

Appendix S1

SUPPLEMENTARY METHODS

Data sources and searches

The main systematic literature searches were unrestricted and carried out using the online bibliographic databases PubMed, Embase, CINAHL, PsycINFO and Scopus for papers published between 1995 (the date of introduction of the first paediatric dermatology-specific HRQOL questionnaire (25)) and January 2016. Additional searches for studies published from January until July 2016 were undertaken to ensure that recently published articles were considered. Keywords were established by means of a pilot search, and included the following combinations of keywords and (when possible) MeSH (Medical Subject Headings) terms: "quality of life" AND "psoriasis" AND "children" OR "adolescents". In addition, alternative conceptualizations of the latter 2 keywords were utilized when relevant (i.e. paediatric, juvenile, childhood). In addition to the main literature searches, www.clinicaltrials.gov was searched to identify unpublished studies. Furthermore, a backward search (snowballing) was conducted using reference lists of identified papers and earlier reviews, together with a forward search (citation tracking), until no additional relevant papers were found.

Study selection

Studies were considered eligible if they: (i) reported data on HRQOL in children or adolescents 4–18 years of age with psoriasis; and (ii) used a quantitative approach to measure HRQOL by either generic, dermatology, or disease-specific self-report questionnaires. Studies restricted to only 1 aspect of quality of life, or related constructs, such as health status, depression, or anxiety, were excluded.

Due to the scarcity of studies examining HRQOL in children and adolescents, we included both controlled and uncontrolled studies, as well as published conference proceedings (i.e. "grey literature"). Papers published in English, German, or Scandinavian languages were included. Case studies describing special cases seen in the clinic or studies with sample sizes <5 were excluded (26–28). In case of missing data, authors were contacted by e-mail and asked to forward relevant data. Authors of papers presenting data on mixed populations of patients with various skin diseases that did not present separate data for their paediatric patients with psoriasis were also contacted. Studies were excluded if authors did not respond or were unable to provide the data. Study selection was conducted by 2 independent authors (HR, TT). Disagreements between reviewers were resolved by discussion. A third author (RZ) was consulted if initial agreement could not be reached.

Quality assessment

To assess the quality of the included studies, we used a revised version of the 5 appraisal-criteria developed by van der Linden and colleagues to assess quality of HRQOL-measurement in studies of HRQOL in patients with cutaneous rosacea (29), together with 9 additional criteria developed specifically for the present study. To adapt van der Linden et al.'s criteria to our population of interest, we replaced the word "rosacea" with "psoriasis", redefined the sample size cut-offs with categories corresponding to the ability to detect small, medium, and large effect sizes (Cohen's *d*) with 80% statistical power (30), and expanded criterion 4 (comparison with persons with other skin diseases or healthy controls) by differentiating between healthy control groups (4a) and comparison groups with other skin or chronic diseases (4b). The 5 revised criteria and their scoring were as follows: 1) sample size ($n < 27$: 0, $n \geq 27$: 1, $n \geq 84$: 2, or $n \geq 795$: 3), 2) grading of clinical disease severity (no: 0, yes: 1), 3) classification into subtypes of psoriasis

(no: 0, yes: 1), 4a) comparison with healthy controls (s) (no: 0, yes: 1), 4b) comparison with patients with other diseases (no: 0, yes: 1), and 5) the specificity of the applied HRQOL instrument (generic: 0, dermatology: 1, or psoriasis-specific: 2).

The following 9 additional criteria were specifically developed for the present study: 6) eligibility criteria well-defined (no: 0, partly: 1, yes: 2), 7) description of recruitment setting and time frame (no: 0, partly: 1, yes: 2), 8) confirmation of psoriasis diagnosis (no: 0, by non-dermatologist physician or not specified: 1, by dermatologist/specialist: 2), 9) reported whether patients were in treatment (no: 0, yes: 1), 10) relevant demographic characteristics of study participants provided (no: 0, partly: 1, yes: 2), 11) relevant clinical characteristics provided (no: 1, partly: 1, yes: 2), 12) use of HRQOL instrument with age-appropriate versions separating between children and adolescents (no: 0, yes: 1), 13) data analysed within clearly defined age-groups (i.e. children vs. adolescents; no: 0, yes: 1), and 14) response rate reported (no: 0, yes: 1).

A total quality score was obtained by summing the scores of items 1 to 14, yielding a total score ranging from 0 to 23. The strength of agreement between reviewers was evaluated with the between-rater total score correlation. Quality ratings were not used as weights in main analyses, as this is generally discouraged (31). Instead, associations between weighted means HRQOL and study quality were explored using meta-regression, and a detailed overview of study quality assessment was provided.

Data extraction

Data extraction was conducted independently by 2 authors (HR, TT) and included: (i) first author and year of publication, (ii) study design, recruitment setting, and country, (iii) comparison groups, (iv) primary objectives, (v) psoriasis patient characteristics (number of patients; clinical subtype; age; boys/girls-ratio; disease duration; age at onset), (vi) clinical disease severity (e.g. Psoriasis Area and Severity Index (PASI); body surface area (BSA)), and (vii) HRQOL-scores. When summary data (means, SDs) were unavailable, other descriptive statistics (e.g. medians and range) were extracted and converted into the desirable format (32). For studies reporting data from independent subgroups of participants (e.g. children and adolescents), the original subgroups were maintained in our analyses to provide the richest datasets possible (33, 34). If data from more than one instrument was available in a single study, e.g. the Pediatric Quality of Life Inventory (PedsQL) and the Children's Dermatology Quality of Life Index (CDLQI) (34,35), we extracted data for both instruments and used these data in separate analyses.

Analytical strategy

To enable between-study comparisons, studies were grouped according to the specific HRQOL instrument used. Separate random-effects meta-analyses were conducted based on these groupings. Independency was ensured by only combining and comparing studies with independent samples within each analysis. The impact of psoriasis on HRQOL was explored by combining the inverse variance weighted mean results from the included studies, thereby taking the precision of each study into consideration. When interpreting the mean score on the CDLQI/DLQI instrument, we relied on stratification bands developed for the CDLQI from a UK sample: no effect (0–1 points), small effect (2–6 points), moderate effect (7–12) very large effect (13–18 points) and extremely large effect (19–30 points) (36). Interpretation of the PedsQL was based on weighted mean total scores only, as norms across other diseases and healthy populations are available online (www.pedsq.org).

When available for ≥ 3 studies for any HRQOL instrument, the following clinical and demographic characteristics of study participants were explored as possible moderators of HRQOL im-

pairment using meta-regression: (i) percentage girls, (ii) age, (iii) clinical disease severity, (iv) age at onset, (v) duration, (vi) region (Western vs. other), and (vii) study quality. Prior to running the meta-regression analyses, we tested whether any of the proposed moderators were correlated. Highly correlated variables ($r < 0.70$) were entered as covariates in the relevant meta-regression models.

Heterogeneity was assessed using Q , I^2 , and T^2 statistics. Heterogeneity tests are aimed at determining whether results are likely to reflect genuine differences (heterogeneity), or whether the variation is more likely to be due to sampling error (homogeneity) (37). In accordance with recommendations, due to the often low statistical power of heterogeneity tests (38), a p -value ≤ 0.10 was used to indicate statistically significant heterogeneity. The I^2 quantity provides a measure of the degree of inconsistency by estimating the amount of total variance that can be accounted for by heterogeneity in the sample of studies (39). I^2 values of 25, 50, and 75% are considered low, moderate, and high, respectively.

T^2 is the variance of the true effects, and was used to calculate prediction intervals. The prediction interval represents the region in which results of comparable, future studies are expected to be found (40).

Publication bias is a widespread problem when conducting meta-analyses (41). Publication bias was evaluated using funnel plots and Egger's method (42, 43). A funnel plot is a graphic illustration of study results in relation to study size or precision. Egger's test provides a statistic for the skewness of results (44). If the results were suggestive of publication bias, an adjusted point estimate was calculated using Duval and Tweedie's trim and fill method (45), which imputes estimates of missing studies and recalculates it accordingly.

Data were pooled with a random-effects model using Comprehensive Meta-Analysis, version 3. Additional calculations were performed using SPSS version 22 or various formulas in Microsoft Excel.