State of the art of patient-reported outcomes in acromegaly or GH deficiency: a systematic review and meta-analysis
Meulen, M. van der; Najafabadi, A.H.Z.; Broersen, L.H.A.; Schoones, J.W.; Pereira, A.M.; Furth, W.R. van; ...; Biermasz, N.R.

Citation
doi:10.1210/clinem/dgab874

Version: Publisher's Version
License: Creative Commons CC BY-NC-ND 4.0 license
Downloaded from: https://hdl.handle.net/1887/3563108

Note: To cite this publication please use the final published version (if applicable).
State of the Art of Patient-reported Outcomes in Acromegaly or GH Deficiency: A Systematic Review and Meta-analysis

Merel van der Meulen,1,* Amir H. Zamanipoor Najafabadi,1,2,†, Leonie H. A. Broersen,1 Jan W. Schoones,3 Alberto M. Pereira,1,† Wouter R. van Furth,2 Kim M. J. A. Claessen,1 and Nienke R. Biermasz1,†

1Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands
2University Neurosurgical Center Holland, Leiden University Medical Center, Haaglanden Medical Centre and Haga Teaching Hospital, Leiden and The Hague, The Netherlands; and
3Directorate of Research Policy, Leiden University Medical Center, Leiden, The Netherlands

*Shared first authors.
†Correspondence: Merel van der Meulen, BSc, Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands. Email: m.vandermeulen@lumc.nl.

Abstract

Context: Insight into the current landscape of patient-reported outcome (PRO) measures (PROM) and differences between PROs and conventional biochemical outcomes is pivotal for future implementation of PROs in research and clinical practice. Therefore, in studies among patients with acromegaly and growth hormone deficiency (GHD), we evaluated (1) used PROMs, (2) their validity, (3) quality of PRO reporting, (4) agreement between PROs and biochemical outcomes, and (5) determinants of discrepancies.

Evidence Acquisition: We searched 8 electronic databases for prospective studies describing both PROs and biochemical outcomes in acromegaly and GHD patients. Quality of PRO reporting was assessed using the International Society for Quality of Life Research (ISOQOL) criteria. Logistic regression analysis was used to evaluate determinants.

Evidence Synthesis: Ninety studies were included (acromegaly: n = 53; GHD: n = 37). Besides nonvalidated symptom lists (used in 37% of studies), 36 formal PROMs were used (predominantly Acromegaly Quality of Life Questionnaire in acromegaly [43%] and Quality of Life-Assessment of Growth Hormone Deficiency in Adults in GHD [45%]). Reporting of PROs was poor, with a median of 37% to 47% of ISOQOL items being reported per study. Eighteen (34%) acromegaly studies and 12 (32%) GHD studies reported discrepancies between PROs and biochemical outcomes, most often improvement in biochemical outcomes without change in PROs.

Conclusions: Prospective studies among patients with acromegaly and GHD use a multitude of PROMs, often poorly reported. Since a substantial proportion of studies report discrepancies between PROs and biochemical outcomes, PROMs are pivotal in the evaluation of disease activity. Therefore, harmonization of PROs in clinical practice and research by development of core outcome sets is an important unmet need.

Key Words: acromegaly, growth hormone deficiency, patient-reported outcomes, quality of life, trials, prospective studies

Patients with acromegaly have pathologically high growth hormone (GH) and insulin-like growth factor 1 (IGF-1) levels (1). As a result, patients often suffer from multisystem morbidity, including facial changes, acral growth, cardiovascular disease, diabetes mellitus, arthropathy, fragility fractures, and neuropsychological complaints (1-3). Consequently, an increased mortality is observed in untreated patients (4). Although treatment often results in normalization of GH and IGF-1 levels with significant improvement of comorbidity and mortality, a substantial proportion of patients still suffer from extensive (irreversible) late effects of disease and impaired health-related quality of life (HRQoL) (1,3).

On the other end of the spectrum exists GH deficiency (GHD), which can be caused by various pathological processes in the pituitary and hypothalamic region, the most common being pituitary tumors and their treatment (5). Most patients with GHD have multiple pituitary hormone deficiencies. Adult GHD is characterized by an adverse body composition, skeletal fragility, impaired cardiac function, muscle weakness, and a decline in HRQoL (6). In addition, life expectancy is reduced in patients with hypopituitarism in adulthood (7). Although GH replacement therapy improves both biochemical parameters and HRQoL in patients with GHD, many comorbid conditions (partially) persist despite hormonal supplementation (8).

In both acromegaly and GHD, biochemical outcomes do not always correlate with HRQoL and reported symptoms (9-11). Therefore, treatment should not only aim to normalize biochemical outcomes but also improve patient-reported outcomes (PROs) such as HRQoL (3). To capture the patients’ perspective by measuring symptoms and HRQoL, the use of PRO measures (PROMs) in clinical trials and practice has been advocated in addition to biochemical evaluation (3,12). PROMs can cover a wide range of unidimensional...
or multidimensional concepts of HRQoL, ranging from specific symptoms and bodily limitations to restrictions in social participation (13). Moreover, incorporation of PROMs in trials and clinical practice safeguards comprehensive patient-centered outcome measurement. However, no clear criteria nor consensus exist for the use of PROs in clinical trials in patients with acromegaly or GHD. As a result, a great variety of validated and unvalidated generic and disease-specific PROs are currently being used in patients with pituitary disease (14), limiting comparability between trials. Proper reporting of the use, analysis, and outcomes of PROMs in publications is needed to facilitate translation of the PRO results into clinical practice (15-17). Moreover, consensus on the assessment of treatment effectiveness in patients with conflicting biochemical outcomes and PROs is warranted.

Therefore, in prospective studies in patients with acromegaly or GHD, we aimed to assess (1) the current landscape of PROM use, (2) the validity of the used PROMs, (3) the quality of PRO reporting, (4) the concordance between PROs and biochemical outcomes, and (5) determinants for discrepancy between PROs and biochemical outcomes.

Methods

This study was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (18).

Search Strategy and Eligibility Criteria

Prospective randomized and nonrandomized studies assessing both PROs and conventional biochemical outcomes in patients with acromegaly or GHD were searched in the following databases: PubMed, MEDLINE, Embase, Web of Science, COCHRANE Library, Emcare, PsycINFO, and Academic Search Premier. This search was performed in April 2021, in collaboration with a trained clinical librarian (J.W.S.). The search strategy is presented in Supplement 1 (19). There were no restrictions regarding treatment for acromegaly or GHD. Studies were excluded if they (primarily) described children, included less than 10 patients, and were written in other languages than English. If multiple studies described (partially) overlapping populations, the largest study was included. Only original articles were included.

Selection of Articles and Data Extraction

Two independent investigators (A.H.Z.N. and L.H.A.B.) first screened references by title and abstract and subsequently reviewed potentially relevant articles by full-text screening. Moreover, the references of the included studies were screened for additional eligible articles. Inconsistencies were resolved in consensus.

The following data were extracted by 2 independent investigators (L.H.A.B. and M.M.): (1) study design, (2) study objectives, (3) study population [number of patients, age, sex, tumor size, previous treatment (including possible washout period)], (4) criteria for diagnosis, (5) study treatment, (6) follow-up duration, (7) PROs (ie, symptom questionnaires or symptom lists and HRQoL questionnaires), (8) biochemical outcomes [ie, GH, IGF-1, and insulin-binding growth factor-binding protein 3 (IGFBP-3) levels], and (9) comparison between PROs and biochemical outcomes. If data were only presented in a figure without numbers, we made an estimate of the outcome according to this figure. If only stratified data were presented, data were combined into 1 outcome score using a fixed effects meta-analysis. For articles presenting a score before study and a difference between before and after study, we calculated the score after treatment, imputing the SD from before treatment as the best estimate of the SD after treatment. If data were reported with the median and range, the mean and SD were estimated according to the Box-Cox method (20). Furthermore, the biochemical outcomes were converted to µg/L for GH and IGF-1 and to mg/L for IGFBP-3, if they were reported in other units.

Risk of Bias Assessment

The risk of bias of the included studies was assessed using a component approach. We included the following components that could potentially bias a reported association between treatment for acromegaly or GHD and symptoms or HRQoL:

1. Loss to follow-up: <5% was considered low risk of bias.
2. Missing outcome data: missing data in <5% of patients was considered low risk of bias.
3. Inclusion of patients: consecutive inclusion of all eligible patients or a random sample was considered low risk of bias.
4. Criteria for diagnosis: for low risk of bias, at least an oral glucose tolerance test had to be performed for the diagnosis of acromegaly, and at least 1 stimulation test (insulin tolerance test, GH-releasing hormone/arginine test, glucagon stimulation test, or clonidine stimulation test) had to be performed for the diagnosis of GHD.
5. Assay for measurement of serum GH and/or IGF-1 levels: for low risk of bias, the type of assay had to be described.

Study Endpoints

1. List of used PROMs used in acromegaly and/or GHD
2. Validity of the used PROMs
3. Quality of PRO reporting
4. Patient-reported and biochemical outcomes
5. Determinants of discrepancies between patient-reported and biochemical outcomes

Assessment of PROM Validity

To assess the validity of the used PROMs, PubMed was searched for validation studies using the search strategy as provided by the COSMIN research group (21). If at least 1 aspect of validity had been confirmed for a PROM, its validation was judged positively. Since most used PROMs had not been validated in patients with acromegaly or GHD, we also evaluated whether the validity had been assessed in other pituitary diseases.

Assessment of the ISOQOL Criteria

The International Society for Quality of Life Research (ISOQOL) criteria (16) consist of 24 items and were developed for randomized controlled trials (RCTs) assessing PROs. We also used these criteria in a modified form for nonrandomized studies (22). We only assessed the ISOQOL criteria for studies that used an existing PROM, not for studies using self-invented symptom lists. The original ISOQOL criteria and details about scoring of the ISOQOL items are described in Supplement 2a (19), and the modified ISOQOL criteria are...
presented in Supplement 2b (19). These criteria were judged by 1 investigator (M.M.), and uncertainties were discussed with 2 other investigators (A.H.Z.N. and K.M.J.A.C.) until consensus was reached.

**Assessment of Discrepancies**

PROs and biochemical outcomes were considered discrepant if both changed in opposite directions or if only 1 changed while the other remained stable. Since inconsistencies between changes in GH and IGF-1 have been reported in up to 40% of acromegaly patients (23), IGF-1 levels were considered leading when determining the direction of change of the biochemical outcomes.

**Statistical Analysis**

The primary study outcome of the meta-analyses was the pooled mean difference (with 95% CI) before vs after study for the total scores of the Acromegaly Quality of Life Questionnaire (AcroQoL) and Patient Assessed Acromegaly Symptom Questionnaire (PASQ) for studies in acromegaly and the Quality of Life Assessment of Growth Hormone Deficiency in Adults (QoL-AGDHA) for studies in GHD. These were chosen because they were the most frequently used disease-specific PROMs in the included studies. Higher scores on the AcroQoL (scale 0-100) represent better HRQoL, while higher scores on the QoL-AGDHA (scale 0-25) represent worse HRQoL, and higher scores on the PASQ (scale 0-40) represent a greater symptom burden. For the PASQ, only the studies reporting the 5 original symptoms (soft tissue swelling, fatigue, headache, arthralgia, excessive perspiration) on a scale of 0 to 40 were included, because this is the original and most frequently used method to calculate the total PASQ score.

The secondary outcome measures were the pooled mean differences (with 95% CI) before vs after study of GH, IGF-1, and IGFBP-3. For the meta-analysis of GH in patients with acromegaly, studies using pegvisomant were excluded, because GH levels are not a reliable measure of disease control under pegvisomant treatment (24).

A random-effects model was used if at least 5 studies could be included in the analysis of a specific outcome, and a fixed-effects model was used if less than 5 studies could be analyzed, since between-study variance cannot be estimated reliably in that case (25). For all meta-analyses, stratified analyses were performed for treatment-naive patients and patients who had been treated before the start of the study. Mixed groups of treatment-naive and treated patients were coded as treated before the study. For studies including a placebo group, results of placebo-treated patients after the study period were excluded from this meta-analysis. All meta-analyses were visualized using forest plots, both for all included studies and separately for intervention studies and cohort studies. For the random-effects models, the Hartung-Knapp adjustment was used to calculate the CI around the pooled mean difference. Heterogeneity was presented with the $\tau^2$ (using the DerSimonian and Laird estimator) and $I^2$.

Finally, logistic regression analyses were used to assess which study characteristics were associated with a discrepancy between biochemical outcomes and PROs.

All analyses and visualizations were performed in R (version 4.0.3) (26) using the packages tidyverse (version 1.3.1) (27) and meta (version 4.18-2) (28).

**Results**

**Identification and Selection of Literature**

A total of 90 studies were included in the final analysis, published between 1990 and 2021 (Fig. 1), including 53 studies among patients with acromegaly (n = 3667 patients, 11-358 per study) and 37 studies among patients with GHD (n = 6795 patients, 10-4110 per study) (Fig. 2). We included 18 cohort studies describing the natural course of several outcome parameters without initiation of new therapies. Furthermore, we included 38 nonrandomized and 34 randomized studies evaluating end points both before and after initiation of a new treatment. A list of references of all included studies is provided in Supplement 3 (19).

**Study Characteristics**

**Acromegaly**

Study characteristics of the 53 studies in acromegaly patients are shown in Supplement 4a and 4b (19). Sixteen of these studies were RCTs, 3 were multi-arm, nonrandomized trials; 24 studies were single-arm trials; and 10 were cohort studies.
that evaluated patients over time without initiation of a new treatment (protocol) per se. In 9 of the cohort studies, treatment was initiated before study inception. Overall, in 13 studies (25%) all patients were treatment-naive at baseline.

**GH deficiency**

Study characteristics of the included 37 studies in patients with GHD are shown in Supplement 4c and 4d (19). In all studies among GH-deficient patients except 1 cohort study, treatment with recombinant human GH (rhGH) replacement therapy was evaluated. Sixteen studies were placebo-controlled randomized trials, 2 studies were randomized controlled studies comparing different replacement strategies, 11 studies were single-arm trials, and 7 studies were cohort studies prospectively evaluating the effects of rhGH therapy. The single cohort study that did not evaluate rhGH treatment, assessed outcomes in untreated patients with GHD. In total, 26 studies (70%) only included treatment-naive patients.

**Risk of Bias Assessment**

Acromegaly

Loss to follow-up (range 0%-53%) was reported by 29 (55%) studies. Only 14 (26%) reported <5% loss to follow-up. Twenty-three (43%) studies reported missing outcome data, which ranged from 0% to 73% and totaled <5% in 5 (9%) studies. Two (4%) studies explicitly stated that consecutive patients had been recruited. To diagnose acromegaly, 30 (57%) studies had used at least an oral glucose tolerance
test. For the measurement of serum GH and IGF-1 levels, 45 (85%) studies reported which assay was used. A detailed risk of bias assessment is presented in Supplement 5a (19).

**GH deficiency**

Loss to follow-up (range 0%-54%) was reported by 13 (35%) studies. Five studies (14%) reported <5% loss to follow-up. Eight (22%) studies reported missing outcome data, which ranged from 0% to 44% and totaled <5% in 2 studies (5%). Three (8%) studies explicitly stated that consecutive patients had been recruited. To diagnose GHD, 31 (84%) studies used at least 1 GH stimulation test. For the measurement of serum GH and IGF-1 levels, 33 (89%) studies reported which assay was used. A detailed risk of bias assessment is presented in Supplement 5b (19).

**PROMs Used**

**Acromegaly**

PROMs for acromegaly were measured with the AcroQoL in 23 (43%) studies and with the PASQ in 8 (15%) studies. Twelve other PROMs were used in 10 (19%) studies (Table 1). In total, 32 (60%) studies used other nonvalidated symptom lists, including a variation of the PASQ with different answer options, and 21 (40%) of these studies did not use a validated PROM besides these nonvalidated symptom lists.

**GH deficiency**

The QoL-AGDHA was the most commonly used PRO for GHD, used in 16 (43%) studies. A total of 27 other PROMs were used less commonly in 28 (76%) studies, and nonvalidated symptom lists were used in 2 (5%) studies (in combination with validated PROMs) (Table 1).

**Validity of the PROMs**

**Acromegaly**

Of the 14 existing PROMs that were used in acromegaly studies, only 1 (AcroQoL) has been validated in patients with acromegaly (Table 1) (29). While the AcroQoL and the PASQ are both disease-specific and frequently used PROMs, the PASQ has not been validated in acromegaly or any other patient population. Moreover, 3 of the other PROMs used in acromegaly have been validated in patients with GHD (ie, Hospital Anxiety and Depression Scale (30), Nottingham Health Profile (31), Short Form-36 (31)).

**GH deficiency**

Of the 28 PROMs that were used in studies among GH-deficient patients, 9 have been validated in this population. Two of those were developed specifically for patients with GHD (ie, QoL-AGDHA (32) and Questions on Life Satisfaction-Hypopituitarism (33)), while for the other 7, validity was only assessed to a (very) limited extent. The other PROMs have not been validated in any other pituitary patient population before (Table 1).

**Quality of PRO Reporting**

**Acromegaly**

Since 23 studies among acromegaly patients did not use an existing PROM, the ISOQOL criteria were assessed for 30 (57%) studies (20 cohort studies and 10 RCTs). The median percentage of ISOQOL items reported in these studies was 47% (range 5%-65%) for cohort studies and 37% (range 26%-54%) for RCTs [Table 2; Supplement 6a and 6b (19)].

**GH deficiency**

All 37 studies among patients with GHD used at least 1 previously published PROM and were therefore assessed using the ISOQOL criteria. The median percentage of ISOQOL items reported in these studies was 44% (range 11%-65%) for cohort studies and 42% (range 16%-62%) for RCTs (Table 2; Supplement 6c and 6d (19)).

**PROs and Biochemical Outcomes**

**Acromegaly**

The median follow-up on study level in studies with acromegaly was 12 months. Over this period, the AcroQoL total scores [quantitative data available for 17 studies; Fig. 3; Supplement 7a (19)] improved significantly in intervention studies (mean difference 4.3, 95% CI 2.5 to 6.0) but not in cohort studies (mean difference 1.7, 95% CI −0.8 to 4.2). The PASQ scores [quantitative data available for 6 studies; Fig. 3; Supplement 7b (19)] showed an overall improvement (mean difference −3.7, 95% CI −6.9 to −0.6; no separate analysis was done since only 1 cohort study used the PASQ).

This was accompanied by a significant decrease in IGF-1 levels [quantitative data available for 34 studies; Supplement 7c (19)], both in intervention studies (mean difference −292 µg/L, 95% CI −372 to −211) and cohort studies (mean difference −326 µg/L, 95% CI −496 to −157). GH levels [quantitative data available for 28 studies; Supplement 7d (19)] only decreased significantly in intervention studies (mean difference −10.7 µg/L, 95% CI −13.2 to −8.3) and not in cohort studies (mean difference −1.6 µg/L, 95% CI −4.7 to 1.5). The improvement in biochemical outcomes was most pronounced in treatment-naïve patients.

The PROs and biochemical outcomes are presented per study in Supplement 7e and 7f, respectively, and per study aim in Supplement 9 (19).

**GH deficiency**

The median follow-up on study level in studies with GH-deficient patients was 12 months. Over this period, the QoL-AGHDA [quantitative data available for 13 studies; Fig. 3; Supplement 8a (19)] improved for both intervention studies (mean difference −3.4, 95% CI −5.4 to −1.3) and cohort studies (mean difference −5.4, 95% CI −5.6 to −5.2). Similarly, IGF-1 [quantitative data available for 23 studies; Supplement 8b (19)] significantly increased in both intervention studies (mean difference 130 µg/L, 95% CI 85 to 174) and cohort studies (mean difference 180 µg/L, 95% CI 174 to 186), a significant increase was also seen in IGFBP-3 [quantitative data available for 3 studies; Supplement 8c (19), a significant increase was also seen (mean difference 0.4 mg/L, 95% CI 0.3 to 0.5). Improvements were again most pronounced in treatment-naïve patients.

The PROs and biochemical outcomes are presented per study in Supplements 8d and 8e, respectively, and per study aim in Supplement 9 (19).

**Discrepancy Between PRO and Biochemical Outcomes**

**Acromegaly**

Eighteen (34%) studies among patients with acromegaly reported discrepant results between PROs and biochemical outcomes [Table 3; Supplement 10 (19)]. The percentage of discrepant results was slightly higher among studies measuring HRQoL (38%) compared to studies measuring...
Table 1. List of patient-reported outcome measures used by the included studies, including validation in acromegaly and growth hormone deficiency

<table>
<thead>
<tr>
<th>Used patient-reported outcome measures</th>
<th>Description</th>
<th>Validated in acromegaly or GHD (yes/no)</th>
<th>Validated in other pituitary conditions?</th>
<th>Confirmed measurement properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acromegaly studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acromegaly Quality of Life Questionnaire (29)</td>
<td>Disease-specific, self-rating questionnaire consisting of 22 items covering a physical scale (8 items) and a psychological scale (14 items, divided equally over the 2 subscales appearance and personal relations), assessing quality of life in acromegaly patients</td>
<td>Yes, acromegaly (29)</td>
<td>No</td>
<td>Internal consistency, content validity</td>
</tr>
<tr>
<td>Appearance Self-Esteem scale (44)</td>
<td>Self-reported questionnaires consisting of 5 items assessing satisfaction with appearance. Part of the State Self-Esteem scale</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Australian/Canadian Hand Osteoarthritis Index (45)</td>
<td>Self-rating questionnaire consisting of 15 questions covering 3 domains (pain, disability, and joint stiffness) used to assess hand osteoarthritis</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>EuroQol 5 Dimensions (46)</td>
<td>Self-rating questionnaire consisting of 6 items (5 multiple choice questions and 1 visual analogue scale) covering 5 domains to assess utility and health-related quality of life</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Marks’ Social Situation Questionnaire (47)</td>
<td>Self-rating questionnaire consisting of 30 items describing situations concerned with social phobia</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Patient Assessed Acromegaly Symptom Questionnaire (48)</td>
<td>Disease-specific, self-rating questionnaire consisting of 5 items assessing symptoms and signs of acromegaly (soft-tissue swelling, arthralgia, headache, excessive perspiration, and fatigue)</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Research And Development-36 (49)</td>
<td>Self-rating questionnaire consisting of 36 items covering 9 domains (general health, vitality, physical functioning, bodily pain, physical role functioning, emotional role functioning, social role functioning, mental health, change in health) which yield a physical component score and mental component score to assess health-related quality of life. This scale is similar to the Short Form-36, but with slightly different scoring algorithm for body pain and general health, and with addition of the change in health domain.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Western Ontario and McMaster Universities Osteoarthritis Index (50)</td>
<td>Self-rating questionnaire consisting of 24 questions covering 3 domains (pain, disability, and joint stiffness) used to assess hip and knee osteoarthritis.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Growth hormone deficiency studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive Psychopathological Rating Scale (51,52)</td>
<td>Questionnaire to be completed by the clinician, consisting of 65 items that assess a wide range of mental symptoms and from which 3 subscales have been developed (depression [Montgomery Asberg Depression Rating Scale anxiety], anxiety [Brief Scale for Anxiety], obsessive-compulsive disorder [CPRS-Obsessive Compulsive Disorder])</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Defense Style Questionnaire (53)</td>
<td>Self-rating questionnaire consisting of 40 items, yielding 20 individual defense scores and 3 higher-order factor scores (mature, neurotic, immature) used to assess defense style</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Depression Scale of the Munich Psychiatric Information System (54)</td>
<td>Self-rating questionnaire consisting of 16 items assessing depression</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Disease Impact Scale (55)</td>
<td>Self-rating questionnaire consisting of 8 items, assessing the impact of disease on different areas of everyday life</td>
<td>Yes, GHD (30)</td>
<td>No</td>
<td>Reliability, internal consistency, construct validity</td>
</tr>
<tr>
<td>General Health Questionnaire (56)</td>
<td>Self-rating questionnaire consisting of 28 items covering 4 subscales (severe depression, anxiety and insomnia, somatic complaints, social dysfunction) to assess psychiatric symptoms</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hamilton Depression Rating Scale (57)</td>
<td>Structured interview conducted by a clinician, consisting of 17 items assessing depression</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Used patient-reported outcome measures</th>
<th>Description</th>
<th>Validated in acromegaly or GHD (yes/ no)</th>
<th>Validated in other pituitary conditions?</th>
<th>Confirmed measurement properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopkins Symptoms Checklist (58)</td>
<td>Self-rating questionnaire consisting of 58 items, covering 5 symptom domains (somatization, obsessive-compulsive, interpersonal sensitivity, anxiety, depression) to assess psychiatric symptoms</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>KIMS Patient Life Situation Form (59)</td>
<td>Self-rating questionnaire collecting patient-reported outcomes, the patient’s personal situation, and use of social care and healthcare resources, for a metabolic database (KIMS).</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Life Fulfillment Scale (60)</td>
<td>Self-rating questionnaire consisting of 12 items, divided over 2 subscales (personal fulfillment and material fulfillment) assessing life fulfillment.</td>
<td>Yes, GHD (30)</td>
<td>No</td>
<td>Reliability, internal consistency, construct validity</td>
</tr>
<tr>
<td>Mental Fatigue Questionnaire (61)</td>
<td>Questionnaire to be completed by the clinician, assessing 5 aspects of mental fatigue.</td>
<td>Yes, GHD (30)</td>
<td>No</td>
<td>Reliability, internal consistency</td>
</tr>
<tr>
<td>Minnesota Multiphasic Personality Inventory-2 (62)</td>
<td>Self-rating questionnaires consisting of 567 items covering clinical 10 scales (hypochondria, depression, hysteria, psychopathologic deviate, masculinity/femininity, paranoia, psychasthenia, schizophrenia, hypomania, social introversion) to assess personality</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Montgomery-Asberg Depression Rating Scale (63)</td>
<td>Structured interview conducted by a clinician, covering 10 items of the CPRS that assess depression.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Pittsburgh Sleep Quality Index (64)</td>
<td>Self-rating questionnaire consisting of 19 items generating 7 component scores (subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, daytime dysfunction) and a global score, to assess sleep quality and disturbances</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Profile of Mood States (65,66)</td>
<td>Self-rating questionnaire consisting of 65 items covering 6 domains (tension-anxiety, depression-dejection, anger-hostility, fatigue-inertia, vigor-inertia, confusion-bewilderment), used to assess transient, distinct mood states. The shortened version consists of 37 items.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Psychological General Well-Being Index (67)</td>
<td>Self-rating questionnaire consisting of 22 items covering 6 domains (anxiety, depressed mood, positive well-being, self-control, general health, and vitality) to assess health-related quality of life</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Quality of Life Assessment of Growth Hormone Deficiency in Adults (32)</td>
<td>Disease-specific, self-rating, unidimensional questionnaire consisting of 25 items assessing quality of life of growth hormone-deficient adults</td>
<td>Yes, GHD (32)</td>
<td>No</td>
<td>Reliability, internal consistency, construct validity</td>
</tr>
<tr>
<td>Questions on Life Satisfaction-Hypopituitarism (33)</td>
<td>Disease-specific, self-rating, unidimensional questionnaire consisting of 9 items assessing quality of life of adults with hypopituitarism</td>
<td>Yes, GHD (33)</td>
<td>Yes, patients with multiple pituitary hormone deficiencies (33)</td>
<td>Reliability, internal consistency, construct validity, responsiveness</td>
</tr>
<tr>
<td>Schedules for Clinical Assessment in Neuropsychiatry (68)</td>
<td>Structured interview conducted by a clinician, to assess psychiatric symptoms. Gathers both patient-reported information and information from clinicians and case records.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Self-Esteem Scale (69)</td>
<td>Self-reported questionnaire consisting of 10 items assessing self-esteem</td>
<td>Yes, GHD (30)</td>
<td>No</td>
<td>Reliability</td>
</tr>
<tr>
<td>Social Adjustment Scale (70,71)</td>
<td>Self-rating questionnaire consisting of 42 items that measure role performance in 6 areas of functioning (work, social and leisure, extended family, marital, parental, family unit), yielding 6 subscores and a total score</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>State-Trait Anxiety Inventory (72)</td>
<td>Self-rating questionnaire consisting of 40 items assessing state and trait anxiety</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Symptom Checklist-90 (73)</td>
<td>Self-rating questionnaire consisting of 90 items, covering 9 symptom dimensions (somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychotism), rated on a 5-point scale, to assess symptomatic behavior of psychiatric outpatients</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
symptoms (32%). Ten (56%) of the discrepant studies reported an improvement in biochemical outcomes, without improvement in PROs. Most of the studies with consistent results reported improvement in both PROs and biochemical outcomes.

GH deficiency

Twelve (32%) studies among patients with GHD reported discrepant results between PROs and biochemical outcomes [Table 3; Supplement 10 (19)]. The percentage of discrepant results was higher among studies measuring

<table>
<thead>
<tr>
<th>Used patient-reported outcome measures</th>
<th>Description</th>
<th>Validated in acromegaly or GHD (yes/no)</th>
<th>Validated in other pituitary conditions?</th>
<th>Confirmed measurement properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both acromegaly and growth hormone deficiency studies</td>
<td>Epworth Sleepiness Scale (74)</td>
<td>Self-rating questionnaire consisting of 8 items assessing daytime sleepiness</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Hospital Anxiety and Depression Scale (75)</td>
<td>Self-rating questionnaire consisting of 14 items covering an anxiety scale (7 items) and a depression scale (7 items) to assess anxiety and depression in general hospital outpatients care</td>
<td>Yes, GHD (30)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Kellner Symptom Questionnaire (76,77)</td>
<td>Self-rating questionnaire consisting of 92 items assessing 4 scales (depression, anxiety, hostility, somatization), divided into 4 psychological scales (depression, anxiety, hostility, somatization) and 4 well-being scales (contentment, relaxation, friendliness, physical well-being).</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Multidimensional Fatigue Inventory (78)</td>
<td>Self-rating questionnaire consisting of 20 items covering 5 domains, to assess fatigue.</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Nottingham Health Profile (79)</td>
<td>Self-rating questionnaire consisting of 45 items in 2 parts: 38 items covering 6 domains (physical mobility, pain, sleep, emotional reactions. social isolation, energy) to assess subjective health status; and 7 items assessing health-related problems in daily life.</td>
<td>Yes, GHD (31)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Short Form-36 (80)</td>
<td>Self-rating questionnaire consisting of 36 items covering 8 domains (general health, vitality), physical functioning, bodily pain, physical role functioning, emotional role functioning, social role functioning, mental health) which yield a physical component score and mental component score to assess health-related quality of life. Sometimes, the role-social component scores are reported as well.</td>
<td>Yes, GHD (31)</td>
<td>No</td>
</tr>
</tbody>
</table>

Abbreviation: GHD, growth hormone deficiency.

Table 2. Quality of reporting of patient-reported outcomes by studies among acromegaly and growth hormone–deficient patients

<table>
<thead>
<tr>
<th></th>
<th>Acromegaly</th>
<th>Growth hormone deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median % of ISOQOL items reported (range)</td>
<td>47 (5-65)</td>
<td>44 (11-65)</td>
</tr>
<tr>
<td>Cohort studies</td>
<td>37 (26-54)</td>
<td>42 (16-62)</td>
</tr>
<tr>
<td>RCTs</td>
<td>Intended PRO collection schedule</td>
<td>Evidence of PRO reliability and validity</td>
</tr>
<tr>
<td>Items reported by most (≥75%) studies</td>
<td>Patient characteristics including baseline PRO scores</td>
<td>Intended PRO collection schedule</td>
</tr>
<tr>
<td></td>
<td>Appropriate statistical analysis for the PROs</td>
<td>Discussion of the PROs in the context of other clinical studies</td>
</tr>
<tr>
<td></td>
<td>Summary of the PROs in the context of other clinical studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clinical significance of the PROs</td>
<td></td>
</tr>
<tr>
<td>Items reported by few (≤25%) studies</td>
<td>PRO hypothesis and the relevant domains</td>
<td>PRO hypothesis and the relevant domains</td>
</tr>
<tr>
<td></td>
<td>Mode of administration of the PRO</td>
<td>Mode of administration of the PRO</td>
</tr>
<tr>
<td></td>
<td>Rationale for the choice of PRO instrument</td>
<td>Rationale for the choice of PRO instrument</td>
</tr>
<tr>
<td></td>
<td>Windows for valid PRO responses</td>
<td>Identification of the PRO in the trial protocol</td>
</tr>
<tr>
<td></td>
<td>Statistical approaches for missing data and the extent of missing data</td>
<td>Status of the PRO as primary or secondary outcome</td>
</tr>
<tr>
<td></td>
<td>Limitations of the PRO</td>
<td>Windows for valid PRO responses</td>
</tr>
<tr>
<td></td>
<td>Generalizability issues related to the PRO results</td>
<td>Power calculation</td>
</tr>
<tr>
<td></td>
<td>Clinical significance of PRO findings</td>
<td>Statistical approaches for missing data and the extent of missing data</td>
</tr>
<tr>
<td></td>
<td>Flow diagram or description of the allocation of participants and those lost to follow-up for PROs specifically</td>
<td>Clinical significance of PRO findings</td>
</tr>
</tbody>
</table>

Abbreviations: ISOQOL, the International Society for Quality of Life Research; PRO, patient-reported outcome; RCT, randomized controlled trial.
symptoms (47%) compared to studies measuring HRQoL (23%). In most cases, discrepant studies reported an improvement in biochemical outcomes, without improvement in PROs. Most of the studies with consistent results reported improvement in both PROs and biochemical outcomes.

Determinants of Discrepancy
Taking all studies together, logistic regression analysis revealed no significant determinants of a discrepancy between PROs and biochemical outcomes. Although not significant, studies that included participants who had been treated previously showed a tendency toward higher odds of discrepancy.
Studies measuring GHD
All studies in acromegaly (n = 53)†
Studies measuring symptoms (n = 44)
Studies measuring HRQoL. (n = 24)

GHD
All studies in GHD (n = 37)§
Studies measuring symptoms (n = 17)
Studies measuring HRQoL. (n = 28)

<table>
<thead>
<tr>
<th></th>
<th>Both improvement</th>
<th>Both no change</th>
<th>Biochemical improvement, no PRO improvement</th>
<th>PRO improvement, no biochemical improvement</th>
<th>Biochemical no change, PRO deterioration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acromegaly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies in acromegaly (n = 53)†</td>
<td>32 (60)</td>
<td>4 (8)</td>
<td>10 (19)</td>
<td>7 (13)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Studies measuring symptoms (n = 44)</td>
<td>27 (61)</td>
<td>3 (7)</td>
<td>6 (14)</td>
<td>7 (16)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Studies measuring HRQoL. (n = 24)</td>
<td>11 (46)</td>
<td>4 (17)</td>
<td>7 (29)</td>
<td>2 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>GHD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies in GHD (n = 37)§</td>
<td>26 (70)</td>
<td>0 (0)</td>
<td>10 (27)</td>
<td>1 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Studies measuring symptoms (n = 17)</td>
<td>9 (53)</td>
<td>0 (0)</td>
<td>7 (41)</td>
<td>0 (0)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Studies measuring HRQoL. (n = 28)</td>
<td>21 (75)</td>
<td>0 (0)</td>
<td>5 (18)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
</tbody>
</table>

Data are given as n (%). Due to rounding, not all percentages add up to 100%.
Abbreviations: GHD, growth hormone deficiency; HRQoL, health-related quality of life; PRO, patient-reported outcome.
†Fifteen studies in acromegaly reported both symptoms and HRQoL. Bronstein 2016 (81) was discrepant for HRQoL, but concordant for symptoms; therefore, the sum of the studies in this row is 54 instead of 53.
§Eight studies in GHD reported both symptoms and HRQoL. Beshyah 1995 (82) was discrepant for HRQoL, but concordant for symptoms; therefore, the sum of the studies in this row is 38 instead of 37.

Discussion

The results of this systematic review indicate that a substantial number of trials among patients with acromegaly and GHD use PROs alongside biochemical outcomes. However, many different, often unvalidated, PROs are used, limiting comparability between trials. Reporting of these PROs is not according to current standards, hampering proper interpretation, comparability, and implementation of results in clinical practice.

Interestingly, we found that in a third of studies, discrepancies exist between PROs and biochemical outcomes. In the studies with discrepant results, biochemical outcomes generally improved with treatment, while patients’ symptoms and HRQoL remained stable across most domains. Therefore, PROs have added value and should be incorporated in the evaluation of treatment efficacy, and new treatment options and better markers of disease activity are still warranted to decrease the symptom burden and optimize patients’ HRQoL.

Use and Validity of PROs

A multitude of PROs were used in the described studies among acromegaly and GH-deficient patients. While some studies used well-defined PROs to measure PROs, a large proportion of the studies in patients with acromegaly assessed a standardized set of symptoms, without clear definition of a PRO. The use of nondisease-specific or unvalidated PROs has also been observed in other areas, such as diabetes mellitus (34), and oncology (35). This is a point of concern, since validation of PROs is important to ensure their relevance, validity, reliability, and sensitivity to change (ie, responsiveness) (36). Moreover, the use of a validated PROM improves interpretability of the results if minimal clinically important differences have been established (36), and comparability of study results increases if multiple studies use the same validated PROM.

While the majority of the used PROMs have not been validated in these populations, some disease-specific HRQoL PROMs have been developed and validated specifically in patients with acromegaly or GHD, such as the AcroQoL (29) and the QoL-AGHDA (32). Moreover, some PROMs have been developed for pituitary diseases in general, including GH-related diseases, such as the Leiden Bother and Needs Questionnaire for pituitary disease (37). Since some of the used PROMs that were originally developed in other patient populations measure specific dimensions that are also relevant for patients with acromegaly or GHD, such as joint complaints (eg, Australian/Canadian Osteoarthritis Hand Index) or sleep problems (eg, Epworth Sleepiness Scale), it would be of great value to validate these PROMs in these patient populations, too.

Quality of Reporting of PROs

The use of PROs in clinical trials in addition to traditional biochemical outcomes has been advocated to assess treatment efficacy in a patient-centered way (15). However, for proper interpretation, adequate and transparent reporting of these PROs is necessary. Therefore, various guidelines have been published for almost all study types and outcome measures to improve the quality of reporting (38). For this study, we used the ISOQOL reporting guideline (16) and found that reporting of PROs in studies among patients with acromegaly or GHD is generally of poor quality. For example, the rationale for the choice of the PROM and the clinical significance of the findings were often poorly reported. This not only limits reliability of these studies but also hampers their utility in and impact on clinical practice (38).

Analysis and Interpretation

A major drawback of the included studies and an unmet need in the field of endocrinology is the lack of interpretation of
the PRO results, separately as well as in the context of the biochemical outcomes. Minimal important clinical differences were reported by only 2 (39,40) of the included studies, limiting assessment of clinical relevance besides assessment of statistical significance. A minimal important clinical difference can not only be used to assess a difference in mean scores between groups but also, more importantly, to determine the percentage of patients who report having experienced a clinically relevant benefit from treatment in terms of HRQoL.

Furthermore, results of our study indicate that in 33% of studies a discrepancy was reported between the biochemical and PROs, which means that improvement in biochemical outcomes is not necessarily accompanied by improvement in PROs. Although we did not identify clear determinants of these discrepancies, we observed that especially some of the smaller studies did show a trend but may have been underpowered to detect statistically significant changes. However, besides methodological causes, discrepancies may also result from the fact that PROMs and biochemical outcome parameters measure different aspects of health and are therefore complementary. It remains difficult to judge treatment efficacy on biochemical outcome parameters alone, partly because every patient has an individual optimal hormonal setpoint, but also because some symptoms and reduced HRQoL may be caused by irreversible damage that is unresponsive to treatment. Therefore, both PROMs and biochemical outcome parameters are needed to obtain a comprehensive view of disease activity. Nevertheless, most studies did not report how they incorporated both outcomes for the reported conclusion on treatment efficacy. To improve interpretability and usability of PRO results, international efforts have been made to evaluate and standardize sound and comprehensive analysis of PROs in cancer clinical trials, which could be adapted and used for patients with pituitary diseases (17).

## Combination of Outcomes

To comprehensively assess the response of treatment in clinical practice, standardized outcome measures for acromegaly have been developed, such as the ACRODAT® (ACROmegaly Disease Activity Tool) and the SAGIT instrument (41-43). The ACRODAT provides clinically relevant information for acromegaly care, focusing on 5 parameters: IGF-1 levels, tumor status, presence of comorbidities, symptoms, and HRQoL as measured with the AcroQoL. The SAGIT instrument is multidimensional, comprising 5 sections that assess key features of acromegaly: signs and symptoms, associated comorbidities, GH levels, IGF-1 levels, and tumor profile. While both tools already combine PROs with clinician-observed outcomes, they still lack comprehensive PRO measurement, especially to evaluate the efficacy of targeted therapy, which is receiving increasing attention in the treatment of acromegaly (3,9).

Importantly, no similar instruments have been developed for GHD. Therefore, we recommend the development of core outcome sets for both patient populations, covering comprehensive PRO measurement, including recommendations for specific PROMs for different outcomes of interest. Based on the results of this systematic review, we recommend the use of a disease-specific HRQoL questionnaire (such as AcroQoL or QoL-AGHDA) and a validated disease-specific symptom questionnaire, in combination with a generic HRQoL questionnaire, and possibly domain-specific questionnaires depending on the specific study aim. For the development and widespread implementation of these core outcome sets, extensive international collaboration between experts is needed, in which the European Reference Network on Rare Endocrine Conditions may play an important role.

### Strengths and Limitations

This is the first systematic review addressing the comparison between changes in PROs and biochemical outcomes in...
acromegaly and GHD. Strengths of this review include the inclusion of studies using a wide variety of PROMs, thereby providing a comprehensive overview of the PROMs used in studies among patients with acromegaly or GHD and of the PRO results in these populations. In addition, the extensive assessment of the quality of PRO reporting using the ISOQOL guideline provides more in-depth insight into the state of the art of PROs in these populations. However, it should be noted that some ISOQOL items can be interpreted in multiple ways. We therefore aimed to judge all items consistently in all studies and thoroughly discussed uncertainties until consensus was reached. Another limitation is that all included studies had a risk of bias, mainly due to (failure to report) loss to follow-up, missing data, and the method of patient recruitment. Since none of the studies scored low risk of bias on all assessed components, no sensitivity analyses using only studies with low risk of bias could be performed. Lastly, some of the quantitative data were estimated from figures, because not all authors provided all values in text. This may have introduced a degree of uncertainty in the data included in the meta-analysis.

Conclusions
Studies among patients with acromegaly or GHD use a large variety of PROMs, which emphasizes the need for consensus on relevant outcome parameters in these populations, not only for use in trials evaluating treatment efficacy but also in clinical practice. To accomplish this, international collaboration is necessary, for example, within the European Reference Network on Rare Endocrine Conditions. In addition, reporting of the PROs is of poor quality, stressing the necessity of more methodological awareness when reporting PROs. Since discrepancies between PROs and biochemical outcomes exist in a substantial proportion of studies, future studies should pay special attention to the interpretation of study results in case of such discrepancies.

Not only do the results of this systematic review emphasize the need to standardize outcomes, they also function as a starting point for improvement of the use and reporting of PROs next to conventional clinical outcomes in clinical trials with patients with acromegaly or GHD. This may improve interpretability of PRO results and consequently facilitate implementation of these outcomes in clinical practice.

Author Contributions
The study protocol was written by A.H.Z.N. and N.R.B. The search strategy was designed by M.v.d.M., L.H.A.B., J.W.S., and A.H.Z.N. The articles were selected by A.H.Z.N. and L.H.A.B. Data extraction and analyses were performed by M.v.d.M., A.H.Z.N., L.H.A.B., K.M.J.A.C. Data interpretation was performed by M.v.d.M., A.H.Z.N., L.H.A.B., J.W.S., A.M.P., W.R.v.F., K.M.J.A.C., and N.R.B. M.v.d.M. and A.H.Z.N. wrote the first version of the manuscript, which was revised for intellectual content by L.H.A.B., J.W.S., A.M.P., W.R.v.F., K.M.J.A.C., and N.R.B. M.v.d.M., A.H.Z.N., L.H.A.B., J.W.S., A.M.P., W.R.v.F., K.M.J.A.C., and N.R.B. approved the final version of the manuscript. All authors had full access to all the data and had final responsibility for the decision to submit for publication. Chiasma, Inc. had no role in the study conduct, data collection, analyses, data interpretation, and the decision to submit the manuscript.

Financial Support
This work was supported by Chiasma, Inc. Chiasma, Inc. had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or decision to submit the manuscript.

Disclosure Summary
The authors have nothing to disclose.

Data Availability
All data generated or analyzed during this study are included in this published article or in the data repositories listed in the references.

References


47. Marks IM. Behavioural Psychotherapy. Wright; 1986.


54. von Zersen D, Koeller DM. Klinische Selbstbeurteilungs-Skalen (KSB-S) aus dem Münchener Psychologisches Informations-System (PSYCHIS München), Manuale. (a) Allgemeiner Teil; (b) Die Beschwerden-Liste; (c) Paranoid-Depressivitäts-Skala, Depressivitäts-Skala; (d) Die Befindlichkeits-Skala. Beltz Test; 1976.


