

JA, ICH
HABE MEINE
VITILIGO
AKZEPTIERT.
SCHLIEßLICH
HABE ICH
KEINE
WAHL.

65 % aller Menschen mit Vitiligo wird gesagt, ihre Erkrankung sei nicht behandelbar.¹ Noch gravierender ist, dass nahezu die Hälfte aller Betroffenen eine Behandlung überhaupt nicht mehr in Betracht zieht.¹ Wie Sie wissen, tritt Vitiligo meist im Teenageralter auf – und ohne zugelassene Therapie fühlen sich viele Betroffene in einem Zustand der Ungewissheit gefangen. Deshalb forschen wir an neuen wissenschaftlichen Ansätzen. Denn wenn wir uns alle mehr mit der Erkrankung Vitiligo befassen, haben Ihre Patientinnen und Patienten eines Tages vielleicht wieder eine Wahl.

entdeckevitiligo.de

Incyte
Dermatology




ENTDECKE VITILIGO →

© 2022, Incyte Biosciences International Sàrl. All rights reserved.
Date of preparation: May 2022 DE/OTHR/M/22/0002

1. Bibeau K, et al. Diagnosis and Management of Vitiligo From the Perspectives of Patients and Healthcare Professionals: Findings From the Global VALIANT Study. Maui Derm for Dermatologists. Maui, HI. January 24th–28th 2022.

Refinement and validation of the ItchyQoL using classical test theory and item response theory resulted in a reduction of the response categories from a 5-point to a 3-point scale

M. Gabes ^{1,2} C. Zeidler ³ S. Ständer,³ S. C. Chen⁴ and C. J. Apfelbacher¹

¹Institute of Social Medicine and Health Systems Research, Otto von Guericke University Magdeburg, Magdeburg, Germany

²Medical Sociology, Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

³Department of Dermatology and Center for Chronic Pruritus, University Hospital of Münster, Münster, Germany

⁴Department of Dermatology, Duke University School of Medicine, Durham, NC 27710, USA

Linked Comment: L. Misery. *Br J Dermatol* 2021; **185**:478–479.

Summary

Correspondence

Michaela Gabes.

Email: michaela.gabes@med.ovgu.de

Accepted for publication

18 March 2021

Funding sources

This article was supported by the European Academy of Dermatology and Venereology (no. 2014-022 to S.S.).

Conflicts of interest

C.J.A. has received institutional funding and consultancy fees from Dr Wolff GmbH and consultancy fees from Sanofi Genzyme. He is a member of the executive committee of the Harmonising Outcome Measures for Eczema initiative. C.Z. has received speaker honoraria/travel fees from Beiersdorf and Dermasence.

DOI 10.1111/bjd.20082

Background The ItchyQoL is an itch-specific patient-reported outcome measure used to assess quality of life in patients with chronic pruritus (CP).

Objectives We aimed to assess and extend the psychometric properties of the ItchyQoL using classical test theory (CTT) and item response theory (IRT).

Methods Item characteristic curves were analysed to investigate whether the response categories were functioning optimally. Confirmatory factor analyses were carried out on the ItchyQoL prior to and after rescoring of the response categories. We conducted a Rasch analysis for the ItchyQoL with revised response options and assessed the mean fit residuals in addition to the assumptions of unidimensionality and local independence.

Results In total, 551 patients with CP from nine European countries completed the 22 items of the ItchyQoL. IRT analysis supported the revision of response options from five points to three. This revision was supported by excellent structural validity using CTT. The overall fit to the Rasch model was adequate. Unidimensionality was supported by the ItchyQoL overall scale and by the single subscales; however, local independence was violated in eight cases.

Conclusions We suggest a revision of the response categories of the ItchyQoL from a 5-point to a 3-point scale. When this revision was applied, the ItchyQoL showed excellent structural validity according to CTT and IRT/Rasch. The calculation of an overall ItchyQoL sum score is allowed.

What is already known about this topic?

- The ItchyQoL is an instrument used to assess quality of life in patients with chronic pruritus.
- The ItchyQoL has already shown excellent reliability, validity and responsiveness.

What does this study add?

- This study provides further information about the structural validity of the ItchyQoL.
- This analysis provides support for a revision of the response categories of the ItchyQoL from a 5-point to a 3-point scale.
- This revision allows for the calculation of an overall ItchyQoL sum score owing to unidimensionality of the ItchyQoL.

What are the clinical implications of this work?

- A revision of the ItchyQoL would be a more reliable instrument for recording the quality of life in patients with chronic pruritus.

Chronic pruritus (CP) lasts for at least 6 weeks and could be classified as one of the most common symptoms in dermatology.¹ A CP point prevalence of 13.5–16.8% was reported in a German general population and a German working population.^{2,3} Most of the patients affected by CP are severely burdened and almost three-quarters of affected patients are impaired in their everyday life.⁴ Therefore, the quality of life (QoL) of patients with CP is negatively affected.⁵ In order to measure QoL, patient-reported outcome measures (PROMs) are used in clinical trials to capture the patient's perspective.⁶ In addition to generic PROMs, such as the 12-item Short-Form Health Survey⁷ and skin-specific PROMs, such as the Dermatology Life Quality Index,⁸ there is an itch-specific PROM, the ItchyQoL, which has been validated and is currently available for use.⁹

The ItchyQoL measures the QoL in patients with CP. It was developed in the USA with patient input and its content validity was ensured with subsequent patient interviews. In its initial validation in the USA and in a cross-European validation in Austria, France, Germany, Italy, Poland, Russia, Spain, Switzerland and Turkey, the ItchyQoL showed excellent reliability, validity and responsiveness.¹⁰ It consists of 22 items with five response options ranging from 1 (never) to 5 (all the time) for each item. The original US version reports the following three domains of the ItchyQoL: symptoms (items 1–6), functioning (items 7–13) and emotions (items 14–22).⁹ During the validation of a German version of the ItchyQoL, the 'GerItchyQoL', a confirmatory factor analysis supported a better model fit using the following four domains: symptoms (items 1–6), functioning (items 7–13), feelings (items 14, 16–18, 20) and self-perception (items 15, 19, 21, 22).¹¹

Structural validity is an important measurement property that describes the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured. Item response theory (IRT), Rasch analyses or factor analyses as part of the classical test theory (CTT) are the preferred statistical methods to assess structural validity.¹² The structural validity of the ItchyQoL using IRT/Rasch has not been assessed so far. IRT is a powerful tool for the construction of a PROM and its refinement as it describes the relationship between a latent trait (e.g. QoL) and the probability of choosing a particular item response that indicates the trait being measured. IRT can enable researchers to create a visual representation of how well items are performing and therefore allows for an in-depth analysis at both the item level and the person level using the same metric.¹³ However, a large sample size is required to perform IRT analyses ($n \geq 200$).¹² The

Rasch model is very similar to the one-parameter logistic IRT model. IRT and Rasch are both latent trait models that are mainly based on probabilistic models and are useful for the construction or refinement of a PROM.¹³ The Rasch model is used to evaluate how well the data fit the model. It is based on assumptions and requirements, such as unidimensionality, local dependence and monotonicity of the latent trait.^{14,15} The factor structure of the ItchyQoL was analysed using the more traditional CTT to evaluate structural validity.¹²

In this study, we aimed to assess the structural validity of the ItchyQoL using CTT and IRT/Rasch and refine the ItchyQoL if necessary.

Patients and methods

Participants and study design

The sample population consisted of patients with CP who had contact dermatitis, atopic dermatitis, chronic nodular prurigo, psoriasis vulgaris, lichen planus or mycosis fungoides/Sézary syndrome who were included in the The European Network on Assessment of Severity and Burden of Pruritus (PruNet) project after written informed consent had been received.¹⁶ Participants were recruited during routine appointments at the departments of dermatology of the Medical University of Graz (Austria), Das Kurhaus Bad Gleichenberg (Austria), University Hospital of Brest (France), Center for Chronic Pruritus at the University Hospital of Münster (Germany), Catholic University of the Sacred Heart (Italy), Wrocław Medical University (Poland), Moscow Scientific and Practical Center (Russia), Hospital Sant Pau (Spain), Kantonsspital Aarau (Switzerland) and Adnan Menderes University (Turkey). Patients completed the ItchyQoL in their national language which was generated from the original ItchyQoL (see <https://emoryott.technology/publisher.com/technology/1853>) by a professional translation office in accordance with standard protocol.¹⁰ A preliminary version was generated by three independent professional translators and this version was reviewed by the investigator of the country-specific site in order to avoid language difficulties and cultural differences between countries. This version was back-translated by another professional translator and compared with the original version to identify any discrepancies, which were discussed with the local investigator, translators and authors of the original text. The ethics committees of all participating clinical sites approved the PruNet project, which is also registered at the German Clinical Trials Register as DRKS00007958.

Analyses

The distribution of the data was assessed to investigate whether there was a pattern of missing data.

Item response theory: analysis of the response categories and rescoring if necessary (refinement)

Item characteristic curves (ICCs) are a useful feature of IRT. They are helpful for determining the appropriate number of response options based on the patients' responses and their latent trait.¹³ ICCs were analysed to investigate whether the response categories were functioning optimally and whether a rescoring of the ItchyQoL was indicated. Response categories function optimally when each category has a distinct peak on the category probability curve and category thresholds monotonically increase with the category.^{14,17}

Classical test theory: confirmatory factor analyses prior to and after rescoring (validation)

We performed two confirmatory factor analyses (CFAs) to confirm either the three-factor structure or the four-factor structure, or both, prior to and after rescoring of the response categories. The goodness-of-fit of the CFA models was evaluated utilizing the following indices and cut-off levels: comparative fit index (CFI)/Tucker–Lewis index (TLI) (≥ 0.95 adequate fit, ≥ 0.90 acceptable fit), root mean square error of approximation (RMSEA) (< 0.05 good fit, < 0.08 acceptable fit), standardized root mean square residual (SRMR) (< 0.05 adequate fit, < 0.8 still acceptable fit) and χ^2 -test/degrees of freedom ratio (χ^2/df) (< 2 adequate fit, < 3 acceptable fit).^{18–20} According to the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative, the structural validity of an instrument is rated as sufficient if either the CFI or TLI is > 0.95 or the RMSEA is < 0.06 or the SRMR is < 0.08 .¹⁵

Performing classical test theory on the ItchyQoL with revised response categories (validation)

In a second step, an exploratory factor analysis (EFA) was performed for the ItchyQoL with transformed response categories. A CFA was used to confirm the factor structure and therefore the structural validity of the ItchyQoL.

Individual items should not have a strong correlation with one another. Polychoric correlations were estimated in order to assess multicollinearity. Multicollinearity is identified when correlation coefficients are ≥ 0.8 .

Performing Rasch analysis on the ItchyQoL with revised response categories

The overall model fit was tested. A mean fit residual of 0 and an SD of 1 indicate overall model fit.^{14,21} Infit and outfit mean squares were also analysed. These values should be ≥ 0.5 to

avoid overfit (redundancy) and ≤ 1.5 to avoid underfit (too much measurement error).¹⁵ The Person Separation Index (PSI) was calculated. This reliability index gives information about how well an instrument can differentiate persons according to disease severity.¹⁴

Unidimensionality was assessed using a principal component analysis (PCA) on the residuals after fitting the Rasch model. The first component should not account for more than 30% of the variance in the data and should have an eigenvalue ≤ 3 to support unidimensionality.¹⁴ Furthermore, unidimensionality refers to a factor analysis per subscale. In addition, a bifactor model using the weighted least square mean and variance was carried out. According to the COSMIN criteria, unidimensionality is not violated if either the CFI or TLI is > 0.95 or the RMSEA is < 0.06 or the SRMR is < 0.08 .¹⁵

Local independence was assessed by examining the correlations of the item residuals. Residual correlations exceeding 0.2–0.3 violate this assumption, residual correlations > 0.4 indicate redundancy.^{14,22}

The assessment of differential item functioning of the ItchyQoL with revised response categories by sex (validation)

We further carried out differential item functioning (DIF) by sex. The following categorization of DIF was used (Table 1).²³ All data analyses were conducted via IBM SPSS Statistics 25, Mplus 8.4 software (Muthen & Muthen, Los Angeles, CA, USA), SAS 9.4 (SAS Institute Inc., Cary, NC, USA) and Winsteps 4.5.5 (<https://www.winsteps.com/>).

Results

The ItchyQoL and sociodemographic items were completed by 551 patients with CP. Overall, 57% of the patients were female and the mean age was 51.6 years (SD 18.1, range 18–98).

The percentage of missing data per item was between 0.5% and 1.8%. There was no pattern to the missing data, thus data were missing at random. Nevertheless, the rates of missing data slightly increased towards the end of the instrument (Table 2).

Refinement of the ItchyQoL using item response theory

All items showed poorly functioning response categories, i.e. no distinct peak on the continuum for each response option. This was addressed by combining adjacent categories. A rescoring of the ItchyQoL response options from 0-1-2-3-4 to

Table 1 Categorization of the DIF size

DIF category	Category explanation	DIF
C	Moderate-to-large	DIF ≥ 0.64 logits
B	Slight-to-moderate	DIF ≥ 0.43 logits
A	Negligible	DIF < 0.43 logits

DIF, differential item functioning.

Table 2 Pattern of missing data

Item	1	2	3	4	5	6	7	8	9	10	11
N	547	548	542	543	548	546	546	546	543	544	543
Missing	4	3	9	8	3	5	5	5	8	7	8
Item	12	13	14	15	16	17	18	19	20	21	22
N	543	543	542	541	543	542	541	543	541	543	541
Missing	8	8	9	10	8	9	10	8	10	8	10

N, sample size.

0-0-1-1-2 resolved the disordering of the categories in 20 of the 22 items (Figure 1).

Performing classical test theory on the ItchyQoL with revised response categories (validation)

Furthermore, the overall model fit improved using either the three-factor structure or the four-factor structure (Table 3). Three factors could be extracted using an EFA as they had eigenvalues > 1. This dataset supports the three-factor structure found in the validation of the original ItchyQoL rather than the four-factor structure found in the validation of the GerItchyQoL. Polychoric correlations among the items ranged from 0.17 to 0.77, thus no multicollinearity issues were found.

Performing Rasch analysis on the ItchyQoL with revised response categories (validation)

A mean fit residual of 0 (SD 0.64) reflected overall model fit. Infit and outfit mean squares are presented in Table 4. We report an underfit of item 13 (body care) as the infit mean square is slightly greater than 1.5. This indicates too much measurement error, which means that this item might not be sensitive enough. However, the fit of this item should be further investigated in future studies.

Table 3 Goodness-of-fit indices obtained by the confirmatory factor analyses before and after the revision of the response scaling

	Prior to rescoring		After rescoring	
	Three-factor structure	Four-factor structure	Three-factor structure	Four-factor structure
CFI	0.409	0.344	0.914	0.924
TLI	0.338	0.253	0.904	0.914
RMSEA	0.180	0.191	0.063	0.060
SRMR	0.182	0.184	0.045	0.043
χ^2 -test/ df	18.756	21.023	3.050	2.846

CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; df, degrees of freedom.

We found a separation of 2.91 at the person, or rather, patient level. This means that the ItchyQoL was able to identify 2.91 statistically significantly different severity groups. It is able to reliably distinguish between patients, differentiating at least two, and almost three, groups of severity. This is reflected by a PSI of 0.89. For the items, a separation of 7.42 with a reliability of 0.98 was found, i.e. we had more than seven levels of item difficulty in our data. This is good news as the ItchyQoL has items matching many severity levels.

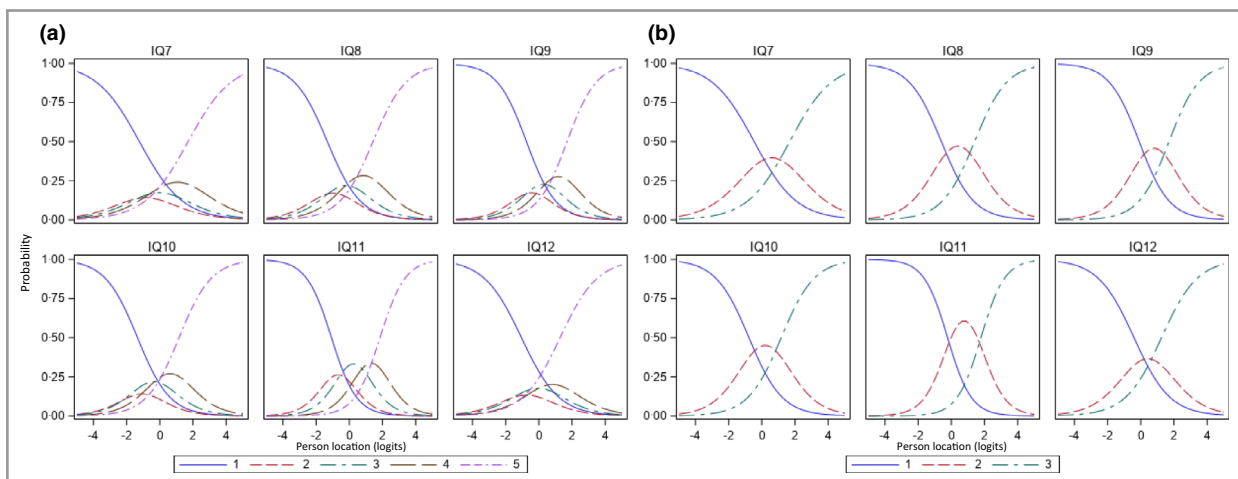


Figure 1 Item characteristic curves for items 7–12 (a) prior to rescoring and (b) after rescoring using three response categories.

Table 4 Infit and outfit mean square for the 22 items of the ItchyQoL

Items	Infit MNSQ	Outfit MNSQ
1. Bleeding	1.07	1.19
2. Pain	0.91	0.91
3. Burning	0.91	0.95
4. Scars	1.40	1.31
5. Scratching	0.85	0.95
6. Season	1.34	1.56
7. Money	1.11	1.12
8. Work	0.92	0.89
9. Interaction	0.95	0.92
10. Sleep	0.96	0.95
11. Concentration	0.66	0.63
12. Clothes	1.17	1.13
13. Body care	1.56	1.64
14. Frustrated	0.88	0.88
15. Self-esteem	0.96	0.90
16. Loss of control	0.87	0.92
17. Angry	0.74	0.75
18. Depressed	0.81	0.76
19. Image	1.09	0.98
20. Never ending	1.09	1.11
21. Embarrassed	0.82	0.75
22. Personality	0.92	0.79

MNSQ, mean squares.

The first component of the PCA had an eigenvalue of 2.997 and accounted for 7.9% of the variance. The fit indices for each of the three ItchyQoL subscales are presented in Table 5. Unidimensionality was supported by the PCA and by the factor analyses per (sub)scale (Table 5). Both the overall ItchyQoL scale and the three subscales fulfilled at least one COSMIN criterion for sufficient structural validity. Using a bifactor model, no COSMIN criterion was fulfilled.

The largest standardized residual correlations for the identification of dependent items are presented in Table 6. Residual correlations exceeded 0.2 eight times, 0.3 twice, and did not exceed 0.4. There were no redundant items; however, the assumption of local independence was violated eight times.

We found negligible DIF ($|DIF| < 0.43$; category A) for 19 of the 22 items. For three items (item 1, item 12, item 13) DIF by sex was slight-to-moderate (DIF 0.45–0.48; category B). This was not surprising as these three items, item 1 (bleeding), item 12 (clothes) and item 13 (body care) are

Table 5 Results of the factor analysis per (sub)scale to support unidimensionality including a bifactor model

Fit indices	Symptoms	Functioning	Emotions	Overall	Bifactor model
CFI	0.952 ^a	0.929	0.954 ^a	0.822	0.874
TLI	0.921	0.894	0.939	0.803	0.848
RMSEA	0.089	0.103	0.086	0.090	0.127
SRMR	0.033 ^a	0.044 ^a	0.034 ^a	0.064 ^a	0.104

CFI, comparative fit index; TLI, Tucker–Lewis Index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual. ^aSufficient according to the COSMIN-based Standards for the selection of health Measurement Instruments (COSMIN) criteria.

Table 6 Largest standardized residual correlations

Residual correlation	Items
0.37	2 (pain) & 3 (burning)
0.31	19 (image) & 21 (embarrassed)
0.29	18 (depressed) & 21 (embarrassed)
0.28	1 (bleeding) & 2 (pain)
0.24	15 (self-esteem) & 19 (image)
0.24	1 (bleeding) & 4 (scars)
0.21	17 (angry) & 18 (depressed)
0.21	8 (work) & 9 (interaction)

The assessment of differential item functioning of the ItchyQoL with revised response categories by sex (validation).

related to outer appearance, which might be more relevant to women.

Discussion

Using IRT, we suggest a revision of the response categories of the ItchyQoL from a 5-point to a 3-point scale. Response options 1 and 2 ('never' and 'rarely'), and 3 and 4 ('sometimes' and 'often') should be combined and response option 5 ('all the time') should be left as a single category to provide visually adequate graphs. We suggest the following wording for the new 3-point scale: 'never', 'often' and 'all the time'. The new wording should be used in future studies to assess whether newly generated ItchyQoL data also fit the Rasch model. When this revision was applied, the ItchyQoL showed excellent structural validity according to CTT and IRT/Rasch. No multicollinearity issues were found. Our findings support the three-factor structure of the ItchyQoL rather than the four-factor structure of the GerItchyQoL; however, both factor structures showed sufficient structural validity results. We prefer the three-factor structure of the ItchyQoL as, from a clinical perspective, it is more feasible to calculate three domain scores rather than four. As the overall ItchyQoL scale fulfilled the assumption of unidimensionality, the calculation of an overall sum score is allowed.

According to the COSMIN group, a PROM is placed in the best category (category A) and is therefore recommended for use if there is evidence for sufficient content validity and at least low-quality evidence for sufficient internal consistency. A

sufficient internal consistency rating is given only if there is at least low-level evidence for sufficient structural validity. A sufficient structural validity rating is assigned if the CFI, the TLI, or a comparable measure is > 0.95 or if the RMSEA is < 0.06 or if the SRMR is < 0.08 .¹⁵ With an SRMR of 0.045, the ItchyQoL fulfils the criterion for sufficient structural validity according to CTT. Using IRT and Rasch analysis, we demonstrated that the ItchyQoL now provides adequate-looking response graphs for most of the items, thus a violation of monotonicity can be denied. Furthermore, we reported an adequate overall model fit. Infit and outfit mean squares were inside the predefined range except for three values that were slightly above 1.5 [items 6 (season) and 13 (body care)]. Unidimensionality was supported by the PCA and factor analyses for the three subscales. At least one of the COSMIN criteria was fulfilled for each subscale.¹⁵ Local independence was violated in eight cases as residual correlations were > 0.2 . However, there was no residual correlation > 0.4 , which indicates that there were no redundant items. Explainable slight-to-moderate DIF by sex was found for three items. As at least low-level evidence for sufficient structural validity was given and Cronbach's α was ≥ 0.70 for each scale,¹⁰ one requirement for a placement in category A was met. As there was also sufficient content validity reported in the original development and validation paper,⁹ the ItchyQoL may fulfil both criteria for category A and could therefore be recommended for future use according to the COSMIN criteria;¹² however, the content validity of the revised response categories should be further assessed.

From a clinical perspective, it is increasingly of interest to measure the impairment of QoL, which is often a secondary endpoint in clinical trials. A revision of the ItchyQoL led to an excellent structural validity with the effect that a revision would provide a more reliable instrument for recording the QoL in patients with CP.

One strength of this study is the large sample size that is required for the application of IRT and Rasch analysis. Furthermore, patients were recruited at the clinic, i.e. CP was physician-confirmed and not self-reported. The percentage of missing data was very low, which may indicate a high level of compliance among the patients and that the items were comprehensible.

There were some limitations to this research. This study included patients with pruritic skin diseases; however, CP can also occur without the presence of visible skin changes. Data were collected in different countries using different translations of the ItchyQoL. It is unclear whether there was DIF by country, different translations or variations in procedures. Another potential limitation is the fact that both CFA and EFA were performed on the same dataset, which might give rise to dependence.

Future studies should focus on the evaluation of the appropriateness of the new response categories in order to have a proper content validity assessment of the refined ItchyQoL. Further validation studies should be carried out to replicate and extend the measurement properties of the refined ItchyQoL.

We suggest a revision of the response categories of the ItchyQoL from a 5-point to a 3-point scale. When this revision was applied, the ItchyQoL showed excellent structural validity according to CTT and IRT/Rasch. The calculation of an overall ItchyQoL sum score is allowed.

Acknowledgments

We thank all members in the PruNet group, especially those who participated as an investigator at the clinical sites: S. Bobko, F.J. Legat, N. Potekaev, A. Lvov, L. Misery, W. Weger, A. Reich, E. Savk, M. Streit, E. Serra-Baldrich and J.C. Szepietowski. Open Access funding enabled and organized by Projekt DEAL.

References

- 1 Ständer S, Weisshaar E, Mettang T *et al.* Clinical classification of itch: a position paper of the International Forum for the Study of Itch. *Acta Derm Venereol* 2007; **87**:291–4.
- 2 Mattered U, Apfelbacher CJ, Loerbroeks A *et al.* Prevalence, correlates and characteristics of chronic pruritus: a population-based cross-sectional study. *Acta Derm Venereol* 2011; **91**:674–9.
- 3 Ständer S, Schäfer I, Phan NQ *et al.* Prevalence of chronic pruritus in Germany: results of a cross-sectional study in a sample working population of 11,730. *Dermatology* 2010; **221**:229–35.
- 4 Kopyciok ME, Ständer HF, Osada N *et al.* Prevalence and characteristics of pruritus: a one-week cross-sectional study in a german dermatology practice. *Acta Derm Venereol* 2016; **96**:50–5.
- 5 Weisshaar E, Apfelbacher C, Jäger G *et al.* Pruritus as a leading symptom: clinical characteristics and quality of life in German and Ugandan patients. *Br J Dermatol* 2006; **155**:957–64.
- 6 Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clinical Pract.* 2006; **12**:559–68.
- 7 Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996; **34**:220–33.
- 8 Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; **19**:210–6.
- 9 Desai NS, Poindexter GB, Monthrope YM *et al.* A pilot quality-of-life instrument for pruritus. *J Am Acad Dermatol* 2008; **59**:234–44.
- 10 Zeidler C, Steinke S, Riepe C *et al.* Cross-European validation of the ItchyQoL in pruritic dermatoses. *J Eur Acad Dermatol Venereol* 2019; **33**:391–7.
- 11 Krause K, Kessler B, Weller K *et al.* German version of ItchyQoL: validation and initial clinical findings. *Acta Derm Venereol* 2013; **93**:562–8.
- 12 Mokkink LB, Prinsen CA, Patrick DL *et al.* COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual. Available at <https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/> (last accessed 22 April 2021).
- 13 Stover AM, McLeod LD, Langer MM *et al.* State of the psychometric methods: patient-reported outcome measure development and refinement using item response theory. *J Patient Rep Outcomes* 2019; **3**:50.
- 14 Kamudoni P, Nutjaree J, Salek S. *Living with Chronic Disease: Measuring Important Patient-Reported Outcomes*. Singapore: Springer, 2018.
- 15 Prinsen CAC, Mokkink LB, Bouter LM *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; **27**:1147–57.

- 16 Ständer S, Zeidler C, Riepe C *et al.* European EADV network on assessment of severity and burden of Pruritus (PruNet): first meeting on outcome tools. *J Eur Acad Dermatol Venerol* 2016; **30**:1144–7.
- 17 Linacre JM. Investigating rating scale category utility. *J Outcome Meas* 1999; **3**:103–22.
- 18 Brown TA. *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press, 2015.
- 19 Moosbrugger H, Schermelleh-Engel K. Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In: *Testtheorie und Fragebogenkonstruktion* (Moosbrugger H, Kelava A, eds). Berlin, Heidelberg: Springer, 2012; 326–43.
- 20 Schmitt TA. Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. *J Psychoeduc Assess* 2011; **29**:304–21.
- 21 Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry* 2009; **9**:21.
- 22 Andrich D, Humphry SM, Marais I. Quantifying local, response dependence between two polytomous items using the Rasch model. *Appl Psych Meas* 2012; **36**:309–24.
- 23 Zwick R, Thayer DT, Lewis C. An empirical Bayes approach to Mantel-Haenszel DIF analysis. *J Educ Meas* 1999; **36**:1–28.