Treatment Outcome Measurement Instruments for Port Wine Stains: A Systematic Review of Their Measurement Properties

M. Ingmar van Raath, Sandeep Chohan, Albert Wolkerstorfer, Chantal M.A.M. van der Horst, Jacqueline Limpens, Xuan Huang, Baoyue Ding, Gert Storm, Rene R.W.J. van der Hulst, Michal Heger

Department of Plastic, Reconstructive and Hand Surgery, Maastricht University Medical Center, Maastricht University, Maastricht, The Netherlands; NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Center, Maastricht University, Maastricht, The Netherlands; Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; Department of Pharmaceutics, Jiaxing Key Laboratory for Photonanomedicine and Experimental Therapeutics, College of Medicine, Jiaxing University, Jiaxing, PR China; Department of Dermatology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; Department of Plastic, Reconstructive and Hand Surgery, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

Keywords
Capillary malformation · Clinical assessment · Clinimetric study · Psychometric evaluation · Quality of life

Abstract
Background: A plethora of outcome measurement instruments (OMIs) are being used in port wine stain (PWS) studies. It is currently unclear how valid, responsive, and reliable these are. Objectives: The aim of this systematic review was to appraise the content validity and other measurement properties of OMIs for PWS treatment to identify the most appropriate instruments and future research priorities. Methods: This study was performed using the updated Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) methodology and adhered to PRISMA guidelines. Comprehensive searches in Medline and Embase were performed. Studies in which an OMI for PWS patients was developed or its measurement properties were evaluated were included. Two investigators independently extracted data and assessed the quality of included studies and instruments to perform qualitative synthesis of the evidence. Results: In total, 1,034 articles were screened, and 77 full-text articles were reviewed. A total of 8 studies were included that reported on 6 physician-reported OMIs of clinical improvement and 6 parent- or patient-reported OMIs of life impact, of which 3 for health-related quality of life and 1 for perceived stigmatization. Overall, the quality of OMI development was inadequate (63%) or doubtful (37%). Each instrument has undergone a very limited evaluation in PWS patients. No content validity studies were performed. The quality of evidence for content validity was very low (78%), low (15%), or moderate (7%), with sufficient comprehensibility, mostly sufficient comprehensiveness, and mixed relevance. No studies on responsiveness, minimal important change, and cross-cultural validity were performed. Conclusions: The available data indicate the need for more rigorous evaluation of the existing OMIs and development of new OMIs to better assess the treatment outcomes of PWS patients. PROSPERO Registration No.: CRD42019119252.
Port wine stains (PWS) are congenital capillary malformations resulting from differentiation-impaired endothelial cells with a progressive dilation of immature, venule-like vasculature [1, 2]. These lesions occur in 0.3–0.5% of the population, enlarge and darken proportionally to age [3–5], and can present a significant psychological burden [6, 7]. Pulsed dye laser (PDL) treatment is the current gold standard but fails to achieve optimal results in a significant proportion of patients [8].

Although much effort has been devoted to developing technical tools to objectively quantify PWS blanching [9–12], (physician- or patient-reported) clinical outcome measurement instruments (OMIs) appear to have undergone far less scrutiny, despite the fact that robust and uniform clinical PWS OMIs are crucial in clinical trials to steer the field towards progress. To date, progress has not materialized [8] notwithstanding the medical need and patient demand [13]. A recent systematic review of all prospective PWS studies performed since 2005 by our group unveiled the wide variety of clinical scoring systems [14], a phenomenon that has hampered study comparisons and meta-analyses. An appropriate OMI must have good measurement properties: it has to be valid (i.e., it measures what it purports to measure), reliable, and responsive to change. This information is derived from (high-quality) clinimetric studies.

The aim of this systematic review was to critically appraise the content validity and other measurement properties of all patient-, parent-, and physician-reported OMIs for the evaluation of clinical outcome or life impact of PWS treatments using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) framework [15]. Secondary aims were to establish the most appropriate OMI for measuring effectiveness in PWS and to identify areas for future research.

Introduction

The aim of the included studies had to be the development of an OMI or the evaluation of one or more of its measurement properties. The studies had to have either an observer-/clinician-reported OMI or patient-/parent-reported outcome measure (PROM) for the evaluation of effectiveness in PWS. This included clinical OMIs as well as measures of life impact (in accordance with COMET classification; life impact includes the domains (psychosocial) functioning, perceived health status, and quality of life) [18], even if these OMIs had not yet been used as an outcome of effectiveness but could be so used (i.e., cross-sectional studies were also eligible). Further exclusion and inclusion criteria as well as data extraction procedures are addressed in the online supplementary Methods for purposes of brevity.

The updated COSMIN methodology was employed to assess the measurement properties (listed and defined in accordance with COSMIN taxonomy in online suppl. Table 2) [15, 23–26]. Although the COSMIN checklists have been developed primarily for the evaluation of PROMs, they are widely implemented for clinician-reported OMIs too [27–29]. Accordingly, some standards were ignored or adapted where necessary (online suppl. Methods). Studies on the comprehensibility of included OMIs in other, yet similar populations were also retrieved and considered inasmuch as these can provide meaningful evidence.

For each included OMI the original development study, manual, and/or additional resources were also retrieved online and by contacting authors (online suppl. Methods) and evaluated because proper item construction during development helps to ensure content validity. It was decided a priori that there is no gold standard in the assessment of PWS, therefore criterion validity and criterion approach for responsiveness could not be assessed (except to compare short versions to the original (long) questionnaires). For the assessment of construct validity, the generic COSMIN hypotheses were applied [15]. In order to help establish whether OMIs adhered to a formative and/or a reflective model, we used the checklist of Fleuren et al. [30].

Assessment of the methodological quality of the OMI’s development study, additional content validity studies, and studies on measurement properties was performed independently by 2 reviewers (M.I.R. and S.C.) using the corresponding COSMIN risk of bias checklists [23] (and rated “very good,” “adequate,” “doubtful,” or “inadequate”). A score per section per study was determined using the lowest rating of any item. The content validity and measurement properties themselves were assessed using the predefined COSMIN criteria for good content validity and the updated criteria for good measurement properties, respectively, and
rated as sufficient (+), insufficient (−), or indeterminate (?) (online suppl. Table 2) [31, 32]. For all reviewer ratings of content validity, an expert panel comprising a plastic surgeon (C.M.A.M.H.) and dermatologist (A.W.) was consulted.

For each OMI the measurement properties and corresponding quality of evidence were summarized and pooled if possible. The overall ratings for measurement properties were rated as sufficient (+), insufficient (−), indeterminate (?), or inconsistent (±) [32]. Then, the overall rating of the quality of evidence was given using the modified GRADE approach (“high,” “moderate,” “low,” or “very low” quality of evidence) [33].

Results

Study Characteristics

The search retrieved 1,030 unique hits (Fig. 1). After reviewing 77 full-text articles, 8 studies were included [34–41]. Reference checking yielded 4 additional eligible articles: 1 was not analyzed because it was a preliminary version of another included study [42] and 3 articles with some clinimetric assessment (albeit limited in reporting and/or quality) of self-constructed questionnaires regarding PWS-related stress and stigmatization were excluded because the questionnaires proved impossible to source [6, 43, 44]. Searches for studies on the comprehensibility of included OMIs performed in other populations yielded 1 additional study [45].

The characteristics of included studies are presented in Table 1. In these studies, 6 physician-reported clinical OMIs and 6 PROMs of life impact were evaluated (summarized in Table 2).

Included Clinical (Physician-Reported) OMIs

Koster et al. [39] have developed the most comprehensive PWS-specific instrument (Table 2). This physician-reported questionnaire contains 8 items (PWS color, patchiness, boundary, pigmentation, size, shape, surface, and hypertrophy) and requires pre- and post-treatment measurement.

Sajan et al. [37] developed a questionnaire for facial infantile hemangiomas and PWS in children that covers...
Table 1. Characteristics of the included studies and study populations

| OMI(s) | Reference | Population | Disease characteristics | Study design | Instrument administration |
|--------|-----------|------------|-------------------------|--------------|---------------------------|
| Koster’s PWS questionnaire | Koster et al. [39] (1998) | 70 | 11.4 (8.6–0–30) years | Prospective | PDL (5 sessions) |
| Sajan’s facial PWS and IH questionnaire | Sajan et al. [37] (2013) | 22 | Mean 10 months (exclusively pediatric patients) | Retrospective | PDL (average 3–6 sessions) |
| Poor/fair/good/excellent; 0–24; 25–49; 50–74 and 75–100% lightening; class 0–IV (Achauer) | | | | | |
| Percentage PWS clearance | Pier et al. [36] (1997) | 80 | NL | Prospective | PDL (1 or several sessions) |
| Percentage PWS treatment success | Szychta et al. [35] (2013) | 48 | Adults | Retrospective | PDL |
| Lighter/darker/did not change | Naam et al. [38] (2008) | 21 | Median 5 (range 1–15) years | Prospective | PDL |
| DLQI | Wang et al. [40] (2017) | 197 | >16 years, 29.4% >60 years | Cross-sectional | Questionnaires |
| DLQI¹ | Safikhani et al. [45] (2013) | 21 | 48.8 (16.1, 18–85) years | Prospective | Questionnaires |
| PSQ (child-² and parent-reported), CBCL/1.5–5², CBCL/4–18², TAPQOL², KIDSCREEN-27 | Marani et al. [41] (2013) | 88 | 6.91 (466.6, 75–1575) years | Face | Questionnaires |

CBCL, Child Behavior Checklist; DLQI, Dermatology Life Quality Index; IH, infantile hemangioma; NA, not applicable; NL, not listed; OMI, outcome measurement instrument; PSQ, Perceived Stigmatization Questionnaire; PDL, pulsed dye laser; PWS, port wine stain; SD, standard deviation; SWS, Sturge-Weber syndrome; TAPQOL, TNO-AZL questionnaire for preschool children’s quality of life; 1 Study performed in non-PWS patients, included because it provides evidence on comprehensibility. ² Questionnaires were completed in an age-specific subset of patients only.
**Table 2. Characteristics of the included outcome measurement instruments**

| OMI (presenting reference) | Construct(s) | Target population | Mode of administration | Recall period | (Sub)scale(s) (number of items) | Response options | Range of scores/scoring options | Original language | Available translations | Completion time | Licensing/costs |
|----------------------------|--------------|-------------------|------------------------|--------------|--------------------------------|------------------|---------------------------------|------------------|---------------------|----------------|----------------|
| Koster’s PWS questionnaire (Koster et al. [39]) | PWS color, patchiness, boundary, hypo-/hyperpigmentation, size, shape, surface, hypertrophy | PWS patients treated with PDL | Physician-reported | NA | (photograph-based) | 8 | Varies per question | Total score | English | None | 2–5 min | Free |
| Sajan’s PWS and IH questionnaire (Sajan et al. [37]) | Improvement in color, thickness, and size, side effects (atrophy, scarring, and hyperpigmentation) | IH and PWS patients treated with PDL | Physician-reported | NA | (photograph-based) | 6 | 0, 1–24, 25–49, 50–75, 76–99, 100% (for improvement in color and thickness) | Individual item scores (no total score) | English | None | 1–2 min | Free |
| Lighter/darker/did not change (Naran et al. [38]) | PWS color improvement | PWS patients treated with PDL | Physician-reported | NA | (photograph-based) | 1 | Lighter, darker, did not change | | English | NA | 1 min | Free |
| Poor/fair/good/excellent (Currie and Monk [34]) | Overall PWS response | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | Poor (defined as almost no change); fair (partly cleared); good (much improved); or excellent (essentially gone) | | English | NA | 1–2 min | Free |
| Percentage lightening [ordinal] (Currie and Monk [34]) | Degree of PWS lightening/success | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | 0–24%, 25–49%, 50–74%, 75–100% lightening | | English | NA | 1 min | Free |
| Percentage clearance [contin.] (Pérez et al. [36]) | PWS clearance | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | 0–100% clearance | | English | NA | 1 min | Free |
| Percentage clearance [ordinal] (Pérez et al. [36]) | PWS clearance | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | 0–24% (absence of clearance); 25–49% (poor clearance); 50–74% good clearance; and 75–100% (excellent clearance) | | English | NA | 1 min | Free |
| Percentage success [contin.] (Spychala et al. [31]) | PWS clearance | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | 0–100% success | | English | NA | 1 min | Free |
| Class 0–IV (Achauer et al., modified by Currie and Monk [34]) | Overall PWS response | PWS patients/undefined | Physician-reported | NA | (photograph-based) | 1 | Class 0 (essentially gone); class I (faint, barely discernible borders); class II (well-defined borders with areas of normal skin interspersed within the lesion); class III (well-defined borders, solid lesion with no areas of normal skin); or class IV (all of class III plus raised and thickened). This was categorized as: class 0; excellent; class I, excellent; class II, good; class III, fair; class IV, poor | | English | NA | 1–2 min | Free |
| OMI (presenting reference) | Construct(s) | Target population | Mode of administration | Recall period | (Sub)scale(s) (number of items) | Response options | Range of scores/scoring options | Original language | Available translations | Completion time | Licensing/ costs |
|----------------------------|--------------|-------------------|------------------------|---------------|--------------------------------|-----------------|-------------------------------|-----------------|-----------------------|----------------|----------------|
| DLQI (Finlay et al. [47])  | Dermatology-specific quality of life | Dermatological patients | Patient-reported | 1 week | 10 | Very much, a lot, a little, not at all | 0-1, no effect; 2-5, small effect; 6-10, moderate effect; 11-20, very large effect; 21-30, extremely large effect | English | >110 translations | 1-2 min | Free for most purposes |
| PSQ (Lawrence et al. [48]) | Perceived stigmatization | People with physical distinction | Parent-/patient-reported | 1 year | 3 (21) | Never, almost never, sometimes, often, always | Total score and subscores (1-5) for absence of friendly behavior, confused/staring behavior, and hostile behavior | English | German, Portuguese | <5 min | Free |
| TAPQOL (Fekkes et al. [49]) | Parent-perceived HRQoL (domains: physical (incl. motor functioning*), social, cognitive, and emotional functioning) | Largely healthy children (9 months to 6 years old) | Parent-reported | 3 months | 12 (48) | Varies per question | T scores per (sub) scale | Dutch | English | 10 min | License required for commercial studies |
| KIDSCREEN-27 (Ravens-Sieberer et al. [50]) | HRQoL | Healthy and chronically ill children/adolescents (6-18 years old) | Parent-/patient-reported | 1 week | 5 (27) | Varies per question | T scores and percentiles per (sub) scale | English | German | 10–15 min | License required for commercial studies |
| CBCL/1.5–5 (Achenbach et al. [52]) | Competencies, adaptive functioning, and problems | Children aged 18 months to 5 years | Parent-/proxy-reported | 2 months | 12 (108) | 0 = not true (as far as you know), 1 = somewhat or sometimes true, 2 = Very true or often true | T scores per (sub) scale | English | >50 languages (full list at aseba.org) | 10–15 min | Fees per form apply |
| CBCL/4–18 (Achenbach et al. [52]) | Competencies, adaptive functioning, and problems | Children aged 4–18 years | Parent-/proxy-reported | 6 months | 8 (122) | 0 = not true (as far as you know), 1 = somewhat or sometimes true, 2 = Very true or often true | T scores per (sub) scale | English | >50 languages (full list at aseba.org) | 10–20 min | Fees per form apply |

*Only for children older than 18 months.
plus nodularity”) [46]. Finally, Achauer et al. also included class V (“all of class IV to poor, fair, good, or excellent (Table 2). Note that original categories (percentage “lightening,” “improvement,” “success,” etc.) and reported as continuous variables or ranges (usually 0–24, 25–49, 50–74, and 75–100%). As these constitute highly intuitive measures, no published development paper exists. Their measurement properties have also been investigated in the studies of Pérez (percentage clearance; both continuous and ordinal scale) and Szychta (percentage success; continuous scale) [35, 36]. Although it could be argued that the construct “color lightening,” “improvement,” or “blanching” is much narrower than “improvement” or “success,” in this review these were considered equivalent and the data were pooled depending on the scales.

The second OMI in Currie’s study assessed percentage lightening. A variety of percentage improvement scales are commonly used in PWS trials and given distinct designations (percentage “lightening,” “improvement,” “success,” etc.) and reported as continuous variables or ranges (usually 0–24, 25–49, 50–74, and 75–100%). As these constitute highly intuitive measures, no published development paper exists. Their measurement properties have also been investigated in the studies of Pérez (percentage clearance; both continuous and ordinal scale) and Szychta (percentage success; continuous scale) [35, 36]. Although it could be argued that the construct “color improvement,” “lightening,” or “blanching” is much narrower than “improvement” or “success,” in this review these were considered equivalent and the data were pooled depending on the scales.

The third OMI assessed by Currie and Monk [34] was originally described by Achauer et al. [46]. Treated PWS were categorized into class 0–IV and then also converted to poor, fair, good, or excellent (Table 2). Note that originally, Achauer et al. also included class V (“all of class IV plus nodularity”) [46].

Naran et al. [38] used a very simple 3-point rating scale (“lighter,” “darker,” or “did not change”) to assess additional lightening after additional PDL treatments in children.

**Summary of Included Life Impact PROMs**

Three instruments for health-related quality of life (HRQoL), 2 for detecting emotional and behavioral problems and 1 for perceived stigmatization, were included (Table 2). None of these PROMs has yet been used as an OMI in published PWS intervention studies. The average Flesch-Kincaid grade level for readability of the English translations was 3.5 (online suppl. Table 3). These and other characteristics of OMI feasibility and interpretability are listed in online supplementary Tables 3 and 4.

The Dermatology Life Quality Index (DLQI) was the first dermatology-specific HRQoL questionnaire [47]. It consists of 10 items (symptoms and feelings, daily activities, leisure, work, and school, personal relationships, and treatment) and has been employed for a range of dermatoses. In the included study using the DLQI’s adult version, PWS patients had a small to moderate HRQoL impairment (online suppl. Table 4). A study that assessed the DLQI’s content validity in adults with psoriasis was also included because its findings regarding comprehensibility are likely to apply to PWS patients as well [45].

The 21-item Perceived Stigmatization Questionnaire (PSQ) ascertains how often people are confronted by certain stigmatizing behavior [48]. Although it was designed for people with visible distinctions, its further development was limited mostly to adult burn survivors. The patient- or parent-reported instrument yields a total score and subscores (range 1–5) for “absence of friendly behavior,” “confused/staring behavior,” and “hostile behavior.” In this study, preschool and school-age children with a facial difference (including PWS) had a PSQ score of 1.66 and 2.10, respectively (compared to, e.g., 2.2 in adult burn survivors [48]).

The TNO-AZL Questionnaire for Preschool Children’s Health-Related Quality of Life (TAPQOL) measures parent’s perception of HRQoL in children aged 9 months to 6 years and incorporates the (perceived) emotional reaction of the child to their health status problems [49]. This multidimensional questionnaire is presented primarily as an instrument for research and group-level data. In this study of children with a facial difference, TAPQOL scores were not impaired (online suppl. Table 4).

The KIDSCREEN-27 is a generic, parent- or self-reported HRQoL measure for children and adolescents (8–18 years old) that covers 5 domains (physical well-being, psychological well-being, autonomy and parents, peers and social support, and school environment) [50]. It is the short version of the KIDSCREEN-52 [51]. Both KIDSCREEN questionnaires were developed in 12 European countries to measure HRQoL in largely healthy children. The included study in children with a facial difference found statistically significant impairment of child- and parent-reported psychological well-being and parent-reported overall HRQoL and physical well-being.

The Child Behavior Checklists (CBCL) gauge parent-reported emotional and behavioral problems, competencies, and adaptive functioning in children and adolescents [52]. Throughout its existence, many revisions have been made. The CBCL/1.5–5 (the successor to the CBCL/2–3; for 18 months to 5 years old) and the CBCL/4–18 (the predecessor of the current CBCL/6–18; for 4–18 years old) were included here. Both offer 3 composite
The quality of OMI development is shown in Table 3. OMI development was rated as very good (V), adequate (A), doubtful (D), or inadequate (I). The CBCL, Child Behavior Checklist; DLQI, Dermatology Life Quality Index; IH, infantile hemangioma; OMI, outcome measurement instrument; PSQ, Perceived Stigmatization Questionnaire; PWS, port wine stain; TAPQOL, TNO-AZL questionnaire for preschool children’s health-related quality of life.

Table 3. Quality of outcome measurement instrument development

| OMI                        | OMI design | Concept elicitation | Total OMI development |
|----------------------------|------------|---------------------|-----------------------|
|                            | clear construct | clear target population | OMI developed in sample representing the target population |
| Koster’s PWS questionnaire | V          | V                   | V                     |
| Sajan’s PWS and IH questionnaire | V          | D                   | V                     |
| DLQI                       | V          | V                   | V                     |
| PSQ                        | V          | V                   | V                     |
| TAPQOL                     | V          | V                   | V                     |
| KIDSCREEN-27              | V          | V                   | V                     |
| CBCL/1-5                   | V          | V                   | V                     |
| CBCL/4-18                  | V          | V                   | V                     |

Development was rated as very good (V), adequate (A), doubtful (D), or inadequate (I). CBCL, Child Behavior Checklist; DLQI, Dermatology Life Quality Index; IH, infantile hemangioma; OMI, outcome measurement instrument; PSQ, Perceived Stigmatization Questionnaire; PWS, port wine stain; TAPQOL, TNO-AZL questionnaire for preschool children’s health-related quality of life. 1 When the patient/parent-reported outcome measure was not developed in a sample representing the target population, concept elicitation was not further rated. 2 Empty cells indicate that a cognitive interview (CI) study (or part of it) was not performed.
indeterminate for the CBCL questionnaires. All PROMs had sufficient comprehensiveness and comprehensi-

**Measurement Properties of PWS OMIs**

The quality and results of the studies on measurement properties are summarized in Table 5. Most studies were of doubtful or inadequate quality.

**Structural Validity and Internal Consistency**

Wang et al. [40] have performed principal component analysis (a form of exploratory factor analysis) of the Chinese translation of the DLQI in patients with exposed PWS and identified 2 common factors (Table 5). This study was found to be of doubtful quality because no factor rotation method was applied. No criteria for good measurement properties have been defined by COSMIN for exploratory factor analysis, resulting in an indeterminate score (and low-grade quality of evidence score; Tables 5, 6). Because of the lack of evidence for sufficient structural validity and calculation of only an overall internal consistency despite suggesting a two-dimensional model (internal consistency can only be interpreted within a unidimensional scale), internal consistency of the DLQI was also scored as indeterminate (very low quality of evidence). Of note, studies in other diseases have yielded inconsistent results regarding unidimensionality of the DLQI [55–57].

Masnari et al. [41, 42] evaluated internal consistency of several PROMs in children with various facial blemishes. Note that internal consistency has no meaningful interpretation for questionnaires with a formative model. Unfortunately, not all OMIs have studied or reported their putative model (reflective vs. formative). Classifying questionnaires correctly can be difficult, and mixed models also exist. We assigned a formative model to the PSQ and CBCL, and a reflective or mixed model to the other PROMs. Even though most (sub-)scales had sufficient internal consistency, there was a lack of information on the structural validity in PWS patients resulting in an indeterminate score. The quality of evidence for these PROMs was downgraded to low or very low due to small sample sizes (imprecision) and indirectness (only 22% of patients had PWS).

**Reliability and Measurement Error**

The inter-rater reliability of Koster’s questionnaire was evaluated for a panel of 5 in children and (young) adults (Table 1). Insufficient reliability was found for all 8 items, particularly pigmentation and surface-structure. This can be explained by the low frequency of hypopigmentation and hyperpigmentation and uneven surface in the study population, as supported by the relatively high percentage of absolute agreement (average of 5 raters): 88 and 79%, respectively (Table 5). The Cronbach α-values on averaged ratings of the 5 panel members suggest that application of this method improves reliability, although
Table 5. Results and quality of studies on measurement properties

| OMI | Language in which the OMI was evaluated | Structural validity | Internal’s consistency | Reliability (inter-rater) | Measurement error (inter-rater) | Measurement error (intrarater) | Construct validity |
|-----|----------------------------------------|---------------------|------------------------|--------------------------|-------------------------------|-------------------------------|------------------|
|     |                                        | n | result (rating) | n | result (rating) | n | result (rating) | n | result (rating) | n | result (rating) | n | result (rating) | n | result (rating) |
| 0–24, 25–49; 50–74 and 75–100% lightening (Currie and Monk [34]) | English | 20 | A | Mean agreement 52.8% (range 50–62) | 20 | A |
| 0–24, 25–49; 50–74 and 75–100% clearance (Pérez et al. [36]) | NL | 80 | A | κw = 0.77 (+) | 48 | I | R = 0.81 (?) | 48 | I | TEM = 9.91% (?) |
| Percentage success (core physicians) (Szychta et al. [35]) | English | 48 | I | R = 0.63 (?) | 48 | I | TEM = 13.84% (?) | 0–24, 25–49; 50–74; 76–99, and 100% color improvement (Sajan et al. [37]) | English | 80 | A | ICC = 0.83 (+) | 22 | I | κw = 0.92 (+) |
| Percentage success (lay people) (Szychta et al. [35]) | English | 48 | I | R = 0.63 (?) | 48 | I | TEM = 13.84% (?) | 0–100% clearance (Pérez et al. [36]) | NL | 80 | A |
| 0; 1–24, 25–49; 50–75; 76–99, and 100% color improvement (Sajan et al. [37]) | English | 24 | I | ICC = 0.92 for color, thickness, size, and scarring (+), ICC = 0.70 for atrophy (+), ICC = 0.10 for hypopigmentation (-) | 70 | A | Mean κw = 0.37 (range 0.11–0.64) (-) | 70 | A | Mean % exact agreement: 47 (color), 30 (patchiness), 64 (boundary), 88 (pigmentation), 66 (size), 51 (shape), 79 (surface structure), 71 (hypertrophy). Mean 64.6% (range 47–88) (?) | Koster’s PWS questionnaire (Koster et al. [39]) | English | 70 | A |
| Sajan’s PWS and IH questionnaire (Sajan et al. [37]) | English | 24 | I | ICC = 0.92 for color, thickness, size, and scarring (+), ICC = 0.70 for atrophy (+), ICC = 0.10 for hypopigmentation (-) | 70 | A | Mean κw = 0.37 (range 0.11–0.64) (-) | 70 | A | Mean % exact agreement: 47 (color), 30 (patchiness), 64 (boundary), 88 (pigmentation), 66 (size), 51 (shape), 79 (surface structure), 71 (hypertrophy). Mean 64.6% (range 47–88) (?) | Sajan’s PWS and IH questionnaire (Sajan et al. [37]) | English | 24 | I |
| Poor/fair/good/excellent (Currie and Monk [34]) | English | 20 | A | Mean agreement 77.1% (range 48.8–96.5) (?) | 20 | A |
| Class 0–IV (Currie and Monk [34]) | English | 20 | A | Mean agreement 64.2% (range 34.7–82.6) (?) | 20 | A | Mean agreement 52.8% (range 50–62) (?) |
| OMI | Language in which the OMI was evaluated | Structural validity | Internal consistency | Reliability (inter-rater) | Measurement error (inter-rater) | Measurement error (intrarater) | Construct validity |
|-----|-----------------------------------------|---------------------|---------------------|--------------------------|-------------------------------|-------------------------------|-------------------|
|     |                                        | n meth. qual. result (rating) | n meth. qual. result (rating) | n meth. qual. result (rating) | n meth. qual. result (rating) | n meth. qual. result (rating) | n meth. qual. result (rating) |
| Lighten/darker/ did not change (Naran et al. [38]) | English | 21 1 Spearman correlations between 3 observers: 0.54 (p < 0.001), 0.41 (p < 0.01), and 0.34 (p = 0.02) (?) | 21 1 Spearman correlation between 3 observers and digital image analysis: −0.33 to −0.09 (p > 0.01) (–) |
| DLQI (Wang et al. [40]) | Chinese | 197 D CFA: 2 common factors with 47.5% cumulative explained variance (?) | 197 D Cronbach α = 0.740 (overall) (?) |
| PSQ; patient-reported (Masnari et al. [42]) | German | 34 D Cronbach α = 0.81 (?) |
| PSQ; parent-reported (Masnari et al. [42]) | German | 86 V Cronbach α = 0.88 (?) |
| TAPQOL (Masnari et al. [42]) | German | 54 V Cronbach α-values = 0.51–1.00 (mean 0.72) (?) |
| KIDSCREEN-27; parent-reported (Masnari et al. [42]) | German | 32 V Cronbach α-values = 0.48–0.92 (mean 0.78) (?) |
| KIDSCREEN-27; child-reported (Masnari et al. [42]) | German | 31 V Cronbach α-values = 0.44–0.86 (mean 0.68) (?) |
| CBCL/1.5–5 (Masnari et al. [42]) | German | 21 V Cronbach α-values = 0.73 (internalizing), 0.90 (externalizing), 0.92 (total) (?) |
| CBCL/4–18 (Masnari et al. [42]) | German | 51 V Cronbach α-values = 0.67 (internalizing), 0.89 (externalizing), 0.95 (total) (?) |

Study quality was very good (V), adequate (A), doubtful (D), or inadequate (I). Measurement properties were rated as sufficient (+), insufficient (−), inconsistent (×), or indeterminate (?) Results were pooled if possible. CBCL, Child Behavior Checklist; CFA, common factor analysis; DLQI, Dermatology Life Quality Index; ICC, intraclass correlation coefficient; IH, infantile hemangioma; n, weighted data; NL, not listed; OMI, outcome measurement instrument; PSQ, Perceived Stigmatization Questionnaire; PWS, port wine stain; SWS, Sturge-Weber syndrome; TAPQOL, TNO-AZL questionnaire for preschool children’s health-related quality of life; TEM, technical error of measurement. 1 Mean agreement for all outcome categories for all 6 assessors. 2 The results of Szychta et al. for reliability could not be pooled because it was reported using an incompatible parameter. 3 PWS (n = 5, of which 1 with SWS) and IH (n = 19) patients.
Table 6. Overall quality of evidence per measurement property

| Outcome measurement instrument: | Class 0–IV DLQI (Chinese version) | TAPQOL (German version) | KIDSSCREEN-27 (parent- and self-reported; German version) | PSQ (parent- and self-reported; German version) | CBCL/1.5-5 (German version) | CBCL/4-18 (German version) |
|---------------------------------|-----------------------------------|-------------------------|----------------------------------------------------------|-------------------------------------------------|-----------------------------|-----------------------------|
| Ref(s):                         | Wang et al. [40], Safikhani et al. [45] | Masnari et al. [42]     | Masnari et al. [42]                                      | Masnari et al. [42]                              | Masnari et al. [42]         | Masnari et al. [42]         |
| Overall rating                  | +                                | +                       | +                                                        | +                                               | +                           | +                           |
| Quality of evidence             | very low                         | low                     | low                                                     | low                                             | very low                    | very low                    |

Measurement properties were rated sufficient (+), insufficient (–), inconsistent (±), or indeterminate (?). The level of evidence was graded according to the modified GRADE approach. CBCL, Child Behavior Checklist; DLQI, Dermatology Life Quality Index; IH, infantile hemangioma; PSQ, Perceived Stigmatization Questionnaire; PWS, port wine stain; Ref(s), reference(s); TAPQOL, TNO-AZL questionnaire for preschool children’s health-related quality of life. 1 Sufficient for all items except “hypopigmentation.”
this would have to be properly evaluated in a separate study. Agreement for the other items varied between 47 and 71% (Table 5). Unfortunately, no criteria for good measurement error have been defined by COSMIN for categorical parameters (i.e., percentage agreement) [15], and no alternative criteria were found in the literature resulting in an indeterminate score.

The questionnaire of Sajan was used in PDL-treated children with PWS and infantile hemangiomas. The inter-rater reliability was sufficient for all items except hypopigmentation. Because of the small number of (PWS) patients, this study provided a very low quality of evidence.

Currie and Monk [34] intended to assess intrarater reliability of 3 common PWS OMIs after 1 month but calculated mean agreement (“concordance”), which is a measure of measurement error (see above). Because the exact distribution of scores is unknown, a post hoc calculation of reliability was impossible. Although no criteria for this parameter of measurement error have been defined, the 3 OMIs can be compared to each other. OMI 1 (poor, fair, good, or excellent) had the best overall results (67–89% mean agreement per outcome category; Table 5). OMI 2 (percentage lightening) performed less but was the most consistent across all 4 categories (50–62%). OMI 3 (class 0–IV) performed relatively well (52–76%) for “excellent,” “good,” and “poor” outcomes but exceptionally poorly (17%) in “fair” outcomes. For all 3 OMIs concordance was highest for the best and worst outcome and lowest in the intermediate categories.

Szychta et al. [35] assessed inter-rater reliability and measurement error of “percentage success” after PDL treatment for both core physicians as well as lay people. For this, they used parameters typically used in anthropometry for which also no COSMIN criteria for good measurement properties have been defined (TEM, technical error of measurement, and R, coefficient of reliability). Because TEM was calculated instead of the standard error of measurement (SEM; the gold standard), this study received an inadequate score. Anthropometric literature suggests an $R \geq 0.95$ as sufficient [58], which would yield an insufficient score for the current results for both core physicians and lay people (Table 5). The inter-rater TEM values (9.9% for expert physicians) appear acceptable but, because TEM is not the same as SEM, it is difficult to interpret the results [59]. In addition, no information is available on the minimal important change in PWS patients.

Pérez et al. [36] investigated inter-rater reliability of a 0–100% clearance scale among specialized dermatologists (Table 1). The continuous score was also converted to 0–24, 25–49, 50–74, and 75–100% clearance. The study was of only adequate quality because it could only be assumed that test conditions were similar for all 3 assessors. Both the categorized and continuous scores had sufficient reliability (Table 5). Overall, there was low quality of evidence due to the limited sample size and quality of the study.

If the percentage improvement scales are regarded as equivalent and the results of Sajan’s (percentage color improvement) and Pérez’ (percentage clearance) studies are pooled (even though Sajan uses 2 extra categories for 0 and 100% color improvement), there is moderate evidence for sufficient reliability (Tables 5, 6). Because of the differences in scales, no meta-analysis was performed.

Naran et al. [38] used correlation analysis to assess the inter-rater reliability of its OMI. However, correlations are not appropriate for this purpose, particularly because the OMI is an ordinal score.

Construct Validity

The OMI of Naran et al. was compared to ΔE (i.e., the change in color difference in comparison to normal skin) derived from digital image analysis. Because this study lacked a predefined hypothesis, it was considered to be of inadequate quality. Spearman correlations between the 3 observers and digital image analysis varied from −0.33 to −0.09 ($p > 0.01$), which is much smaller than what would be expected (Tables 5, 6).

**COSMIN Recommendations**

In line with COSMIN guidelines, instrument recommendations are provided below based on 2 conclusions of this review: (1) no OMI with evidence for sufficient content validity (any level) and at least low-quality evidence for sufficient internal consistency was included (this OMI would be recommended for use and produce results that can be trusted), and (2) no OMI with high-quality evidence for an insufficient measurement property was found (this OMI would not be recommended for use). Accordingly, all included instruments have the potential to be included as an OMI for PWS treatment but would require further assessment of quality. Based on content validity, the PSQ and DLQI are provisionally recommended. No clinical OMI can be recommended.

**Discussion**

This study was the first to systematically review the measurement properties of PWS OMIs. We have identified 6 physician-reported clinical OMIs and 6 PROMs of
life impact. Each has undergone very limited evaluation in PWS patients. Although not all elements of the COSMIN are applicable for all OMIs, each OMI has important data missing. No study addressed responsiveness, minimal important change, or cross-cultural validity. The quality of the development (if any) of PWS-specific OMIs was inadequate because substantial methods were used to generate relevant items and assess the comprehensibility and comprehensiveness of the OMI. Most importantly, no study has asked PWS patients (or their parents) what matters to them, nor have professionals been adequately interviewed about relevant items. Together with the small number of PWS patients this generally resulted in low levels of evidence (quality was very low and low for 74 and 21% of all evidence, respectively). No study has assessed the content validity of OMIs originally developed for other populations and now used in PWS patients.

Considering the lack of high-quality evidence, it was impossible to recommend a clinical OMI. Koster’s questionnaire is clearly the most elaborate clinical OMI. This approach is justified at least in part by another study of Koster et al. [60], in which Koster’s questionnaire and a disfigurement score were used to study the disfiguring effect of PWS on the head and neck. According to physicians, PWS size, color, and boundary contributed the most to disfigurement, underscoring their relevance in clinical OMIs. The inter-rater reliability of Koster’s questionnaire was insufficient. However, in evaluative studies measurement error is much more important than reliability [61]. This is based on the fact that reliability refers to how well patients can be discriminated, that is, the ability of an instrument to distinguish patients from each other despite measurement error [62]. As a result, reliability could be low because variability in the study population is low. The measurement error (percentage agreement) was <70% for a majority of items, indicating substantial measurement error. The measurement properties of Koster’s questionnaire have not been investigated further since, nor has the instrument been employed in a published trial. The relevance of Sajan’s questionnaire [37] suffered from the fact that it was also designed for infantile hemangiomas and included a question on new atrophy, which was considered irrelevant for PWS. Consequently, it did not meet the COSMIN criterion (≥85% relevant items). Perhaps this criterion is too strict for such brief questionnaires. Nevertheless, this questionnaire could be easily modified and retested. This instrument has not yet been used in a published PWS trial either. The OMI of Naran et al. [38] was considered to be far too limited.

The most common OMI in recent PWS trials is a percentage improvement or lightening, also referred to as blanching, clearance, or other similar terms [14]. It is unclear whether these terms are actually interchangeable. For example, a very lightened but patchy lesion may score better for “lightening” than for “success.” Unfortunately, the content validity of these methods has not been investigated. When all studies are taken together there is moderate evidence for sufficient inter-rater reliability (when performed by experienced physicians). Note that these data are derived almost exclusively from patients with a PWS in the face or neck and may not apply to lesions elsewhere. When the percentages are categorized it is important to be aware of floor and ceiling effects (online suppl. Table 4), as these may obscure treatment superiority. The classification of Perez [36] suffers from such a ceiling effect, which could be prevented by adding another top category (as in Sajan’s questionnaire [37]). Unfortunately, the data on measurement error could not be compared to COSMIN criteria but the results do indicate that experienced physicians perform better than lay people.

In order to achieve treatment results that most closely align with what means the most for patients, it is important to measure corresponding outcomes, that is, PROMs. The emphasis on and inclusion of such patient-reported data has increased greatly, also in dermatology [63, 64], and helps to provide patient-centered care and guide clinical decision making. In terms of PROMs for PWS patients, many studies have demonstrated a significant psychosocial impact of PWS, stigmatization, and reduced HRQoL [6, 7, 43, 44, 53, 65]. However, the use of PROMs in PWS trials is limited [14]. In this study no clinimetric studies of PWS-specific PROMs or a (generic) PROM used as an outcome for a PWS intervention trial were found. Because PROMs not yet used in effectiveness studies were also eligible, we included 6 PROMs. The assessment of these PROMs was very restricted (limited to structural validity and/or internal consistency). Additionally, the information on structural validity was of poor quality or absent, which hampered the interpretability of the internal consistency. The sufficient overall content validity of the DLQI and PSQ do justify a provisional recommendation. The relevance of the TAPQOL, KIDSCREEN-27, and CBCL questionnaires was less certain, primarily because too many questionnaire items were deemed irrelevant for PWS patients as supported by the lack of aberrant TAPQOL and CBCL scores (Table 1 and online suppl. Table 4).

This review was limited in several ways. First, 3 studies of 3 PROMs could not be included because the question-
naires themselves were not accessible. However, these studies were of low quality and would have provided little evidence (data not shown). Second, some OMIs were developed prior to the introduction of current methodological standards. Because of the COSMIN’s worst score counts principle, this negatively impacted OMI development and content validity scores for older OMIs. Third, even though the COSMIN methodology specifies many criteria and possibilities, many questions (particularly for content validity) still require some form of subjective assessment. Other assessors may thus arrive at slightly different results. It should also be emphasized that this review is not a comprehensive index of all OMIs used in PWS studies, as OMIs without any assessment of measurement properties are not covered in this methodology.

Finally, it is important that we address the lack of evidence for the measurement properties of PWS OMIs because this hinders outcome selection, especially since PWS OMIs have thus far been highly heterogeneous [14]. Ideally, a PWS core outcome set is developed. This would drastically improve study quality and comparability across trials and enable interstudy comparison, which is direly needed considering the paucity of high-quality PWS trials [14]. First, consensus on the core domains (constructs) measured in therapeutic trials would need to be reached in a structured Delphi process that includes all stakeholders. Subsequently, new OMIs need to be developed, or existing OMIs could be repurposed. Finally, further assessment of all relevant measurement properties needs to be performed.

Conclusions

There was insufficient evidence to recommend a single clinical OMI or PROM for treatment of PWS. More research into the measurement properties of clinical OMIs and PROMs is needed, preferentially guided by the establishment of a set of core domains.

Key Message

The current literature provides insufficient evidence to guide outcome measurement instrument selection for port wine stains.

Acknowledgments

We are grateful to the COSMIN team for their assistance in performing this review.

Conflict of Interest Statement

C.M.A.M.H. co-authored one of the included papers (Koster’s questionnaire) and was therefore excluded from review decisions regarding this paper. The authors have no other conflicts of interest to declare.

Funding Sources

This project was supported by a Stichting Technologische Wetenschap (STW) valorization grant (12064) and a preseed grant from the Academic Medical Center SKE Fund (Techностarter 20090812) to M.H. M.H. is currently supported by grants from the Dutch Cancer Foundation (KWF project 10666), National Natural Science Foundation of China (81872220), a Zhejiang Provincial Foreign Expert Program Grant, Zhejiang Provincial Key Natural Science Foundation of China (Z20H160031), and a grant for the establishment of the Jiaxing Key Laboratory for Photonanomedicine and Experimental Therapeutics. B.D. is supported by a basic public welfare research project of Zhejiang Province (LGF18H160034). Site-specific pharmaco-laser therapy, which is an experimental treatment modality for PWS and a core research activity of M.H. that thematically inspired this COSMIN study, was patented (M. Heger. Drug delivery system for use in the treatment of vascular and vessel-related pathologies; EP09151332.5, PCT/EP2010/050833). M.H. is cofounder of Nurish.Me and has equity in that company (whose business activities are unrelated to the present work).

Author Contributions

M.I.R., S.C., A.W., C.M.A.M.H., M.H., and J.L. contributed to the study design. J.L. performed all searches. M.I.R. and S.C. conducted qualitative analysis. Reviewer decisions were discussed by M.I.R., S.C., C.M.A.M.H., and A.W. All authors participated in the drafting and writing of the manuscript, contributed to the interpretation and critical revision of the manuscript, and gave final approval of this paper.

References

1 Tan W, Wang J, Zhou F, Gao L, Yin R, Liu H, et al. Coexistence of Eph receptor B1 and ephrin B2 in port-wine stain endothelial progenitor cells contributes to clinicopathological vasculature dilatation. Br J Dermatol. 2017 Dec;177(6):1601–11.
2 Nguyen V, Hochman M, Mihm MC Jr, Nelson JS, Tan W. The Pathogenesis of Port Wine Stain and Sturge Weber Syndrome: Complex Interactions between Genetic Alterations and Aberrant MAPK and PI3K Activation. Int J Mol Sci. 2019 May; 20(9): 2243.
3 Jacobs AH, Walton RG. The incidence of birthmarks in the neonate. Pediatrics. 1976 Aug;58(2):218–22.
4 Pratt AG. Birthmarks in infants. AMA Arch Derm Syphilol. 1953 Mar;67(3):302–5.
5 van Drooge AM, Beek JF, van der Veen JP, van der Horst CM, Wolkerstorfer A. Hyper-trophy in port-wine stains: prevalence and patient characteristics in a large patient cohort. J Am Acad Dermatol. 2012 Dec;67(6):1214–9.
6 Augustin M, Zschoch J, Wiek K, Peschen M, Vanscheidt W. Psychosocial stress of patients with port wine stains and expectations of dye laser treatment. Dermatology. 1998;197(4):353–60.
7 Hagen SL, Grey KR, Korta DZ, Kelly KM. Quality of life in adults with facial port-wine stains. J Am Acad Dermatol. 2017 Apr;76(4):695–702.
8 van Raath MI, Chohan S, Wolkerstorfer A, van der Horst CM, Storm G, Heger M. Port wine stain treatment outcomes have not improved over the past three decades. J Eur Acad Dermatol Venereol. 2019 Jul;33(7):1369–77.
9 Al-Janabi MH, Ismaeel Ali NT, Mohamed Al-Naqib KD, Al-Dhalimi MA, Abdul Wahid SN. A new imaging technique for assessment of the effectiveness of long pulse Nd:YAG 532 nm laser in treatment of facial port wine stain. J Cosmet Laser Ther. 2017 Nov;19(7):418–21.
10 Jung B, Kim SC, Choi B, Kelly KM, Nelson JS. Use of erythema index imaging for systematic analysis of port wine skin stain response to laser therapy. Lasers Surg Med. 2005 Sep;37(3):186–91.
11 Lister T, Wright P, Chappell P. Spectrophotometers for the clinical assessment of port-wine stains: a method of assessment. Br J Plast Surg. 1989 Mar;42(2):125–32.
12 Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol. 2018 Apr;96:84–92.
13 Quaba AA. Results of argon laser treatment of port-wine stains: a method of assessment. Br J Plast Surg. 1989 Mar;42(2):125–32.
14 Ginsbach G. A Tool for the Evaluation of Colour in Port Wine Stains. Lasers Med Sci. 1991;6(1):49–52.
15 Rah DK, Kim SC, Lee KH, Park BY, Kim DW. Objective evaluation of treatment effects on port-wine stains using L*a*b* colour coordinates. Plast Reconstr Surg. 2001 Sep;108(4):842–7.
16 Ouzzani M, Hammad H, Fedorowicz Z, Elmagarmid A. Rayyan: a web and mobile app for systematic reviews. Syst Rev. 2016 Dec;5(1):210.
17 Terveen CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018 May;27(5):1171–9.
18 Terveen CB, Prinsen CA, Chiarotto A, De Vet HC, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs: User manual version 1.0. 2018. pp. 1–72.
19 Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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 Jul;63(7):737–45.
20 Dobson F, Choi YM, Hall M, Hinman RS. A new imaging technique for assessment of port-wine stains to laser treatment be reliably assessed using subjective methods? Br J Dermatol. 2000 Aug;143(2):360–4.
21 Szychta P, Al-Nakib K, Anderson W, Stewart K, Quaba A. Quantitative method for evaluation of aesthetic results after laser treatment for birthmarks. Lasers Med Sci. 2013 Nov;28(6):1567–72.
22 Pérez B, Abraira V, Núñez M, Boixeda P, Perea Corral F, Ledo A. Evaluation of agreement among dermatologists in the assessment of the color of port wine stains and their clearance after treatment with the flashlamp-pumped dye laser. Dermatology. 1997;194(2):127–30.
23 Sajan JA, Tibesar R, Jabbour N, Lander T, Hilger P, Sidman J. Assessment of pulsed-dye laser therapy for pediatric cutaneous vascular anomalies. JAMA Facial Plast Surg. 2013 Nov-Dec;15(6):434–8.
24 Naran S, Gilmore J, Deleyiannis FW. The assessment of port-wine stains in children following multiple pulsed-dye laser treatments. Ann Plast Surg. 2008 Apr;60(4):426–30.
25 Koster PH, Bossuyt PM, van der Horst CM, Gijsbers GH, van Gennip AJ. Assessment of clinical outcome after flashlamp pumped pulsed dye laser treatment of portwine stains: a comprehensive questionnaire. Plast Reconstr Surg. 1998 Jul;102(1):42–8.
26 Wang J, Zhu YY, Wang ZY, Yao XH, Zhang LF, Lv F, et al. Analysis of quality of life and influencing factors in 197 Chinese patients with port-wine stains. Medicine (Baltimore). 2017 Dec;96(51):e9446.
27 Masnari O, Schiesl C, Rössler J, Gütlein SK, Neuhaus K, Weibel L, et al. Stigmatization predicts psychological adjustment and quality of life in children and adolescents with a facial difference. J Pediatr Psychol. 2013 Mar;38(2):162–72.
28 Masnari O, Landolt MA, Roessler J, Weingärtner SK, Neuhaus K, Meuli M, et al. Self- and parent-perceived stigmatization in children and adolescents with congenital or acquired facial differences. J Plast Reconstr Aesthet Surg. 2012 Dec;65(12):1664–70.
43 Lanigan SW, Cotterill JA. Psychological disabilities amongst patients with port wine stains. Br J Dermatol. 1989 Aug;121(2):209–15.

44 Troilius A, Wrangsjö B, Ljunggren B. Patients with port-wine stains and their psychosocial reactions after photothermolysis treatment. Dermatol Surg. 2000 Mar;26(3):190–6.

45 Safikhani S, Sundaram M, Bao Y, Mulani P, Revicki DA. Qualitative assessment of the content validity of the Dermatology Life Quality Index in patients with moderate to severe psoriasis. J Dermatol Treat. 2013 Feb; 24(1):30–9.

46 Achauer BM, Vander Kam VM, Padilla JF. Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. J Abnorm Child Psychol. 1987 Dec;15(4):629–50.

47 Achenbach TM, Edelbrock C, Howell CT. 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; 86(3): 284–5.

48 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 Mar; 18(1):106–11.

49 Fekkes M, Theunissen NC, Brugman E, Veen S, Verrips EG, Koopman HM, et al. 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.

50 Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M, et al.; KIDSCREEN Group. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health. 2008 Jul-Aug;11(4):645–58.

51 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 Pharmacoecon Outcomes Res. 2005 Jun;5(3):353–64.

52 Achenbach TM, Edelbrock C, Howell CT. Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. J Abnorm Child Psychol. 1987 Dec;15(4):629–50.

53 van der Horst CM, de Borge CA, Knoppe JL, Bossuyt PM. Psychosocial adjustment of children and adults with port wine stains. Br J Plast Surg. 1997 Sep;50(6):463–7.

54 Miller AC, Pit-Ten Cate IM, Watson HS, Geromonus RG. Stress and family satisfaction in parents of children with facial port-wine stains. Pediatr Dermatol. 1999 May-Jun;16(3):190–7.

55 He Z, Lu C, Basra MK, Ou A, Yan Y, Li L. Psychological impact of port-wine stain disfigurement. Plast Reconstr Surg. 1998 Sep;102(4):1210–6.

56 Miller AC, Pit-Ten Cate IM, Watson HS, Geromonus RG. Stress and family satisfaction in parents of children with facial port-wine stains. Pediatr Dermatol. 1999 May-Jun; 16(3):190–7.

57 Nijsten T, Meads DM, McKenna SP. Dimensionality of the Dermatology Life Quality Index (DLQI) in 851 Chinese patients with psoriasis. J Eur Acad Dermatol Venereol. 2013 Jan;27(1):109–15.

58 Mazzotti E, Barbaranelli C, Picardi A, Abeni D, Pasquini P. Psychometric properties of the Dermatology Life Quality Index (DLQI) in 900 Italian patients with psoriasis. Acta Derm Venereol. 2005;85(5):409–13.

59 Rao P, Senthilnathan S, Balakrishnan S, et al. The use of TAPQOL in measuring quality-of-life in adolescents with melasma. Acta Derm Venereol. 2016;96(7):844–8.

60 Kurwa H, Mills C, Lanigan S. Improvement in the psychological impact of a port wine stain after successful pulsed dye laser treatment. J Dermatolog Treat. 1999;10(4):277–82.