DOI: 10.1111/jdv.16848 JEADV

SHORT REPORT

IDQoL, CDLQI and the 45-item CADIS received a sufficient content validity rating during the HOME VII meeting in Japan: a group discussion study

M. Gabes,[1,2] C. Apfelbacher,[12,*], for the quality of life working group of the Harmonising Outcome Measures for Eczema (HOME) initiative

[1]Institute of Social Medicine and Health Economics, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany
[2]Medical Sociology, Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany
*Correspondence: C. Apfelbacher. E-mail: christian.apfelbacher@med.ovgu.de

Abstract

Background The Harmonising Outcome Measures for Eczema (HOME) initiative has agreed that quality of life should be measured in all atopic eczema clinical trials. Various candidate instruments exist for this domain but their content validity in atopic eczema is largely unclear.

Objective To assess the content validity of quality-of-life candidate instruments for atopic eczema in infants, children and adults in order to aid the decision on what instrument to include in the core outcome set for the quality-of-life domain.

Methods Six group discussions were conducted at the HOME VII Meeting in Tokyo. Each group was composed of 8–12 patients or parents of patients, clinicians, methodologists and pharmaceutical industry delegates and discussed one or two candidate instruments. The COSMIN criteria on relevance, comprehensiveness and comprehensibility were used to determine the overall content validity rating per instrument.

Results Content validity of the Infant's Dermatitis Quality of Life Index, Children's Dermatology Life Quality Index and the Childhood Atopic Dermatitis Impact Scale (CADIS) long-form was rated as sufficient (+). Results for the CADIS short-form, DLQI and Skindex were inconsistent (±). DISABKIDS, Infants and Toddlers Dermatology Quality of Life and ABS-A were classified as having insufficient content validity.

Conclusions The content validity rating allowed for a comparison of all candidate instruments and informed the consensus-seeking process regarding the core instrument for the quality-of-life domain.

Received: 9 April 2020; Accepted: 23 July 2020

Conflict of interest

Christian Apfelbacher is a member of the executive committee of the Harmonising Outcome Measures for Eczema (HOME) initiative. Michaela Gabes and Christian Apfelbacher both attended the HOME VII meeting in Tokyo, however were excluded from the votings within the group discussions as they were leading on the development of one of the instruments (the CADIS short form). Christian Apfelbacher received consultancy fees from Dr. Wolff GmbH and Sanofi Genzyme, and institutional funding from Dr. Wolff GmbH.

Funding source

No financial support received.

Background

Eczema (also called atopic dermatitis or atopic eczema) is a pruritic skin disease that usually starts in early childhood. It affects all age groups, up to 20% of children and up to 3% of adults.[1] With its chronic course, it has a measurable negative impact on the quality of life (QoL) of both affected children and adults.[2,3] The global Harmonising Outcome Measures for Eczema (HOME) initiative (www.homeforeczema.org) aims to develop a consensus-based core outcome set (COS), a minimum of outcomes that should be measured in every clinical eczema trial. Health-related QoL is one of four domains of the COS and measured by self- or proxy-reported instruments.[4] The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative is an international group with expertise in the development and validation of patient-reported outcome measures (PROMs). The HOME initiative works in accordance with the

[Correction added on 5th January, 2021, after first online publication: M. Gabes was designated as corresponding author]

COSMIN guidelines when selecting PROMs for their COS. To reach consensus on a PROM for the QoL core outcome domain, the content validity of several candidate instruments was rated by the HOME meeting attendees. According to the COSMIN initiative, content validity is considered to be the most important measurement property, as it is the degree to which the content of a PROM is an adequate reflection of the construct to be measured.[5] The single items of a PROM should be relevant, comprehensive and comprehensible with respect to the construct of interest and the target population.[5] The aim of this content validity assessment was to facilitate the decision on what instrument to include as a core instrument for the QoL domain.

Methods

We conducted six group discussions at the HOME VII meeting in Tokyo (8–10 April 2019). Groups were composed of patients or parents of patients, clinicians, methodologists and pharmaceutical industry delegates, at least one of these stakeholders was present in each group. All participants of the HOME VII meeting were allowed to participate in the group discussions. Only the preparers of the task, MG and CA, were excluded from the assessment. Each of the group discussions had a time frame of 60 min to discuss either one or two measurement instruments and to rate aspects of content validity. The COSMIN criteria and rating system for evaluating the content validity of PROMs was used and aspects on relevance, comprehensiveness and comprehensibility were rated.[6] The group discussions were facilitated by members of the executive committee of the HOME initiative. All facilitators were briefed before the task started, they were required to refrain from prompting and a sheet with remarks of the COSMIN manual was handed out and explained (Appendix 1). All facilitators gave written informed consent for the publication of their names. One pretest with a group of patients was conducted the day before.

Eight criteria were used to determine the overall content validity rating (Table 1).

Each criterion was either rated as sufficient (+) or insufficient (−) using a general rule: If ≥85% of the items of the PROM fulfilled the criterion, a sufficient rating (+) was given, and if <85% of the items of the PROM fulfilled the criterion, an insufficient rating (−) was assigned. The group decided by a show of hands if a criterion was fulfilled sufficiently or not.

Following the guidance on the single relevance, comprehensiveness and comprehensibility ratings (Table 2), an either sufficient (+), insufficient (−) or inconsistent (±, if at least one of the ratings was '+' and one was '−' (see user manual, p. 59)[6]) overall content validity rating was possible. The analysis of the results was performed by the facilitator and a second reviewer according to the guidance.

The candidate instruments

An updated systematic review revealed that there are currently five instruments for children and four instruments for adults available with published development and/or validation studies (Table 3).[7] Content validity of the Atopic Dermatitis Impact Scale (ADerm-IS) could not be assessed at the meeting since no version was provided by the authors.

Table 1 The COSMIN criteria on relevance, comprehensiveness and comprehensibility for evaluating the content validity of PROMs (see *rating of reviewers* column in the COSMIN user manual, p. 53)[6]

Relevance
1. Are the included items relevant for the construct of interest?
2. Are the included items relevant for the target population of interest?
3. Are the included items relevant for the context of use of interest?
4. Are the response options appropriate?
5. Is the recall period appropriate?
Comprehensiveness
6. Are all key concepts included?
Comprehensibility
7. Are the PROM items appropriately worded?
8. Do the response options match the question?

COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; PROMs, patient-reported outcome measures.

Table 2 Guidance for determining the relevance, comprehensiveness and comprehensibility rating (see user manual, p. 58)[6]

	Relevance	Comprehensiveness	Comprehensibility
+	At least criteria 1 and 2 are rated + AND at least two of the other three criteria on relevance are rated +	Rating of criterion 6	Both criteria 7 and 8 are rated +
−	At least criteria 1 and 2 are rated − AND at least two of the other three criteria on relevance are rated −	Rating of criterion 6	Both criteria 7 and 8 are rated −
±	All other situations	Rating of criterion 6	One criterion is rated + and one is rated −

Results

Each group was composed of 8–12 participants including one facilitator each. All items of each instrument were discussed. The facilitators tried to involve all participants. The ratings of the single groups are presented in Table 4. The Childhood Atopic Dermatitis Impact Scale (CADIS) short-form was presented first to the group to prevent biased results regarding comprehensiveness. All instruments were presented in English except for the Atopic Dermatitis Burden Scale for Adults (ABS-A) which was only available in French. It was translated to the group by a French-speaking participant.

Infant's Dermatitis Quality of Life Index (IDQoL) and Children's Dermatology Life Quality Index (CDLQI) were both very

Table 3 Overview of the candidate instruments for infants, children and adults

PROMs	Infants/children				
	IDQoL	**CDLQI**	**CADIS**	**DISABKIDS**	**InToDermQoL**
Construct of interest	QoL	QoL	QoL	HRQoL	HRQoL
Target population	Infants with AD (<4 years)	Children with skin diseases (4–16 years)	Very young children with AD (<6 years)	Children/ Adolescents with AD (4–16 y)	Children with skin diseases (<1 year, 1–3 years, >3 years)
Recall period	1 week	1 week	4 weeks	4 weeks	1 week
Mode of administration	Proxy-reported	Self-reported	Proxy-reported	Self-/Proxy-reported	Proxy-reported

PROMs	Adults			
	DLQI	**Skindex-16**	**ABS-A**	**ADerm-IS**
Construct of interest	Disability, QoL	HRQoL	Burden	Impact on QoL
Target population	Adults with skin diseases	Adults with skin diseases	Adults with AD	Adults with moderate-to-severe AD
Recall period	1 week	1 week	1 week	24 h and 1 week
Mode of administration	Self-reported	Self-reported	Self-reported	Self-reported

ABS-A, Atopic Dermatitis Burden Scale for Adults; ADerm-IS, Atopic Dermatitis Impact Scale; CADIS, Childhood Atopic Dermatitis Impact Scale; CDLQI, Children's Dermatology Life Quality Index; DLQI, Dermatology Life Quality Index; HRQoL, health-related quality of life; IDQoL, Infant's Dermatitis Quality of Life Index; InToDermQoL, Infants and Toddlers Dermatology Quality of Life; PROMs, patient-reported outcome measures; QoL, quality of life; yr(s), year(s).

much liked by the group. IDQoL is only for children younger than 4 years of age and CDLQI is a self-reported questionnaire (for children >4 years). It was discussed whether children from 4 to 6 years are able to answer the questions themselves. The group felt that aspects concerning teenagers might not have been covered by the CDLQI. Furthermore, it was important to the group that an instrument considers the impact on the family when assessing QoL in children. The heterogenous and more negative response options of the IDQoL and the recall period of the CDLQI due to the fluctuating nature of the disease were considered as the only issues. Most of the participants felt that both instruments were acceptable.

The CADIS long-form was considered as too lengthy by the group participants. According to the group, the questionnaire was geared rather towards the parental perception of disease than the child's QoL. The group criticized that if the parents had more education about the disease, they might respond differently. Furthermore, a recall period of 1 month was considered to be difficult to remember and accuracy would therefore be compromised. The long-form with 45 items seemed comprehensive; however, the 14-item draft short-form lacked an item concerning the child's sleep. There was general consensus that a sleep item should be included to the draft short-form in any way. The group felt a redundancy in items, especially in the parental emotions domain where some questions could have been expressed in only one question.

For the DISABKIDS, the group felt that key concepts, such as aspects on the eczema treatment (effects on the daily life, time spent etc.) and family effects were not included. The participants noted an overlap on appearance questions. The questionnaire was considered to be more about frequency than intensity. Some wording would be difficult for an 8-year-old child and in general the predominantly negative wording was criticized. Especially participants from Japan perceived the questions as too 'direct'. The group voiced that the English word 'bother' is always an issue for translation. Parents of children with eczema noted that it may be hard for parents to rate the impact on the child's sleep since they may not know.

The version of the Infants and Toddlers Dermatology Quality of Life (InToDermQoL) received the worst rating. There were no headings provided which made it difficult to understand the questions. The different number of items for different age groups was considered as an issue for long-term trials. The group noted an overlap in the response options. The 1-week recall period was considered as appropriate.

Both Dermatology Life Quality Index (DLQI) and Skindex received an insufficient comprehensiveness rating. However, regarding comprehensiveness, the group preferred the Skindex over the DLQI since it includes psychological aspect of QoL. The DLQI was more about doing than emotions and relevant aspects, such as anxiety and mood, were missing. Both lacked sleep items which are really important for eczema patients. Further, treatment burden was missing from the Skindex, but this was considered less important by the group. No complaints about relevance and comprehensibility were mentioned by the participants.

The ABS-A was mainly liked by the group. However, psychological aspects, such as mood, anxiety or embarrassment, were missing. The group did not like the response options since every question could be answered as 'not applicable'. This could be difficult to interpret in clinical trials when there are high levels of missing data. If an instrument is designed well, most should not tick 'not applicable' which is therefore not needed. Furthermore, the group suspected that the ABS-A is likely susceptible to floor effects.

Table 4 Results of the single group discussions for all patient-reported outcome measures for infants, children and adults

	Rating of the groups								
	Group 1 *n* = 10 facilitator: Joanne Chalmers		Group 2 *n* = 10 facilitator: Eric Simpson		Group 3 *n* = 11 facilitator: Phyllis Spuls	Group 4 *n* = 11 facilitator: Jochen Schmitt	Group 5 *n* = 8 facilitator: Hywel Williams		Group 6 *n* = 12 facilitator: Kim Thomas
	IDQoL	CDLQI	CADIS (long-form)	CADIS (short-form)	DISABKIDS (proxy- and self-reported)	InToDermQoL	DLQI	Skindex-16	ABS-A
Relevance									
1. Are the included items relevant for the construct of interest?	+	+	–	–	+	–	+	+	+
2. Are the included items relevant for the target population of interest?	+	+	+	+	+	–	+	+	+
3. Are the included items relevant for the context of use of interest?	+	+	+	+	+	–	+	+	–
4. Are the response options appropriate?	–	+	+	+	–	–	+	+	–
5. Is the recall period appropriate?	+	–	–	–	–	+	+	+	+
Relevance rating	+	+	±	±	±	–	+	+	±
Comprehensiveness									
6. Are all key concepts included?	+	+	+	–	–	–	–	–	–
Comprehensiveness rating	+	+	+	–	–	–	–	–	–
Comprehensibility									
7. Are the PROM items appropriately worded?	+	+	+	+	–	–	+	+	–
8. Do the response options match the question?	+	+	+	+	–	–	+	+	–
Comprehensibility rating	+	+	+	+	–	–	+	+	–
Content validity rating	+	+	+	±	–	–	±	±	–

'–', insufficient; '+', sufficient; '±', inconsistent; ABS-A, Atopic Dermatitis Burden Scale for Adults; ADerm-IS, Atopic Dermatitis Impact Scale; CADIS, Childhood Atopic Dermatitis Impact Scale; CDLQI, Children's Dermatology Life Quality Index; DLQI, Dermatology Life Quality Index; IDQoL, Infant's Dermatitis Quality of Life Index; InToDermQoL, Infants and Toddlers Dermatology Quality of Life; *n*, sample size.

Discussion

This content validity study allowed for a comparison of all candidate instruments for atopic eczema in infants, children and adults and informed the consensus-seeking process regarding a core instrument for the quality-of-life domain. During the plenary discussions, it became clear that there may be redundant items across the different domains of the COS, since other core instruments measure symptoms, signs and long-term control. However, this potential item overlap was more tolerable than a lack of comprehensiveness.

A strength of this content validity rating is the heterogeneous composition of the different group discussions with patients or parents of patients, clinicians, methodologists and pharmaceutical industry delegates. Those different perspectives enriched the discussions and gave a comprehensive view on the instruments. There were also some limitations which should be mentioned. Since the meeting took place in Japan, many patients attending the group discussions were Japanese. Group discussions were held in (academic) English and some of the patient participants had problems to perfectly speak and understand English. Further, because patients did not have the methodological background, there might have been a risk for methodologists and clinicians to dominate the discussion which may have led to an underrepresentation of the patients' opinion. Furthermore, group discussion sessions were not recorded and transcribed verbatim. Only notes were made by hand during the sessions. The CADIS short-form was presented first to the group to prevent biased results regarding comprehensiveness; however, this might have also induced a favourable bias towards the short-form. One instrument, the ABS-A, was only available in French which slightly impeded the content validity rating. However, the idea of using the participants of an international meeting to collect data is an effective, innovative and valuable opportunity and should be further discussed in the future. The development and validation of QoL instruments for infants, children and adults with atopic eczema is an ongoing process and new instruments are already on the horizon. This study can serve as an example for future content validity assessments of new emerging instruments.

Conclusion

This content validity rating allowed for a comparison of all candidate instruments and informed the consensus-seeking process regarding the core instrument for the quality-of-life domain.

Acknowledgement

We are grateful to all the participants of the group discussions who discussed the topics and provided valuable and constructive input. We would like to express out special thanks to the group facilitators for their professional guidance of the groups. Furthermore, we would like to thank Joanne Chalmers who provided the meeting minutes for the preparation of this manuscript. Open access funding enabled and organized by Projekt DEAL.

References

1 Nutten S. Atopic dermatitis: global epidemiology and risk factors. *Ann Nutr Metab* 2015; **66**(Suppl. 1): 8–16.
2 Drucker AM, Wang AR, Li WQ *et al*. The burden of atopic dermatitis: summary of a report for the national eczema association. *J Invest Dermatol* 2017; **137**: 26–30.
3 Eichenfield LF, Tom WL, Chamlin SL *et al*. Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. *J Am Acad Dermatol* 2014; **70**: 338–351.
4 Schmitt J, Spuls P, Boers M *et al*. Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting. *Allergy* 2012; **67**: 1111–1117.
5 Mokkink LB, Prinsen CA, Patrick DL*et al*. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual, 2018. URL https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/ (last accessed: 09 April 2020).
6 Terwee CB, Prinsen CA, Chiarotto A*et al*.COSMIN methodology for assessing the content validity of Patient-Reported Outcome Measures (PROMs). User manual, 2017. URL https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/ (last accessed: 09 April 2020).
7 Gabes M, Tischer C, Apfelbacher C. Measurement properties of quality-of-life outcome measures for children and adults with eczema: an updated systematic review. *Pediatr Allergy Immunol* 2020; **31**: 66–77.
8 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guideline to Their Development and Use, 4th edn. Oxford University Press, Oxford, UK, 2008.

Appendix 1

Guidance for giving a sufficient (+) rating for good content validity – Remarks[6]

1	Every PROM item should measure a defined facet of the construct of interest, within the conceptual framework. PROM items should also be specific for the construct of interest, i.e. they should not measure a coexisting, but separate construct. For example, an item in a fatigue questionnaire such as 'my muscles are weak' is relevant to fatigue but not specific for fatigue. Someone with MS may answer this question in the affirmative but not experiencing fatigue. There should also be no unnecessary items (too many items, except for a large scale item bank that will be used for computer adaptive testing). When a total PROM score is evaluated, each subscale (domain) should be relevant for the construct that the total PROM intends to measure. Professionals can best ensure that items are consistent with the theory, conceptual framework or disease model that was used to define the construct of interest.
2	The relevance of the items for the target population can best be judged by patients. Some items may be relevant to only a small number of patients but they are necessary to capture the full range of patient experiences.
3	It should especially be clear whether the PROM is suitable for use in research and/or clinical practice. Professionals are considered to be more knowledgeable about the context of use of the PROM than patients.
4	The response options should be appropriate for the construct, population, and context of use of interest. For example, if the construct is pain intensity, the response options should measure intensity, not frequency. Also, a reasonable range of responses should be provided for measuring the construct of interest.
5	The recall period can be important for measuring the construct, for example, whether there is no recall period (do you feel depressed now?) or whether the recall period is 1 week (did you feel depressed last week?). Different recall periods may be important, depending on the context. However, sometimes is does not matter whether the recall period is e.g. 1 or 2 weeks.
6	The items should cover the full breadth of the construct of interest. However, there are often good reasons for not including all content suggested by patients in a PROM, for example because an item (or domain) is considered to be outside the scope of the PROM. When a total PROM score is evaluated, the subscales (domain) together should cover the full breath of the construct that the total PROM intends to measure.
7	Consider aspects such as reading level (a scale should not require reading skills beyond that of a 12-year old), ambiguous items, double-barrelled questions, jargon, value-laden words, and length or items[8].
8	The response options should be appropriate to the question asked and should be linguistically linked to the item content.