Br J Dermatol 2024; **190**:867–875
https://doi.org/10.1093/bjd/ljae017
Advance access publication date: 23 January 2024

BJD
British Journal of Dermatology
Outcomes and Qualitative Research

The construct validity, responsiveness, reliability and interpretability of the Recap of atopic eczema questionnaire (RECAP) in children

Aviël Ragamin[1,2] Junfen Zhang[3] Suzanne G.M.A. Pasmans[1,2] Renske Schappin[1,2] Geertruida L.E. Romeijn[3] Maroos A. van Reusel[1] Jart A.F. Oosterhaven[3] and Marie L.A. Schuttelaar[3]

[1]Department of Dermatology, Erasmus MC University Medical Centre Rotterdam, Rotterdam, the Netherlands
[2]Department of Dermatology, Centre of Paediatric Dermatology, Sophia Children's Hospital, Erasmus MC University Medical Centre Rotterdam–Sophia Children's Hospital, Rotterdam, the Netherlands
[3]Department of Dermatology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands
Correspondence: Marie L.A. Schuttelaar. Email: m.l.a.schuttelaar@umcg.nl

Plain language summary

Atopic dermatitis (AD) is a skin disease that affects children and adults. People with AD (eczema) and other stakeholders have identified perceived 'eczema control' as an important outcome to investigate in research. For this purpose, the Recap of atopic eczema (RECAP) questionnaire was developed, consisting of seven items to measure eczema control in people with AD. However, when developing questionnaires, they must be examined to ensure they are relevant, reliable and sensitive enough to detect meaningful change before and after any new treatment. Prior studies have demonstrated that the RECAP is suitable for adults with AD, but studies investigating whether the RECAP is suitable for children are lacking.

A study of 231 children (under 12 years old) with AD and their caregivers was conducted in the Netherlands. Caregivers completed the RECAP questionnaire at three time points: at the start of the study, after 1–7 days and after 4–8 weeks. The researchers assessed AD severity and eczema control using other measures for comparison. RECAP scores from children whose caregivers reported unchanged eczema control were used to assess how reproducible this questionnaire was. RECAP scores of caregivers who reported change in eczema control were used to examine sensitivity to change. Statistical tests were used to analyse findings.

The researchers found that RECAP accurately measures changes in eczema control over time and was sensitive enough to detect small changes in eczema control.

Overall, the authors concluded that the RECAP questionnaire is valid, reproducible and responsive. Furthermore, they consider an improvement of at least 6 points to represent a genuine improvement in Dutch children.

Linked Article: von Kobyletzki and Svensson *Br J Dermatol* 2024; **190**:785.

Abstract

Background The Recap of atopic eczema questionnaire (RECAP) was developed to measure eczema control in patients with atopic dermatitis (AD). The measurement properties of RECAP have not yet been validated in caregivers of children with AD.

Objectives To assess the construct validity, responsiveness, reliability and interpretability of the Dutch proxy version of RECAP.

Methods A prospective validation study was conducted in children (aged < 12 years) with AD and their caregivers (in a Dutch tertiary hospital). At three timepoints (T_0 = baseline; T_1 = after 1–7 days; T_2 = after 4–8 weeks) RECAP and multiple reference instruments were completed by caregivers of child patients. Single- and change-score validity (responsiveness) were tested with a priori hypotheses on correlations with reference instruments. Intraclass correlation coefficients ($ICC_{agreement}$) and standard error of agreement ($SEM_{agreement}$) were reported. Bands for perceived eczema control were proposed. The smallest detectable change (SDC) and minimally important change (MIC) were determined. Two anchor-based methods based on receiver operating characteristic curve (ROC) and predictive modelling were used to determine the MIC.

Results A total of 231 children with AD and their caregivers participated. Of our a priori hypotheses for single-score and change-score validity, 77% and 80% were confirmed, respectively. A stronger correlation than hypothesized was found for all rejected hypotheses.

Excellent reliability was found ($ICC_{agreement}$ = 0.94, 95% confidence interval 0.90–0.96). The $SEM_{agreement}$ was 1.9 points. The final banding was 0–1 (completely controlled), 2–7 (mostly controlled), 8–12 (moderately controlled), 13–18 (a little controlled) and 19–28 (not at all controlled). A cutoff point of ≥ 8 was selected to identify children whose AD is not under control. The SDC was 5.3 and the MIC values were 1.5 and 3.6 for the ROC and predictive modelling approaches, respectively. No floor or ceiling effects were observed.

Accepted: 8 January 2024

Conclusions The proxy version of RECAP is a valid, reliable and responsive measurement instrument for measuring eczema control in children with AD. An improvement of ≥6 points can be regarded as a real and important change in children with AD.

What is already known about this topic?

- The Recap of atopic eczema questionnaire (RECAP) was developed to measure 'eczema control' in patients with atopic dermatitis (AD).
- Initial studies have found promising validity and reliability of RECAP; however, the validity of RECAP in children remains uncertain.

What does this study add?

- This study found good construct validity and responsiveness, and excellent reliability of the proxy version of RECAP.
- For single scores, bands for eczema control are proposed: 0–1 (completely controlled); 2–7 (mostly controlled); 8–12 (moderately controlled); 13–18 (a little controlled); 19–28 (not at all controlled).
- A cutoff point of ≥8 was selected to identified children whose AD is not under control; furthermore, an improvement of ≥6 points represents real and important change.

What are the clinical implications of this work?

- The Dutch proxy version of RECAP can now be used to measure 'eczema control' in children with AD.

Atopic dermatitis (AD) is a common chronic inflammatory skin disorder that affects up to 20% of children.[1] AD has a chronic relapsing course characterized by periods of intense inflammation and pruritus (flares), and remission. During meetings of the Harmonising Outcome Measures for Eczema (HOME) initiative, patients, physicians and other stakeholders have agreed to measure AD control as part of the core outcome set for AD.[2,3] For this purpose, together with patients, a new outcome measurement instrument – the Recap of atopic eczema (RECAP) questionnaire – was developed.[4] RECAP aims to capture a patients' perspective of AD control based on seven items. Total RECAP scores range between 0 and 28, with higher scores indicating less 'AD control'.

Previous studies have found promising validity and reliability of RECAP for reporting AD control.[5–7] Recently, a comprehensive study in an academic hospital in the Netherlands of 200 adults with AD found good validity, moderate responsiveness and excellent reliability.[7] This study also assessed the interpretability of RECAP for both single and change scores. In children, RECAP has been less extensively studied. A study in dermatological clinics in the UK assessed the validity of RECAP among 16 caregivers and found strong correlation between patient-reported AD severity and perceived AD control, suggesting good validity.[6] An online survey study in the UK reported good validity, reliability and responsiveness of the RECAP in caregivers.[5] However, this study only contained a few reference instruments and a limited number of caregivers ($n=33$) were included in the reliability analysis. Furthermore, the interpretability of the RECAP in children has not been studied thus far. In order to investigate and interpret the effectiveness of treatment in children in clinical practice and research, validation of outcome measures in designated populations is needed. In this study, we reported the single-score validity, change-score validity (responsiveness), reliability and interpretability of the Dutch proxy version of the RECAP in children.

Materials and methods

Study design

This study was conducted in accordance with the guidelines developed by the COSMIN group.[8]

A prospective study was used to assess measurement properties of the RECAP. The RECAP questionnaire was developed and initially tested by Howells *et al.* and can be found on the website of the Centre of Evidence Based Dermatology, University of Nottingham.[4,9] Caregivers of children with AD were instructed to complete the RECAP questionnaire and reference instruments at three timepoints [T_0 = baseline; T_1 = after 1–7 days; T_2 = after 4–8 weeks (Table 1)]. Caregivers were encouraged to complete the RECAP and reference instruments together with their child where possible, as recommended by Gabes *et al.*[10]

Study population

Children (aged < 12 years) with AD and their caregivers consulting KinderHaven – an outpatient expert clinic for children with atopic diseases of Erasmus MC University Medical Centre–Sophia Children's Hospital in Rotterdam in the Netherlands – were randomly invited to participate in this study between April 2021 and December 2022. All children diagnosed with AD by a dermatologist according to the UK Working Party criteria, regardless of clinical severity and treatment, were eligible to participate except when children and caregivers were unable to understand the Dutch language or when children had a coexisting condition (e.g. urticaria) affecting outcomes of the RECAP or reference instruments.[11]

Reference instruments

An overview of all reference instruments is provided in Appendix S1 (see Supporting Information).

Table 1 Overview of study procedures

	T_0 (on paper)	T_1 (electronic)	T_2 (electronic)
Completed by participants			
Demographics	X	–	–
RECAP	X	X	X
Patient-reported symptoms			
POEM	X	–	X
NRS peak pruritus	X	–	X
NRS sleep disturbance	X	–	X
PtGA of AD severity	X	–	X
Skin-specific HRQoL			
IDQOL (< 4 years)	X	–	X
CDLQI (≥ 4 years)	X	–	X
Generic HRQoL			
EQ-5D-Y	X	–	X
PtGA of AD control	X	X	X
GRC scale	–	X	X
Completed by professional			
EASI	X	–	–
vIGA-AD	X	–	–

AD, atopic dermatitis; CDLQI, Children's Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; EQ-5D-Y, EuroQol Five Dimensions Health Questionnaire Youth; GRC, Global Rating of Change; HRQoL, health-related quality of life; IDQOL, Infants' Dermatitis Quality of Life Index; NRS, numerical rating scale; POEM, Patient Oriented Eczema Measure; PtGA, patient global assessment; RECAP, Recap of atopic eczema questionnaire; T_0, baseline; T_1, after 1–7 days; T_2, after 4–8 weeks; vIGA-AD, Validated Investigator Global Assessment for Atopic Dermatitis.

Table 2 Patient characteristics at baseline (T_0)

Item	Total ($n=231$)	Male ($n=129$)	Female ($n=102$)
Age (years)	3 (0–6)	3 (1–5)	4 (2–7)
Missing (*n*)	0	0	0
Fitzpatrick skin type I–III	136 (59.6)	73 (57.9)	62 (60.8)
Missing (*n*)	3	3	0
Follow-up consultation	146 (63.5)	84 (65.1)	62 (61.4)
Missing (*n*)	1	0	1
EASI	4.25 (1.6–10.5)	4.2 (1.2–10.7)	4.3 (2.1–10.4)
Missing (*n*)	97	56	41
vIGA-AD	2 (1–3)	2 (1–3)	2 (1–3)
Clear	10 (7.0)	5 (6.5)	5 (7.7)
Almost clear	35 (24.6)	22 (28.6)	13 (20.0)
Mild	40 (28.2)	22 (28.6)	18 (27.7)
Moderate	50 (35.2)	26 (33.8)	24 (36.9)
Severe	7 (4.9)	2 (2.6)	5 (7.7)
Missing (*n*)	89	52	37
POEM	11.5 (6.0–18.3)	10 (5–18)	13 (7–19)
Missing (*n*)	41	22	19
NRS peak pruritus	5 (2.0–7.3)	5 (2–8)	6 (2–7)
Missing (*n*)	45	23	22
NRS sleep disturbance	2 (0–7)	2 (0–6)	3 (0–7)
Missing (*n*)	49	26	23
PGA disease severity	2 (1–3)	2 (1–3)	2 (1–3)
Missing (*n*)	41	20	21
PGA disease control	2 (1–3)	2 (1–3)	2 (1–3)
Missing (*n*)	41	20	21
CDLQI	6 (3.0–11.3)	6 (3–10)	7 (4–14.5)
Missing (*n*)	24	9	15
IDQOL	6 (2.0–10.0)	6 (2–10)	7.5 (2–10)
Missing (*n*)	30	19	11
EQ-5D-3Y value	0.84 (0.72–1.0)	0.85 (0.70–1.0)	0.84 (0.72–1.0)
Missing (*n*)	24	10	14
EQ-5D-3Y VAS	80 (70–90)	80 (67.5–92.5)	80 (70–82.5)
Missing (*n*)	28	13	15
RECAP	11 (6–19)	10 (5–19)	12 (6–18.5)
Missing (*n*)	35	18	17

Data are presented as *n* (%) or median (interquartile range) unless otherwise stated. No significant differences between sexes were found according to Mann–Whitney *U* tests and χ^2 tests. CDLQI, Children's Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; EQ-5D-Y, EuroQol Five Dimensions Health Questionnaire Youth; IDQOL, Infants' Dermatitis Quality of Life Index; NRS, numerical rating scale; PGA, patient global assessment; POEM, Patient Oriented Eczema Measure; RECAP, Recap of atopic eczema questionnaire; VAS, visual analogue scale; vIGA-AD, Validated Investigator Global Assessment for Atopic Dermatitis.

Anchors

Patient Global Assessment of atopic dermatitis control

As an anchor, patients were asked the following question (in Dutch): 'To what extent was your child's eczema under control, in the past 7 days?', with five options for answering: 'not at all controlled'; 'a little controlled'; 'moderately controlled'; 'mostly controlled'; and 'completely controlled'.

Global Rating of Change scale

The Global Rating of Change (GRC) scale was used to assess which patients were unchanged at T_1 and which patients had changed (worsened or improved) at T_2. Participants first responded to the question: 'Overall, has there been any change in degree of control in your child's AD since the last time you completed the RECAP?', with the answer options 'yes' or 'no'. If a caregiver answered 'yes', two additional questions were asked. The first question was used to determine the direction and degree of change: 'To what degree has the disease control of your child's AD changed?' The response options to this question were 'much improvement', 'moderate improvement', 'minor improvement', 'minor deterioration', 'moderate deterioration' and 'much deterioration'. Finally, caregivers reported the importance of this change: 'Was this change (improvement/deterioration) important?'; the answer options were 'yes' and 'no'. Based on these questions, participants were classified into seven groups: no important change; important improvement (minor, moderate, much); and important deterioration (minor, moderate, much).

Statistical analysis

All analyses were conducted using SPSS version 28.0 (IBM, Armonk, NY, USA). For all analyses, missing values were not imputed.

Construct validity

Correlation between the RECAP and reference instruments was assessed on single scores (T_0) and change scores (T_2) using Spearman's rank correlation coefficient (ρ). Strong correlation (+++) was defined as a positive or negative $\rho \geq 0.7$; moderate (++) as $\rho \geq 0.4$ to < 0.7; and weak (+) as $\rho < 0.4$. For change scores, we tested whether correlations of changes in the RECAP and changes in instruments measuring similar constructs were ≥ 0.5, and whether correlations of changes in the RECAP with changes in instruments measuring related but dissimilar constructs were lower (0.3–0.5). Finally, we tested whether correlations between changes in RECAP and unrelated constructs were low (< 0.3). Validity was considered high if $< 25\%$ of hypotheses were rejected, moderate if 25–50% were rejected and poor if $> 50\%$ were rejected. Our a priori hypotheses are formulated in Tables 3 and 4. A sample size of at least 70 (item/participant ratio of 1 : 10) was deemed necessary to assess validity.[12]

Reliability

Reliability (test–retest) was assessed in unchanged participants at T1, using intraclass correlation (ICC) metrics for a two-way mixed-effects model for absolute agreement. An $ICC_{agreement}$ value of > 0.70 was considered acceptable. Measurement error was reported with the standard error of measurement ($SEM_{agreement}$) among

Table 3 Single-score validity (at baseline, i.e. T_0) correlations between Recap of atopic eczema and reference measures

Item	Hypothesized correlation	Correlation found	R^2	Hypothesis confirmed
Patient-reported AD symptoms				
POEM ($n=186$)	+++	0.84	0.71	Yes
NRS peak pruritus ($n=183$)	+++	0.83	0.69	Yes
NRS sleep disturbance ($n=179$)	+++	0.72	0.52	Yes
PtGA disease severity ($n=186$)	+++	0.84	0.71	Yes
PtGA disease control ($n=186$)	+++	–0.80[a]	0.64	Yes
Skin-specific QoL				
IDQOL ($n=100$)	+++	0.87	0.76	Yes
CDLQI ($n=79$)	+++	0.76	0.58	Yes
Generic QoL				
EQ-5D-3Y, value ($n=79$)	+	–0.65[a]	0.42	No
EQ-5D-3Y, VAS ($n=76$)	+	–0.40[a]	0.16	Yes
Clinician-reported AD severity				
EASI ($n=115$)	++	0.70	0.49	No
vIGA-AD ($n=123$)	++	0.77	0.59	No
Discriminant measures				
Age (years) ($n=196$)	±	–0.03	< 0.01	Yes
Participant ID number ($n=196$)	±	0.07	< 0.01	Yes
Total no. hypotheses rejected				3/13 (23%)

AD, atopic dermatitis; CDLQI, Children's Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; EQ-5D-Y, EuroQol Five Dimensions Health Questionnaire Youth; IDQOL, Infants' Dermatitis Quality of Life Index; NRS, Numerical Rating; POEM, Patient Oriented Eczema Measure; PtGA, patient global assessment; QoL, quality of life; VAS, Visual analogue scale; vIGA-AD, Validated Investigator Global Assessment for Atopic Dermatitis. [a]Negative value, as both the PGA disease control and EQ-5D-Y questionnaire are inversely scored to the RECAP.

Table 4 Change-score validity at T_2 (i.e. after 4–8 weeks)

Item	Correlation found	Hypotheses confirmed?
Instruments measuring similar constructs (≥ 0.5)		
Change RECAP – change PtGA disease control ($n=100$)	0.83	Yes
Change RECAP – GRC ($n=103$)	0.63	Yes
Change RECAP – change POEM ($n=105$)	0.80	Yes
Change RECAP – change NRS peak pruritus ($n=100$)	0.77	Yes
Change RECAP – change NRS sleep disturbance ($n=99$)	0.70	Yes
Change RECAP – change PtGA disease severity ($n=100$)	0.82	Yes
Instruments measuring related but dissimilar constructs (0.3–0.5)		
Change RECAP – change CDLQI ($n=41$)	0.66	No
Change RECAP – change IDQOL ($n=60$)	0.90	No
Change RECAP – change EQ-5D-3Y value ($n=40$)	–0.41	Yes
Change RECAP – EQ5D-3Y, VAS ($n=39$)	–0.36	Yes
Total no. of hypotheses rejected		2/10 (20%)

CDLQI, Children's Dermatology Life Quality Index; EQ-5D-Y, EuroQol Five Dimensions Health Questionnaire Youth GRC, global rating scale of change; IDQOL, Infants' Dermatitis Quality of Life Index; NRS, numerical rating scale; POEM, Patient Oriented Eczema Measure; PtGA, patient global assessment; RECAP, Recap of atopic eczema questionnaire; VAS, visual analogue scale.

the same participants. Based on COSMIN guidelines, a sample size of at least 50 unchanged participants was deemed adequate.[13]

Interpretability

For single scores, cutoff values for bands indicating perceived AD-related control were evaluated based on the agreement between RECAP scores and patient global assessment (PtGA) for AD control (linear weighted kappa). For change scores, we first calculated the smallest detectable change (SDC) using the following formula.[14] Next, we evaluated the minimal important change (MIC) for improvement based on the predictive modelling ($MIC_{predict}$) and the receiver operating characteristic (MIC_{ROC}) method.[8,15] The GRC at T_2 was used to determine importantly changed patients. For participants who completed T_1, RECAP change scores between T_1 and T_2 were used, while for participants who did not participate at T_1, change scores between T_0 and T_2 were used. A sample size of 100 patients, ideally 50 of whom should be in the improved group, was deemed sufficient.[16] For the MIC based on predictive modelling, an adjusted MIC was calculated, as the proportion of improved patients was 41%.[17]

Floor and ceiling effects

Floor and ceiling effects were considered present if the percentage of participants who achieved the lowest or highest RECAP scores was > 15%.[18]

Results

Patient characteristics

A total of 235 children were included in the study at baseline (T_0), of whom 4 were later excluded based on exclusion criteria (Figure 1). Of the remaining 231 children, 44.2% were female and the median age was 3 years old (interquartile range 0–6). Participant characteristics are presented in Table 2. Based the Eczema Area and Severity Index and Validated Investigator Global Assessment for Atopic Dermatitis distribution, most participants had mild-to-moderate AD. No differences in AD severity or patient-reported outcome measures were found between the sexes.

Single-score and change-score validity

Of our a priori hypotheses for the single-score validity, 77% were confirmed, which indicated high validity of the Dutch RECAP (Table 3). Up to 108 participants were included for analyses of change-score validity. Of the a priori formulated hypotheses for this analysis, 20% were rejected, indicating high change-score validity. For both single- and change-score validity, a stronger correlation than hypothesized was found for all rejected hypotheses.

Reliability

Fifty-seven unchanged participants between T_0 and T_1 were included in the reliability analysis. The $SEM_{agreement}$ was 1.9 points and the $ICC_{agreement}$ 0.94 (95% confidence interval 0.90–0.96), indicating excellent reliability.

Interpretability

Single scores

An overview of the distribution between RECAP scores and PtGA of AD control is presented in Figure S1 (see Supporting Information). Strong correlation ($\rho = -0.80$) between PtGA of AD control and RECAP scores was found. For single RECAP scores, 26 bands for severity of AD control were tested (Table S1; see Supporting Information). The band with the highest kappa value ($\kappa = 0.65$) was selected as the final banding (completely controlled 0–1; mostly controlled 2–7; moderately controlled 8–12; a little controlled 13–18; not at all controlled 19–28) (Table S2; see Supporting Information). After categorizing patients who reported moderately, a little and not at all controlled AD, a single cutoff point of ≥ 8 was determined to identify patients whose AD is not under control. In our study population, three patients (1.6%) had a PtGA of AD control score > 2 points beyond the proposed banding. Three patients (1.6%) had a PtGA of AD control score 2 points higher than the proposed banding and four (2.2%) lower.

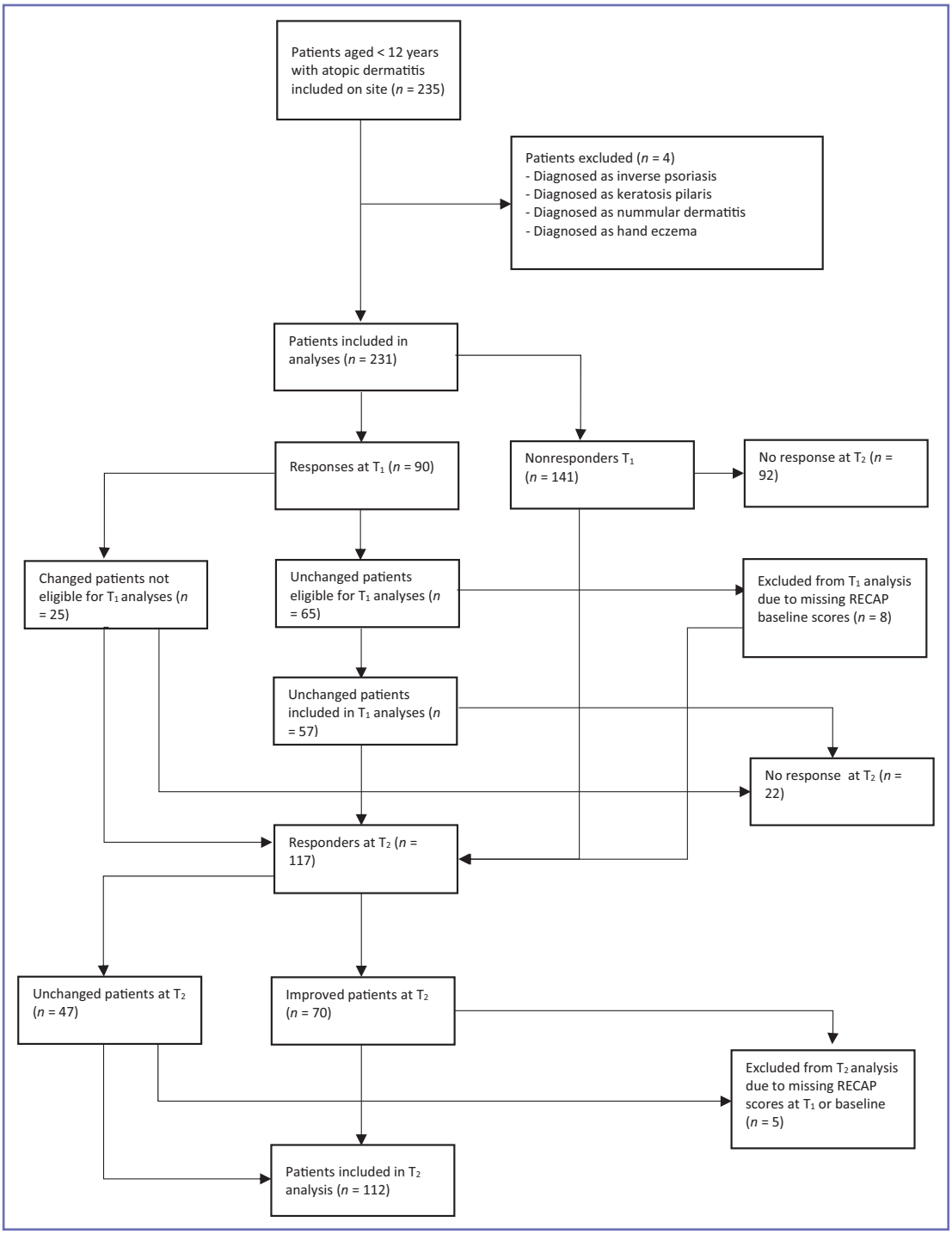

Figure 1 Study flow diagram. RECAP, Recap of atopic eczema; T_1, after 1–7 days; T_2, after 4–8 weeks.

Change scores

The SDC in unchanged participants between T_0 and T_1 was 5.3. The MIC using the ROC method was 1.5 and the MIC using the adjusted predictive modelling method was 3.6 (unadjusted MIC = 3.2).

Floor and ceiling effects

No floor (up to 10%) or ceiling effects (up to 1.0%) were found for all timepoints.

Sensitivity analysis

A post hoc analysis was conducted to compare the baseline characteristics of patients included in the T_2 analysis with patients that could not be included in the interpretability analysis (T_2 analysis) owing to loss of follow-up (Table S3; see Supporting Information). Small but significantly higher patient-reported AD severity, less perceived eczema control and older age at baseline was found in patients lost to follow-up.

Discussion

In this study, we investigated the validity, reliability and interpretability of the proxy version of RECAP in Dutch children. We found high single- and change-score validity, and excellent reliability of the RECAP questionnaire. For the interpretability of single scores, bands for the RECAP scores were determined (completely controlled: 0–1; mostly controlled: 2–7; moderately controlled: 8–12; a little controlled: 13–18; not at all controlled: 19–28). A single cutoff point of ≥ 8 was proposed to identify children whose AD is not under control. Furthermore, for the interpretability of change scores, an improvement of ≥ 6 points should be considered as a real and important improvement in children with AD.

An improvement of ≥ 6 points should be regarded as a real and important change for the proxy version of the RECAP. In our study, we used two anchor-based methods endorsed by COSMIN to assess the MIC, the MIC_{ROC} and $MIC_{predict}$. We found a notable discrepancy between these methods, with a smaller MIC derived from the ROC method (MIC_{ROC} 1.5 vs. unadjusted $MIC_{predict}$ 3.2). However, it should be noted that the sum of the percentage of misclassifications, used to retrieve the optimal cutoff point for the MIC_{ROC}, is close to the sum of the percentage of misclassifications for a cutoff point of 2.5 (difference 0.004). The discrepancy between the MIC_{ROC} and $MIC_{predict}$ was likely caused by the distribution (both the variance and skewness) of change scores.[15] In most studies, including ours, the degree of improvement reported by patients is not normally distributed, with patients more often reporting having experienced 'much' improvement (35/47) than 'minor' or 'moderate' improvement.[19] These patients are more likely to have greater change scores, which can cause more variance. Compared with the predictive modelling approach, the MIC_{ROC} is more sensitive to issues with data distribution, causing a difference between these methods.[15] Moreover, in our study the portion of improved patients was 41%, causing a bias toward a lower MIC than the 'genuine' MIC.[17] To correct this bias, a formula has been published to adjust the $MIC_{predict}$. As yet, no similar correction for the MIC_{ROC} has been proposed. For these and other reasons, the predictive modelling method is endorsed by the COSMIN group as the preferred method to find the most accurate MIC.[20] Therefore, we assumed the adjusted $MIC_{predict}$ to be the best representation of the 'genuine' MIC.

In addition to important improvement, a change in RECAP should be detectable beyond the measurement error. Therefore, the SDC needs to be considered when interpreting change in eczema control on the RECAP questionnaire. In our study, we found an SDC (5.3) larger than the adjusted $MIC_{predict}$ (3.6). Although, the MIC in other patient-reported outcome measures is generally larger than the SDC, an SDC larger than a MIC is not uncommon.[21] While the difference between MIC and SDC in our study was relatively small, ideally the SDC should be smaller than the MIC to reduce the risk of measurement error. In any case, a relatively large SDC indicates that there may be factors influencing the reliability. In addition to the ability to recall and comprehend, judging the health condition of someone else (i.e. children) may influence reliability and thereby the SDC of a questionnaire. Unfortunately, limited research has investigated how the reliability of proxy questionnaires can be improved. Regardless, an improvement of ≥ 6 points can be interpreted as real and important change.

The validation of proxy questionnaires is essential to understanding the effectiveness of interventions and improving the care of young patients and their families. Children are a vulnerable population and, unfortunately, dedicated validation studies following the COSMIN approach for investigating measurement properties of outcome measures in children are scarce.[3,22] As a result, triallists are compelled to use less optimal data for sample size calculation and the interpretation of results, hindering the quality of research. With >70 compounds currently in development for AD, of which many will eventually be investigated in children, validated outcome measures in children are needed.[23] Compared with previous studies that investigated RECAP in adults, similar correlations between RECAP and other measurement instruments were found, indicating that RECAP captures the same construct in children as adults.[5–7] It is worth noting that although we found excellent reliability in children, the reliability was slightly lower than in adults. This is probably due to the difficulty in reporting symptoms by proxy. Consequently, improvements in RECAP scores need to be larger in order to be certain to detect a change, as is reflected in the difference in SDC (3.2 in adults vs. 5.3 in our study).[7] The interpretability of single scores in children was similar to that of adults with only minor differences in banding of the RECAP, most notably between the 'mostly' and 'moderately controlled' groups. As a result, the binary cutoff point for uncontrolled AD differed (≥ 6 points in adults vs. ≥ 8 in children). The MIC scores in adults and children were comparable when using the predictive modelling, with a MIC of 3.9 in adults and 3.6 in children. However, as previously described, the larger SDC in children means that greater improvement in the RECAP ($\geq$ 4 points in adults vs. ≥ 6 in children) is needed to identify real improvement in individual children. Overall, differences in measurement properties of the proxy and adult versions of RECAP were small, indicating the adequacy of RECAP for children. The results of this study can be integrated within HOME. For now, the proposed banding can be used in all children with AD and our data on interpretability in comparable populations. Further validation in other populations (e.g. adolescents and other languages) should still be encouraged.

To our knowledge, our study is one of few to have investigated measurement properties in children with AD following COSMIN guidelines. Further strengths of this study are that all analyses were adequately powered and that the study was directly conducted in an outpatient clinic, which is the intended population of use for many studies. The main limitation of this study was the low follow-up rate (39% at T_1 and 51% at $T_{2)}$, which could have led to selection bias affecting the reliability and interpretability. To gain a better understanding of how this may have affected our results, we conducted a sensitivity analysis to compare the characteristics of responders at T_2 with patients lost to follow-up and found small but significant differences in age and patient-reported AD severity. In general, the MIC tended to be greater in the 'more affected' population; however, we assumed that this would not affect our recommendation of real and important improvement because the SDC in our study was larger than the MIC. Additionally, we assumed that the older age of children in the nonresponder group would likely increase the reliability as these children would be more capable of reflecting on their experienced 'eczema control' with their caregivers. Next, in comparison to a study

investigating the reliability of RECAP in adults, the longer time interval (+1–3 days vs. + 1–7 days) for the reliability analysis could have increased the likelihood of memory and recall bias, potentially leading to lower reliability. Finally, the use of both paper questionnaires at T_0 and electronic questionnaires at T_1 and T_2 could influence measurement properties. However, research shows conflicting results and no consensus has been reached on this subject.[24,25]

The proxy version of RECAP shows high construct validity and responsiveness, and excellent reliability for measuring 'eczema control' in children with AD. Our results suggest a threshold of ≥ 8 to identify patients whose AD is not under control. Furthermore, an improvement of ≥ 6 indicates a clinically important and real change. Further validation of RECAP in other populations such as adolescents and discussion on the role of RECAP in AD are necessary.

Acknowledgements

We gratefully thank all the children and their parents and families for their participation. Furthermore, we would like to thank Tan Nguyen and Anouk Nouwen, in particular, for their support and contributions to this study. Finally, we would like to thank Janneke Verhoeven and Coline Meyer for the logistical support during this study.

Funding sources

J.Z. is supported by the China Scholarship Council (CSC; grant #201806200089).

Conflicts of interest

S.G.M.A.P. is an advisor, consultant, speaker and/or investigator for LEO Pharma, Regeneron Pharmaceuticals, Sanofi Genzyme, Novartis and Pierre Fabre; and has received grants from Novartis and Pierre Fabre, BAP Medical, D&M and DeclaCare (part of BENU Netherlands). M.L.A.S. is an advisor, consultant, speaker and/or investigator for AbbVie, Pfizer, LEO Pharma, Regeneron, Sanofi Genzyme, Eli Lilly and Galderma. She has received grants from Regeneron, Sanofi Genzyme, Novartis and Pfizer. The other authors declare no conflicts of interest.

Data availability

The data that support the findings of this study are available on request from S.G.M.A.P. The data are not publicly available due to privacy restrictions.

Ethics statement

This study was exempt from the Dutch Medical Research Involving Human Subjects Act according to the Institutional Review Board of Erasmus MC (MEC-2020-0417).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website.

References

1 Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *Lancet* 2020; **396**:345–60.
2 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**:1111–17.
3 Williams HC, Schmitt J, Thomas KS *et al.* The HOME Core outcome set for clinical trials of atopic dermatitis. *J Allergy Clin Immunol* 2022; **149**:1899–911.
4 Howells LM, Chalmers JR, Gran S *et al.* Development and initial testing of a new instrument to measure the experience of eczema control in adults and children: Recap of atopic eczema (RECAP). *Br J Dermatol* 2020; **183**:524–36.
5 Bhanot A, Peters TJ, Ridd MJ. Assessing the validity, responsiveness and reliability of the Recap measure of eczema control. *Br J Dermatol* 2021; **184**:955–7.
6 Bhanot A, Vincent R, Peters TJ *et al.* Validation of the RECap of AtoPic eczema measure of eczema control for use in dermatology clinics. *Clin Exp Dermatol* 2022; **47**:440–2.
7 Zhang J, Ragamin A, Romeijn GLE *et al.* Validity, reliability, responsiveness, and interpretability of the Recap of atopic eczema (RECAP) questionnaire. *Br J Dermatol* 2023; **189**:578–87.
8 de Vet HCW, Terwee CB, Mokkink LB *et al. Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press, 2011.
9 Centre of Evidence Based Dermatology. Recap of atopic eczema (RECAP): an outcome measurement instrument to capture 'eczema control'. Available at: https://www.nottingham.ac.uk/research/groups/cebd/resources/recap.aspx (last accessed 30 January 2024).
10 Gabes M, Ragamin A, Baker A *et al.* Content validity of the Recap of atopic eczema (RECAP) instrument in Dutch, English and German to measure eczema control in young people with atopic eczema: a cognitive interview study. *Br J Dermatol* 2022; **187**:919–26.
11 Williams HC, Burney PG, Pembroke AC *et al.* The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. *III. Independent hospital validation. Br J Dermatol* 1994; **131**:406–16.
12 Anthoine E, Moret L, Regnault A *et al.* Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; **12**:176.
13 Mokkink LB, Prinsen CA, Patrick DL *et al.* COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf (last accessed 30 January 2024).
14 de Vet HC, Terwee CB, Knol DL *et al.* When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; **59**:1033–9.
15 Terluin B, Eekhout I, Terwee CB *et al.* Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J Clin Epidemiol* 2015; **68**:1388–96.
16 Terwee CB, Peipert JD, Chapman R *et al.* Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Quality of Life Research* 2021; **30**:2729–54.
17 Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol* 2017; **83**:90–100.
18 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; **4**:293–307.
19 Oosterhaven JAF, Ofenloch RF, Schuttelaar MLA. Validation of the Dutch Quality of Life in Hand Eczema Questionnaire (QOLHEQ). *Br J Dermatol* 2020; **183**:86–95.

20 Terwee CB, Peipert JD, Chapman R *et al.* Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Qual Life Res* 2021; **30**:2729–54.
21 van Kampen DA, Willems WJ, van Beers LW *et al.* Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res* 2013; **8**:40.
22 Gabes M, Donhauser T, Piontek K *et al.* Measurement properties of quality-of-life outcome measures for children and adults with eczema: a systematic review update 2.0. *Pediatr Allergy Immunol* 2023; **34**:e13934.
23 Bieber T. Atopic dermatitis: an expanding therapeutic pipeline for a complex disease. *Nat Rev Drug Discov* 2022; **21**:21–40.
24 Bjorner JB, Rose M, Gandek B *et al.* Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol* 2014; **67**:108–13.
25 Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. *Respir Med* 2009; **103**:932–4.