reliab\* or unreliab\* or valid\* or coefficient or homogeneity or homogeneous or ((internal or external) adj3 (consistency or inconsistency)) or cronbach\* or (item and (correlation\* or selection\* or reduction\*)) or ((item or items) adj3 (discriminant\* or convergent\* or divergent\*)) or agreement or precision or imprecision or (precise adj values) or (test and retest) or accuracy test\* or stability or interrater or intrarater or intertester or intratester or interobserver or intraobserver or intertechnician or intratechnician or interexaminer or intraexaminer or interassay or intraassay or interindividual or intraindividual or interparticipant or intraparticipant or ((inter or intra) adj (rater or tester or observer or technician or examiner or assay or individual or participant)) or kappa or kappa's or kappas or repeatab\* or ((replicab\* or repeated) and (measure or measures or findings or result or results or test or tests)) or generaliza\* or generalisa\* or concordance or (intraclass and correlation\*) or discriminative or (known adj group) or (factor adj (analy\* or structure\*)) or dimension\* or interscale or inter-scale or interscales or inter-scales or subscale\* or sub-scale\* or ((multitrait\* or multi-trait\*) and (scaling or scale\*)) or error or errors or ((individual or interval or rate) adj variability) or (variability adj5 (analy\* or values)) or (uncertainty and (measurement or measuring)) or sensitiv\* or responsive\* or ((limit or limits) and detection) or ((minim\* or lowest) adj2 detectable adj2 (concentration\* or dose\* or level\* or amount\*)) or interpretab\* or (small\* and (real or detectable) and (change or difference)) or meaningful change\* or ((minimal\* or minimum) adj2 (meaningful or important or detectable or real or identifiable or relevant) adj3 (change\* or difference\* or improvement\*)) or ((minimal or minimally or clinical or clinically) and (important or significant) and (change\* or difference\* or improvement\*)) or (MDC adj2 value\*) or MCID or MCIDs or MICD or MICDs or MCII or MCIC or MCICs or ((ceiling or floor) adj2 effect\*) or item response\* or IRT or rasch or ((differential or fit) adj2 item\*) or DIF or computer adaptive test\* or item bank\* or cross-cultural equivalen\*).tw,ot,kf,kw. **[ COSMIN FILTER adapted for OVID MEDLINE ]** | 8851762 |
| 14 | feasibility studies/ | 61565 |
| 15 | (feasibilit\* or practicab\* or practicalit\* or practibil\* or intuitiv\* or accept?bility or workab\* or viability or expedien\* or usefulness or (complet\* adj2 time) or quality criter\*).tw,ot,kf. | 506859 |
| 16 | ((easy or ease or fast or simple or practical or feasible) adj6 ("use" or apply or perform\* or method\* or measure or measurement or outcome or model or instrument\* or tool or index or indices or score\* or scoring or scale\* or subscale\* or assess\* or evaluat\* or fill in or value)).tw,ot,kf. | 336912 |
| **17** | **or/14-16 [feasibility filter]** | **838454** |
| **18** | **13 or 17 [ COSMIN filter expanded with feasibility filter ]** | **9187717** |
| **19** | **12 and 18 [ measurement properties PWS ]** | **680** |
| **20** | **remove duplicates from 19 [ measurement properties PWS -deduplicated ]** | **680** |

Database(s): **EMBASE Classic+EMBASE**1947 to 2019 March 29. Search Strategy: **2019-04-01**

|  |  |  |
| --- | --- | --- |
| **#** | **Searches** | **Results** |
| 1 | nevus flammeus/ | 2581 |
| 2 | ((port win\* or portwin\*) adj6 l?esion\*).tw,ot,kw. | 173 |
| 3 | (port-win\* or portwin\*).tw,ot,kw. and (angioma/ or exp hemangioma/ or congenital blood vessel malformation/ or capillary/ or (PWS or stain\* or birthmark\* or mark or marks or n?evus or n?evi or h?emangiom\* or angiom\* or malform\* or anomal\* or SWS or Sturge or Weber or facial or capillar\*).tw,ot,kw.) | 2130 |
| 4 | (PWS\* adj3 (birthmark\* or birth-mark\* or n?evus or n?evi or h?emangiom\* or angiom\* or malformat\* or capillar\*)).tw,ot,kw. | 101 |
| 5 | ((n?evus or n?evi) and (flamm?eus or vinos\*)).tw,ot,kw. | 554 |
| 6 | (vascular adj (ne?vus or n?evi)).tw,ot,kw. | 167 |
| 7 | (birthmarks or vascular birthmark\*).ti. | 153 |
| 8 | ((capillar\* adj4 malformat\*).tw,kw. or (capillaries/ and vascular malformations/)) and (laser\* or video\* or PDL or photo\*).mp. | 266 |
| 9 | ((facial or face or head or neck or lip or lips or trunk or arm or arms or leg or legs or skin or derm\* or cutaneous\*) adj3 (capillar\* adj2 malformat\*)).tw,kw. | 186 |
| **10** | **or/1-9 [PWS]** | **3738** |
| 11 | (animal.hw. or nonhuman/) not human/ | 6149326 |
| **12** | **10 not 11 [ human PWS ]** | **3662** |
| 13 | methodology/ or exp health status indicator/ or Sickness Impact Profile/ or clinical assessment/ or clinical assessment tool/ or outcome assessment/ or outcomes research/ or medical assessment/ or measurement/ or exp measurement precision/ or exp measurement accuracy/ or measurement error/ or exp systematic error/ or exp performance measurement system/ or exp measurement repeatability/ or intermethod comparison/ or data collection method/ or system analysis/ or validation study/ or feasibility study/ or exp quality control/ or rating scale/ or scoring system/ or summated rating scale/ or qualitative analysis/ or quantitative analysis/ or correlation analysis/ or "constants and coefficients"/ or correlation coefficient/ or cronbach alpha coefficient/ or kappa statistics/ or correlation function/ or exp reliability/ or discriminant analysis/ or exp validity/ or valid\*.hw. or factorial analysis/ or observer variation/ or psychometry/ or (audit or audits or psychometr\* or clin?metr\* or ((outcome\* or clinical or observer\* or utility or satisfaction or QoL or quality of life or score or scores or method or methods or physicians or dermatologists or modelling or objective) adj3 assessm\*) or clinical asses\* or outcome measure\* or observer variation\* or reproducib\* or reliab\* or unreliab\* or valid\* or coefficient or homogeneity or homogeneous or ((internal or external) adj3 (consistency or inconsistency)) or cronbach\* or (item and (correlation\* or selection\* or reduction\*)) or ((item or items) adj3 (discriminant\* or convergent\* or divergent\*)) or agreement or precision or imprecision or (precise adj values) or (test and retest) or accuracy test\* or stability or interrater or intrarater or intertester or intratester or interobserver or intraobserver or intertechnician or intratechnician or interexaminer or intraexaminer or interassay or intraassay or interindividual or intraindividual or interparticipant or intraparticipant or ((inter or intra) adj (rater or tester or observer or technician or examiner or assay or individual or participant)) or kappa or kappa's or kappas or repeatab\* or ((replicab\* or repeated) and (measure or measures or findings or result or results or test or tests)) or generaliza\* or generalisa\* or concordance or (intraclass and correlation\*) or discriminative or (known adj group) or (factor adj (analy\* or structure\*)) or dimension\* or interscale or inter-scale or interscales or inter-scales or subscale\* or sub-scale\* or ((multitrait\* or multi-trait\*) and (scaling or scale\*)) or error or errors or ((individual or interval or rate) adj variability) or (variability adj5 (analy\* or values)) or (uncertainty and (measurement or measuring)) or sensitiv\* or responsive\* or ((limit or limits) and detection) or ((minim\* or lowest) adj2 detectable adj2 (concentration\* or dose\* or level\* or amount\*)) or interpretab\* or (small\* and (real or detectable) and (change or difference)) or meaningful change\* or ((minimal\* or minimum) adj2 (meaningful or important or detectable or real or identifiable or relevant) adj3 (change\* or difference\* or improvement\*)) or ((minimal or minimally or clinical or clinically) and (important or significant) and (change\* or difference\* or improvement\*)) or (MDC adj2 value\*) or MCID or MCIDs or MICD or MICDs or MCII or MCIC or MCICs or ((ceiling or floor) adj2 effect\*) or item response\* or IRT or rasch or ((differential or fit) adj2 item\*) or DIF or computer adaptive test\* or item bank\* or cross-cultural equivalen\*).tw,ot,kw. [ COSMIN FILTER adapted for OVID EMBASE ] | 8832635 |
| 14 | (feasibilit\* or practicab\* or practicalit\* or practibil\* or intuitiv\* or accept?bility or workab\* or viability or expedien\* or usefulness or (complet\* adj2 time) or quality criter\*).tw,ot,kw. | 688323 |
| 15 | ((easy or ease or fast or simple or practical or feasible) adj6 ("use" or apply or perform\* or method\* or measure or measurement or outcome or model or instrument\* or tool or index or indices or score\* or scoring or scale\* or subscale\* or assess\* or evaluat\* or fill-in or value)).tw,ot,kw. | 444124 |
| **16** | **or/13-15 [ COSMIN filter expanded with feasibility filter ]** | **9358970** |
| 17 | 12 and 16 [ measurement properties PWS ] | 729 |
| 18 | remove duplicates from 17 [ measurement properties PWS -deduplicated ] | 722 |
| **19** | **18 not medline.cr. [ measurement properties PWS -deduplicated - EMBASE records only ]** | **633** |

**Online supplementary Table 2.** COSMIN definitions and updated criteria for good measurement properties.

|  |  |  |  |
| --- | --- | --- | --- |
| **Measurement property** | **COSMIN definition** | **Rating** | **Criteria** |
| Content validity | The degree to which the content of an OMI is an adequate reflection of the construct to be measured, i.e., all items, response options, and the recall period are considered relevant for the construct to be measured, the target population, and the intended context of use AND the OMI is considered to be comprehensive AND the OMI is comprehensible by the target population |  | 1 |
| Structural validity | The degree to which the scores of an OMI are an adequate reflection of the dimensionality of the construct to be measured | + | **CTT:** CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.082**IRT/Rasch:** No violation of unidimensionality3 (CFI or TLI or comparable measure >0.95 OR RMSEA < 0.06 OR SRMR < 0.08) AND no violation of local independence (residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37) AND no violation of monotonicity (adequate looking graphs OR item scalability > 0.30) AND adequate model fit (**IRT**: χ2 > 0.01 / **Rasch:** infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z‐standardized values > ‐2 and < 2) |
| ? | **CTT:** Not all information for ‘+’ reported **IRT/Rasch:** Model fit not reported |
| - | Criteria for ‘+’ not met |
| Internal consistency | The degree of the interrelatedness among the items | + | At least low evidence4 for sufficient structural validity5 ANDCronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale6 |
| ? | Criteria for “At least low evidence4 for sufficient structural validity5” not met |
| - | At least low evidence4 for sufficient structural validity5 AND Cronbach’s alpha(s) < 0.70 for each unidimensional scale or subscale6 |
| Reliability | The proportion of the total variance in the measurements which is due to ‘true’ differences between patients | + | ICC or weighted Kappa ≥ 0.70 |
| ? | ICC or weighted Kappa not reported  |
| - | ICC or weighted Kappa < 0.70 |
| Measurement error | The systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured | + | SDC or LoA < MIC5 |
| ? | MIC not defined |
| - | SDC or LoA > MIC5 |
| (Hypotheses testing for) construct validity | The degree to which the scores of an OMI are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the OMI validly measures the construct to be measured | + | The result is in accordance with the hypothesis7 |
| ? | No hypothesis defined (by the review team) |
| - | The result is not in accordance with the hypothesis7 |
| Cross-cultural validity/measurement invariance | The degree to which the performance of the items on a translated or culturally adapted OMI are an adequate reflection of the performance of the items of the original version of the OMI | + | No important differences found between group factors (suchas age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R2 < 0.02) |
| ? | No multiple group factor analysis OR DIF analysis performed |
| - | Important differences between group factors OR DIF was found |
| Criterion validity | The degree to which the scores of an OMI are an adequate reflection of a ‘gold standard’ | + | Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70 |
| ? | Not all information for ‘+’ reported |
| - | Correlation with gold standard < 0.70 OR AUC < 0.70 |
| Responsiveness | The ability of an OMI to detect change over time in the construct to be measured | + | The result is in accordance with the hypothesis7 OR AUC ≥ 0.70 |
| ? | No hypothesis defined (by the review team) |
| - | The result is not in accordance with the hypothesis7 OR AUC < 0.70 |
| Interpretability8 | Interpretability is the degree to which one can assign qualitative meaning – that is, clinical or commonly understood connotations – to an OMI’s quantitative scores or change in scores |  |  |
| Reproduced and modified from the COSMIN manual (Mokkink *et al.* (2018)). Measurement properties are rated sufficient (+), insufficient (-), or indeterminate (?). 1The criteria for content validity can be found the corresponding COSMIN manual (Terwee *et al.* (2017)). 2To rate the quality of the summary score, the factor structures should be equal across studies. 3Unidimensionality refers to a factor analysis per subscale, while structural validity refers to a factor analysis of a (multidimensional) OMI. 4As defined by grading the evidence according to the COSMIN-modified GRADE approach. 5This evidence may come from different studies. 6The criteria ‘Cronbach’s alpha < 0.95’ was deleted, as this is relevant in the development phase of an OMI and not when evaluating an existing OMI. 7The results of all studies should be taken together and it should then be decided if 75% of the results are in accordance with the hypotheses. 8Not a measurement property but an important aspect related to the OMI. Abbreviations: AUC, area under the curve; CFA, confirmatory factor analysis; CFI, comparative fit index; CTT, classical test theory; DIF, differential item functioning; ICC, intraclass correlation coefficient; IRT, item response theory; LoA, limits of agreement; MIC, minimal important change; OMI, outcome measurement instrument; RMSEA: Root Mean Square Error of Approximation; SEM, Standard Error of Measurement; SDC, smallest detectable change; SRMR: Standardized Root Mean Residuals; TLI, Tucker‐Lewis index. |

**Online supplementary Table 3.** Feasibility of outcome measures.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **OMI** | **Readibility1** | **Type of administration** | **Length of the instrument (number of items)** | **Completion time** | **Ease of score calculation2** | **Copyright/Cost** | **Required equipment** | **Availability in different settings** |
| Koster's PWS questionnaire (Koster *et al.*) |  | PAE | 8 | 2-5 min | ++ | Free | None | Good |
| Sajan's PWS and HOI questionnaire (Sajan *et al.*) |  | PAE | 6 | 1-2 min | ++ (no total score) | Free | None | Good |
| Poor/fair/good/excellent (Currie *et al.*)  |  | PAE | 1 | 1-2 min | +++ | Free | None | Good |
| 0-100% clearance/success (Pérez *et al.,* Szychta *et al.*) |  | PAE | 1 | 1 min | +++ | Free | None | Good |
| 0-24; 25-49; 50-74 and 75-100% lightening/clearance/success |  | PAE | 1 | 1 min | +++ | Free | None | Good |
| Class 0-IV (Currie *et al.*) |  | PAE | 1 | 1-2 min | +++ | Free | None | Good |
| Lighter/darker/did not change (Naran *et al.*) |  | PAE | 1 | 1 min | +++ | Free | None | Good |
| DLQI | 83.6 / 2.7 | PAPI | 10 | 1-2 min | +++ | Free for most purposes | None | Good |
| PSQ | 75.5 / 4.6 (self-reported form) | PAPI/interview | 21 | < 5 min | ++ | Unknown | None | Poor |
| TAPQOL | 80.5 / 4.2 | PAPI | 43 | 10 min | + | License required for commercial use | Computerized scoring advisable | Good |
| KIDSCREEN-27 | 89.9 / 2.6 (child-form); 2.1 / 2.6 (parent-form) | PAPI/interview | 27 | 10-15 min | + | License required for commercial use | Computerized scoring advisable | Good |
| CBCL/1.5-5 | 78.4 / 3.5 | PAPI | 100 | 10-15 min | + | Fees per form apply | Computerized scoring advisable | Good |
| CBCL/4-18  | 79.4 / 3.5 | PAPI | 100 | 10-20 min | + | Fees per form apply | Computerized scoring advisable | Good |
| 1 Flesch reading ease score / grade level of the English versions of patient-/parent-reported questionnaires. 2 Subjective score ranging from + to +++. Abbreviations: CBCL, child behavior checklist; DLQI, dermatology life quality index; HOI, hemangioma of infancy; OMI, outcome measurement instrument; PAE, photograph-assisted evaluation; PAPI, paper and pencil self-administered questionnaire; PSQ, perceived stigmatization questionnaire; PWS, port wine stain; SD, standard deviation; SWS, Sturge-Weber syndrome; TAPQOL, TNO-AZL questionnaire for preschool children’s health-related quality of life. |

**Online supplementary Table 4.** Interpretability of outcome measures.

|  |  |  |  |
| --- | --- | --- | --- |
| **OMI** | **Distribution of scores in the study population** | **Percentage of missing items and percentage of missing total scores** | **Floor and ceiling effects** |
| Koster's PWS questionnaire (Koster *et al.*) |   | Before treatment Mean (SD) | After treatment Mean (SD) | NL | NL/cannot be derived |
| Color (1-7) | 4.8 (11) | 3.2 (1.1) |
| Patchiness (1-4) | 1.8 (0.5) | 1.8 (0.5) |
| Boundary (1-3) | 2.4 (0.5) | 1.5 (0.6) |
| Pigmentation (1-3) | 1.1 (0.1) | 1.2 (0.2) |
| Size (1-6) | 4.5 (0.9) | 3.9 (1.1) |
| Shape (1-3) | 2.1 (0.5) | 1.9 (0.4) |
| Surface (1-3) | 1.1. (0.3) | 1.1 (0.3) |
| Hypertrophy (1-4) | 1.2 (0.4) | 1.2 (0.4) |
| Sajan's PWS and HOI questionnaire (Sajan *et al.*) |   | Mean % improvement | NL       | NL/cannot be derived     |
| Color |  | 1-24%1 |
| Thickness |  | 0% |
| Size |  | 0% |
| Scarring |  | 1-24%1 |
| Atrophy |  | 1-24%1 |
| Pigmentation |  | 1-24%1 |
| Lighter/darker/did not change (Naran *et al.*) | Lighter |  | 33.3-63.9%2 | NL | Floor and ceiling effects  |
| Darker |  | 21.4-40.5%2 |   |
| Did not change |  | 16.7-40.5%2 |   |   |
| Excellent/good/fair/poor (Currie *et al.*) | Excellent, good, fair, poor | 9%, 40%, 40%, 11%1 | NL | Absent |
| Class 0-IV (Currie *et al.*) | Class 0, excellent; class I, excellent; class II, good; class III, fair; class IV, poor | 33%, 31%, 1%, 35%1 | NL | Floor and ceiling effects |
| 0-24, 25-49, 50-74, 75-100% lightening (Currie *et al.*) | <25%, 25-49%, 50-75%, >75% lightening | 19%, 33,5%, 29%, 18,5%1 | NL | Floor and ceiling effects |
| 0-24; 25-49; 50-74; and 75-100% clearance (Pérez *et al.*) | 0-24%, 25-49%, 50-74%, 75-100% clearance | 4.6%, 18.3%, 33.8%, 43.3%1 | NL | Ceiling effect |
| 0-100% clearance (Pérez *et al.*) | 0-100% clearance | 66.2%1 | NL | Could not be derived |
| 0-100% success (Szychta *et al.*) | 0-100% success | Core physicians: mean 70.0% Lay people: mean 75.7% | NL | Could not be derived |
| DLQI (Wang *et al.*) | Score: |  | N (%) |  | Absent |
| 2-5 (small effect) |  | 103 (52.3%) |
| 6-10 (moderate effect) |  | 83 (42.1) |
| 11-20 (very large effect) |   | 11 (5.6) |
| PSQ (patient-reported) (Masnari *et al.*) | Total score | Total 1.82 (0.49); pre-school 1.66 (0.40), school-age 2.10 (0.53) | NL | Could not be derived |
| PSQ (parent-reported) (Masnari *et al.*) | Total score |   | Not reported | NL | Could not be derived |
| TAPQOL (Masnari *et al.*) | Physical functioning |  | NL3 | NL    | Could not be derived |
| Social functioning |  | NL3 |
| Cognitive functioning |  | NL3 |   |
| Emotional functioning |  | NL3 |   |
|   |   |   | Mean (SD) |   |   |
| KIDSCREEN-27 (parent-reported) (Masnari *et al.*) | Physical well-being |  | 46.6 (15.7)4 | NL      | Could not be derived |
| Psychological well-being |  | 45.9 (12.2)4 |
| Autonomy and parents |  | 49.4 (11.5) |   |
| Peers and social support |  | 50.3 (12.7) |   |
| School environment |  | 49.0 (6.6)4 |   |
| Total score |  | 48.8 (9.2)4 |   |
| KIDSCREEN-27 (child-reported) (Masnari *et al.*) | Physical well-being |   | 51.2 (12.5) | NL | Could not be derived  |
| Psychological well-being |  | 49.1 (9.9)4 |   |
| Autonomy and parents |  | 54.7 (11.2) |   |
| Peers and social support |  | 54.0 (9.0) |   |   |
| School environment |  | 53.5 (6.8) |   |   |
| Total score |   | 52.5 (6.5) |   |   |
| CBCL/1.5-5 (Masnari *et al.*) | Internalizing |  | 48.2 (8.8) | NL | Could not be derived |
| Externalizing |  | 48.1 (9.5) |   |
| Total behavior score |  | 48.1 (9.7) |   |   |
| CBCL/4-18 (Masnari *et al.*) | Internalizing |   | 49.8 (11.1) | NL | Could not be derived |
| Externalizing |   | 50.5 (11.5) |   |
| Total behavior score |   | 50.9 (11.5) |   |   |
| Floor and ceiling effects were defined as ≤ 15% of the study population receiving the lowest or highest possible score, respectively. No study reported information on (change) scores for relevant (sub-)groups, minimal important change/difference, or response shift. 1 Mean score of multiple assessors. 2 Range of scores of 3 observers. 3 Scores for all 12 subscales can be found in Masnari *et al.* (2013). 4 Significantly impaired scores compared to reference populations. Abbreviations: IH, infantile hemangioma; NA, not available; NL, not listed; OMI, outcome measurement instrument; PWS, port wine stain. |