Validity, reliability, responsiveness, and interpretability of the Recap of atopic eczema (RECAP) questionnaire

Zhang, Junfen; Ragamin, Aviël; Romeijn, Geertruida L E; Loman, Laura; Oosterhaven, Jart A F; Schuttelaar, Marie-Louise A

Published in:
The British journal of dermatology

DOI:
10.1093/bjd/ljad247

Publication date:
2023

Validity, reliability, responsiveness, and interpretability of the Recap of atopic eczema (RECAP) questionnaire

Junfen Zhang,[1] Aviël Ragamin,[2] Geertruida L.E. Romeijn,[1] Laura Loman,[1] Jart A.F. Oosterhaven[1] and Marie-Louise A. Schuttelaar[1]

[1]Department of Dermatology, University of Groningen, University Medical Center Groningen, the Netherlands.

[2]Department of Dermatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands.

Corresponding author: Marie-Louise Schuttelaar

Email: m.l.a.schuttelaar@umcg.nl

Funding sources: J.Z. is supported by the China Scholarship Council (CSC) Grant #201806200089.

Conflicts of interest: Dr. M.L.A. Schuttelaar 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.

Data availability: The data underlying this article will be shared on reasonable request to the corresponding author.

Ethics statement: Reviewed and approved by the Medical Ethical Review Board of the University Medical Center Groningen (reference: METc 202000915)

What is already known about this topic?

- The RECAP questionnaire has been recommended by the Harmonising Outcome Measures for Eczema (HOME) initiative as a core outcome instrument for measuring eczema control.
- Despite its potential utility, the validity and reliability of the RECAP has been investigated to some extent, but there is a paucity of evidence pertaining to its interpretability.

What does this study add?

- The RECAP has good single-score validity and known-group validity, moderate responsiveness and excellent reliability.

- The RECAP scores were categorized into: 0-1 (completely controlled), 2-5 (mostly controlled), 6-11 (moderately controlled), 12-19 (a little controlled), 20-28 (not at all controlled). For the sake of simplicity, a threshold of ≥6 points was determined to identify patients whose AD is considered 'not under control'. Moreover, an improvement of ≥4-points on the RECAP represents a clinically important change.

What are the clinical implications of this work?

- Outcome data from this study can facilitate the practical usage of RECAP in both clinical practice and research settings. The proposed RECAP banding could help to monitor to what extent patients perceive their AD control status, whilst minimally important change scores could help monitor eczema control over time, and evaluate the treatment effectiveness. These findings, in turn, can support shared decision-making among healthcare providers and patients.

Abstract

Background: Limited research has been conducted on the measurement properties of the Recap of atopic eczema (RECAP) questionnaire, particularly its interpretability.

Objectives: To investigate the validity, reliability, responsiveness, and interpretability of the Dutch RECAP in adults with atopic dermatitis (AD).

Methods: We conducted a prospective study in a Dutch tertiary hospital, recruiting adults with AD between June 2021 and December 2022. Patients completed the RECAP questionnaire, reference instruments, and anchor questions at three time points: baseline, after 1-3 days, and after 4-12 weeks. ***Validity:*** Hypotheses-testing was used to investigate single-score validity and change-score validity (responsiveness). ***Reliability***: Both standard error of measurement ($SEM_{agreement}$) and intraclass correlation coefficient ($ICC_{agreement}$) were reported. ***Interpretability of single score***: Bands for eczema control were proposed. ***Interpretability of change score:*** Both smallest detectable change (SDC) and minimally important change (MIC) scores were determined. To estimate the MIC scores, four different anchor-based methods were employed: the mean change method, 95% limit cut-off point, receiver operating characteristic curve, and predictive modelling.

Results: In total, 200 participants were included (57.5% male, mean age 38.5 years). Of the a priori hypotheses, 82% (single-score validity) and 59% (responsiveness) were confirmed. Known-group analyses showed differences in the RECAP scores between patient groups based on disease severity and impairment of the quality of life. The $SEM_{agreement}$ was 1.17 points, and the $ICC_{agreement}$ was 0.988. The final banding was: 0-1 (completely controlled); 2-5 (mostly controlled); 6-11 (moderately controlled); 12-19 (a little controlled); 20-28 (not at all controlled). Moreover, a single cut-off point of ≥6 was determined to identify patients whose AD is not under control. The SDC was 3.2 points, and the MIC value from the predictive modeling was 3.9 points. Neither floor nor ceiling effects were seen.

Conclusion: The RECAP has good single-score validity, moderate responsiveness and excellent reliability. This study fills a gap in the interpretability of the RECAP. Our results indicate a threshold of ≥6 points to identify patients whose AD is 'not under control', while an improvement of ≥4 points represents a clinically important change. Given its endorsement by the Harmonising Outcome Measures for Eczema (HOME) initiatives, the results of this study support the integration of RECAP into both routine clinical practice and research settings.

Introduction

The Recap of atopic eczema (RECAP), a 7-item patient-reported measurement instrument[1], has been recommended by the Harmonising Outcome Measures for Eczema (HOME) initiative as a core outcome instrument for measuring long-term control of atopic dermatitis (AD) in both clinical trials[2] and clinical practice.[3] RECAP was initially developed in the UK, and has since been translated into multiple languages, including Dutch, Chinese, German, French, and Spanish.[4] It includes both self-reported and proxy versions, with the self-completion version being deemed suitable for patients aged 12 years or above.[5] However, despite its potential utility, limited research has been conducted on the measurement properties of the RECAP. While validity and reliability has been investigated to some extent,[1,6] there is a paucity of evidence pertaining to the interpretability of the RECAP scores or the extent to which changes in scores can be considered as clinically relevant. Its validity has been

demonstrated in the initial validation work[1] and in a clinical population with a small sample size of 43 adults.[6] An online survey study has examined its reliability and responsiveness with a self-report AD diagnosis and a low follow-up rate.[7] These validation studies have been conducted in the UK. In addition, the German and Spanish versions of RECAP have demonstrated content validity and have been deemed linguistically equivalent to the original version.[8,9] However, the RECAP has yet to be validated in the Dutch population.

In the present study, we assessed the validity, reliability, responsiveness and interpretability of the Dutch RECAP in adult patients with AD.

Materials and Methods

Study population and design

This prospective study was conducted at the Department of Dermatology in the University Medical Center Groningen (UMCG), a tertiary referral center for atopic dermatitis in the Netherlands. The study design adhered to the guidelines by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group.[10,11] Briefly, adults (≥18 years) with AD, regardless of disease severity and treatment, as diagnosed by a dermatologist according to the U.K. Working Party Criteria,[12] were recruited between 10 June 2021 and 30 December 2022 from the outpatient clinic. Data was collected via RoQua,[13] a tool integrated in the electronic patient record. Patients completed the RECAP, reference instruments, and anchor questions at three time points: at baseline (T_0), after 1-3 days (T_1), and after 4-12 weeks (T_2). Clinical severity was assessed by dermatologists based on the Eczema Area and Severity Index (EASI)[14,15] and the validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD).[16] See Table 1 for an overview of the longitudinal study design, and Supplementary files (Methods: Studied instrument and reference instruments) for descriptions of the above-mentioned instruments. This study was exempt from the Dutch Medical Research Involving Human Subjects Act according to the institutional review board of UMCG (reference: METc 202000915), and all patients provided written informed consent.

Anchors

Patient Global Assessment (PtGA) of AD control

The PtGA of AD control was used to assess patient's overall perception of their disease control at three time points, by asking "What is your overall impression of your atopic dermatitis control over the last week?", with 5 response options: not at all, a little, moderately, mostly, and completely controlled.[17]

Global Rating of Change (GRC) scale

The GRC scale was implemented at T_1 and T_2 to measure the degree of changes in patients' perception of their disease control. First, patients were asked "Overall, has there been any change in the level of disease control of your atopic dermatitis since the last time you completed the RECAP?" with answer categories: no/yes. If a patient answered 'yes', two followed-up questions were asked. One was asked to determine the direction and extent of a change, "To what extent has the disease control of your atopic dermatitis changed?", with six answer categories: much improvement, moderate improvement, minor improvement, minor deterioration, moderate deterioration, much deterioration. The last one indicated the importance of a change "Was this change (improvement/deterioration) important to you?" with response options: no/yes. Consequently, patients were ultimately classified into seven groups: no important change, important improvement (much/moderate/minor improvement), and important deterioration (minor/moderate/much deterioration).

Single-score validity and responsiveness (change-score validity)

Hypotheses testing was used to investigate the validity of the RECAP, with a-priori hypotheses formulated in the study protocol (2021-01-12) before data collection. For the single scores, tests on correlations between the RECAP and reference instruments were performed at T_0 using Spearman's rho (r). For the change scores, a correlation difference of ≥0.1 was deemed relevant.[18] Additionally, as recommended by COSMIN, we tested whether correlations of changes in the RECAP with changes in reference instruments measuring similar constructs were ≥ 0.5, and whether correlations with changes

in reference instruments measuring related but dissimilar constructs were between 0.3 and 0.5.[18] Validity was appraised as high, moderate and poor, if < 25%, 25-50% and >50% of hypotheses were rejected, respectively.[18]

Known-groups validity

Box plots of the RECAP scores showing differences between patient groups were presented. It facilitates the interpretation of the discriminating potential of the RECAP better than mean and standard deviation (SD).[11]

Reliability

Test-retest reliability was assessed among unchanged patients between T_0 and T_1 according to the GRC scale, by calculating the intra-class correlation coefficient ($ICC_{agreement}$) using a two-way mixed effects model for absolute agreement.[19] An $ICC_{agreement}$ value of > 0.70 was considered to be acceptable.[20] Measure error was reported in the same group with standard error of measurement ($SEM_{agreemnt}$), using the square root of the within-subject total variance of an analysis of variance.[19] Moreover, Bland-Altman plot was drawn to illustrate the agreement between repeated measures (T_0 and T_1) and identify possible outliers.[21]

Interpretability

Single scores

The PtGA of AD control was used as an anchor at T_0 to determine possible cut-off points of the RECAP scores, and a linear weighted kappa (κ) coefficient of agreement was calculated to determine the highest level of agreement. Sensitivity analyses were performed to test if patients falling within versus outside the proposed banding had a similar distribution of sex and age. Moreover, a single cut-off point was estimated to simplify its use, where patients who reported their AD as 'not at all controlled' or 'a little controlled' or 'moderately controlled' were classified as 'not under control'.

Change scores

The smallest detectable change (SDC) was determined in unchanged patients at T_1 according to the GRC scale, using the formula: $SDC = 1.96 \times \sqrt{2} \times SEM_{agreement}$.

The minimally important change (MIC) for improvement was determined in importantly changed patients at T_2 based on the GRC scale. The anchor questions were considered as an appropriate anchor to determine the MIC if their correlation with changes in the RECAP scores was >0.30, but preferably >0.50.[22] Change scores for the RECAP and reference instruments were calculated by subtracting the score at T_2 from that at T_1. Positive scores indicated an improvement in disease control while negative scores indicated a deterioration in disease control. Patients were stratified based on their degree of change, considering the indication of their change as important/not important. Four MIC values were determined:

(1) The mean change method: based on the mean change in the RECAP scores of the group with an important minor improvement on the GRC scale

(2) The 95% upper limit cut-off point: based on the 95% upper limit cut-off point of the not importantly changed patients, which corresponds to $Mean_{change}+1.645\times SD_{change}$ of this group

(3) The receiver operating characteristic (ROC) cut-off point: indicating the point closest to the upper left corner, where the sum of percentage of correctly classified patients was highest

(4) The predictive modelling: using logistic regression to predict if a patient belonged to the importantly improved or not importantly improved group according to the GRC scale, with changes in the RECAP as the predictor.[23] The MIC was calculated based on the equation $(\ln(odds_{pre}) - C) / B_x$, with C representing the intercept and B_x representing the regression coefficient of the changes in the RECAP. The $odds_{pre}$ was calculated using the prevalence of important improvement divided by 1 minus the prevalence based on the GRC scale. Furthermore, an adjusted MIC was reported due to the prevalence of being importantly improved not being equal to 0.5 (0.372) in this study.[24]

Floor and ceiling effects

If the percentage of patients who achieved the highest or lowest RECAP scores was > 15%, floor and ceiling effects were considered to be present.[25]

Statistical analysis

This study meets the following recommendations with regards to the sample size for different analyses: an item/subject ratio of 1:10 (n≥70) for construct validity;[26] ≥50 unchanged patients seen as adequate for reliability;[10] a sample size of ≥100 patients with ≥50 reporting important improvement for interpretability.[11,27] Variables were analysed using descriptive statistics, including mean (SD), median (interquartile range (IQR)) and proportions. To compare the differences between groups, categorical variables were compared using the Chi-Square test, while continuous variables were analysed using either the Mann Whitney U test or the median test. For all analyses, cases with missing values were excluded. IBM SPSS STATISTICS for Windows, Version 28.0 (SPSS Inc. Chicago, IL, U.S.A.) was used for all analyses.

Results

In total 204 patients were recruited at baseline (T_0). Of those, 200 patients were included in the T_0 analyses, after excluding 4 patients due to language barrier or the diagnosis of other types of eczema. A study flow diagram is provided in Figure 1. Of the study population, 57.5% were males and the mean age was 38.5 years. Female patients generally reported greater disease severity, more impairment in health-related quality of life (HRQoL), and worse symptoms related to their AD, than male patients (Table 2).

Single-score validity and responsiveness (change-score validity)

Of the a priori hypotheses for single-score validity, 82% were confirmed, indicating a high single-score validity of the Dutch RECAP (Table 3). In the analyses of responsiveness, 188 patients were included

who completed questionnaires at both T_0 and T_2; 59% of the a priori hypotheses for change scores were confirmed, indicating a moderate responsiveness of the Dutch RECAP (Table 4).

Known-groups validity

According to the known-group analyses, patients with greater disease severity based on all relevant outcome measures had higher RECAP scores (indicating poor AD control). Likewise, subgroups of Dermatology Life Quality Index (DLQI) that reported a greater impact on QoL were associated with higher RECAP scores (Figure S1).

Reliability

There were 112 patients included for the reliability analyses, who filled out the T_1 questionnaires within 1-3 days and indicated no change on the GRC scale at T_1. The $SEM_{agreement}$ was 1.17 points. The $ICC_{agreement}$ was 0.988 [95% confidence interval (CI) 0.983-0.992], indicating an excellent reliability. Furthermore, the Bland-Altman plot revealed that the repeatability for most of the test-retest measures was within the limits of agreement (-3.4 to 3.1), with 5 outliers observed (Figure S2).

Interpretability

Single scores

The distribution of the RECAP scores by the PtGA of AD control is shown in Figure S3. There was a significant, strong correlation between the PtGA of AD control and the RECAP (Spearman's rho = -0.82, $p<0.001$), which was not significantly affected by age nor sex. A total of 24 banding options were tested with details presented in Table S1 and S2. The banding with the highest κ-coefficient of agreement (κ = 0.671) was chosen as the final banding: 0-1 (completely controlled); 2-5 (mostly controlled); 6-11 (moderately controlled); 12-19 (a little controlled); 20-28 (not at all controlled). Moreover, a single cut-off point of ≥6 was determined to identify patients whose AD is not under control.

Overview of RECAP scores falling outside the proposed banding

Of the study population, 1 patient (0.5%) had a PtGA of AD control score > 2 points outside of that predicted by the proposed banding. There were 5 patients (2.5%) whose actual PtGA of AD control score was 2 points lower than the proposed banding would have predicted from their RECAP scores, while in 2 patients (1.0%) it was 2 points higher than the proposed banding would have predicted. The patients falling outside versus within the proposed banding, exhibited a similar distribution of age and sex.

Change scores

Smallest detectable change. The SDC of the RECAP was based on the same unchanged group for the reliability analyses, and it was 3.2 points.

Minimally important change. The correlation between the change in the RECAP scores and the GRC scale (rho =0.66) was higher than the minimally recommended correlation of 0.3-0.5 for estimating MIC values.[22] The GRC scale was thus considered to be a useful anchor. The distribution of raw RECAP change scores was visualized as the anchor-based distribution for patients indicating that they had important improvement or no important change, along with the four MIC values based on different methods (Figure 2). The MIC values derived from different methods were as follows: 4.1 for the mean change method, 7.7 for the 95% upper limit cut-off point, 3.5 for the ROC cut-off point, and 3.9 for predictive modelling after adjustment.

Floor and ceiling effects

Neither floor nor ceiling effects were observed, because <5% of patients achieved the highest or the lowest score at all three time points.

Discussion

In the present study, we demonstrated that RECAP had good single-score validity, excellent test-retest reliability, and moderate responsiveness. Known-groups comparisons indicated the discriminating potential of the RECAP for differences between groups. Moreover, bands for the RECAP scores were determined: 0-1 (completely controlled); 2-5 (mostly controlled); 6-11 (moderately controlled); 12-19 (a little controlled); 20-28 (not at all controlled). For the sake of simplicity, a single cut-off point of ≥6 was determined to identify patients whose AD is not under control. An improvement of ≥4 points should be considered as a clinically important improvement.

Most of our a-priori hypotheses for the single-score validity were confirmed, reflecting a good single-score validity. This also confirms the initial findings of previous validation studies in the UK population.[1,6,7] Furthermore, a valid instrument should also be capable of truly measuring changes in the construct it intends to assess, known as change-score validity or responsiveness. However, we only found moderate responsiveness in the present study. There are two possible explanations for this result. One is that the correlation between the changes in the RECAP and changes in the reference instruments that measure AD-specific symptoms and QoL were greater than anticipated. This may be due to the fact that domains such as symptoms and QoL inevitably became 'subdomains' of eczema control during the development of the RECAP given that eczema control is a multifaceted construct.[1,3,28] Another explanation could be related to the use of the PtGA of AD control as an anchor. The PtGA of AD control is intended to measure the same construct as the RECAP. However, the PtGA of AD control might not fully capture the contribution of AD-specific symptoms to their disease control rating over time when using a stand-alone question, while these are components of the RECAP. This discrepancy may have resulted in a weaker correlation between changes in the RECAP and changes in the PtGA of AD control than anticipated, thereby contributing to the moderate results. It is worth noting that the correlations of changes in RECAP scores with changes in PtGA of disease control were all higher compared to correlations of changes in other reference instruments with changes in PtGA of disease control, but the correlation differences for five of our hypotheses were lower than 0.1, leading to their rejection.

We employed an anchor-based approach to evaluate the interpretability of individual scores, with the PtGA of AD control serving as an anchor. The proposed RECAP banding could help to monitor to what extent patients perceive their AD control status, and thus support shared decision-making regarding treatment plans. For the sake of simplicity, we recommend a threshold of ≥6 points for single scores as a means of identifying patients whose AD is not under control. This optimal threshold may support a treat-to-target approach in clinical trials.

The interpretability of change scores was assessed using a patient-guided anchor, the GRC scale, to evaluate patient-perceived important change in eczema control. The correlation between the anchor and the change in the RECAP scores exceeded 0.50, and the anchor is explicitly linked to the definition of MIC, as defined by patients,[11] suggesting that the GRC scale is a useful anchor. Notably, all MIC values obtained with the four methods exceeded the SDC score, reflecting the ability of the RECAP to detect changes as small as the MIC value at an individual level. Although the MIC estimates varied across the methods in this study, the absolute differences were small except for the 95% limit cut-off point. Of the four MIC estimates, the predictive MIC may be the most accurate. The underlying concept of the 95% limit cut-off point is that the MIC estimate should be beyond measurement error,[29] and thus it doesn't necessarily relate to the importance of the change. The mean change method, which is based on only one subgroup reporting minor improvement with a small sample size of 10 in this study, failed to take the variability of the RECAP scores into account.[11] In many situations, the predictive modelling and the ROC curve produce identical MIC values, but recent insights have shown that the former method is more precise.[23] Meanwhile the percentage of patients who show improvement may affect the MIC, which can be corrected using predictive modelling.[24] In this study, the adjusted MIC from predictive modelling differed slightly from the ROC-based MIC (3.9 vs. 3.5). Therefore, we recommend using a threshold of ≥4 points as a clinically important change. Such outcome data could provide diverse benefits in both clinical care and research. It could help monitor eczema control over a long-term period, evaluate the effectiveness of treatments, and support shared decision-making in

both daily practice and clinical trials. In research, it could help determine the proportions of responders and possibly perform responder analyses.

A strong point of this study is its adherence to the COSMIN guidelines,[10,11] as well as the inclusion of patients across all disease severities and a high response rate. These factors likely contribute to the robustness of our findings. A limitation of this study is the lack of MIC estimates for deteriorated patients due to a small sample size of this group (n=17). In addition, the study population was restricted to adult patients in the Netherlands, which might limit its generalizability. Further research is warranted to evaluate measurement properties of the RECAP in other populations, including children and other language settings. It should be noted that the anchors employed in this study, i.e. PtGA of disease control and GRC, are not validated, as validated instruments specifically designed for these constructs do not exist.

Conclusion

The RECAP shows good single-score validity and excellent reliability. In addition, this study fills a gap on the interpretability of the RECAP. Our results indicate a threshold of ≥6 points to identify patients whose AD is 'not under control', while an improvement of ≥4 points represent a clinically important change. Given its endorsement by HOME, the results of this study support the integration of RECAP into both routine clinical practice and research settings.

References

1. 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). British Journal of Dermatology. 2020;183(3):524–36.
2. Williams HC, Schmitt J, Thomas KS, et al. The HOME Core outcome set for clinical trials of atopic dermatitis. J Allergy Clin Immunol. 2022 Jun;149(6):1899–911.
3. Leshem YA, Chalmers JR, Apfelbacher C, et al. Measuring Atopic Eczema Control and Itch Intensity in Clinical Practice: A Consensus Statement From the Harmonising Outcome Measures for Eczema in Clinical Practice (HOME-CP) Initiative. JAMA Dermatol. 2022 Dec;158(12):1429–35.

4. RECAP - The University of Nottingham [Internet]. [cited 2023 Jul 10]. Available from: https://www.nottingham.ac.uk/research/groups/cebd/resources/recap.aspx

5. 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 Dec;187(6):919–26.

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. Vol. 47, Clinical and experimental dermatology. England; 2022. p. 440–2.

7. Bhanot A, Peters TJ, Ridd MJ. Assessing the validity, responsiveness and reliability of the Recap measure of eczema control. Vol. 184, The British journal of dermatology. England; 2021. p. 955–7.

8. Gabes M, Tischer C, Herrmann A, et al. The German RECAP questionnaire: linguistic validation and cognitive debriefing in German adults with self-reported atopic eczema and parents of affected children. J Patient Rep Outcomes. 2021 Jan;5(1):13.

9. Onteniente-Gomis MM, Ortiz-Romero PL, Tous Romero F, et al. Spanish Version of the RECAP Questionnaire to Assess Control of Atopic Eczema: Translation, Cultural Adaptation, Validation, and Correlations with Other Patient-Reported Outcome Measures. Actas Dermosifiliogr. 2023 Jun;114(6):488–93.

10. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. In 2019. [cited 2023 Jul 10]. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf#

11. de Vet HC, Terwee CB, Mokkink L et al. Measurement in Medicine: a practical guide. Cambridge University Press 2011.;

12. WILLIAMS HC, JBURNEY PG, PEMBROKE AC, et al. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. British Journal of Dermatology. 1994;131(3):406–16.

13. Roqua [Internet]. [cited 2023 Jul 10]. Available from: https://www.roqua.nl/

14. Hanifin JM, Thurston M, Omoto M, et al. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. Exp Dermatol. 2001 Feb;10(1):11–8.

15. Chopra R, Vakharia PP, Sacotte R, et al. Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index and body surface area in adolescents and adults with atopic dermatitis. Br J Dermatol. 2017 Nov;177(5):1316–21.

16. Simpson E, Bissonnette R, Eichenfield LF, et al. The Validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD): The development and reliability testing of a novel clinical outcome measurement instrument for the severity of atopic dermatitis. J Am Acad Dermatol. 2020 Sep;83(3):839–46.

17. Pariser DM, Simpson EL, Gadkari A, et al. Evaluating patient-perceived control of atopic dermatitis: design, validation, and scoring of the Atopic Dermatitis Control Tool (ADCT). Curr Med Res Opin. 2020;36(3):367–76.

18. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147–57.

19. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. J Clin Epidemiol. 2006 Oct;59(10):1033–9.

20. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34–42.

21. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999 Jun;8(2):135–60.

22. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008 Feb;61(2):102–9.

23. 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 Dec;68(12):1388–96.

24. 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 Mar;83:90–100.

25. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995 Aug;4(4):293–307.

26. 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 Dec;12:176.

27. 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 Oct;30(10):2729–54.

28. Howells L, Thomas KS, Sears A V., et al. Defining and measuring 'eczema control': an international qualitative study to explore the views of those living with and treating atopic eczema. Journal of the European Academy of Dermatology and Venereology. 2019;33(6):1124–32.

29. De Vet HCW, Ostelo RWJG, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Quality of Life Research. 2007;16(1):131–42.

Figure legends

Figure 1. Study flow diagram. N, number; GRC, global rating of change. [†]136 patients reported no change based on the GRC scale at T_1; of those 112 patients filled out the T_1 questionnaires within 1-3

days. ‡GRC scale was not included in the package of questionnaires at T_2 at the first 5 months of data collection.

Figure 2. Visual anchor-based distribution of raw RECAP change scores among patients with importantly improved (green line) and unchanged (blue dashed line) scores on the anchor (Global Rating of Change (GRC) scale), along with the minimally important change (MIC) values obtained from four methods. ROC, receiver operating characteristic.

Table 1. Overview of longitudinal study design.

T_0 Baseline (on site) Single-score validity, known-groups validity, Interpretability – single scores	**T_1 after 1-3 days (at home)** Reliability, Interpretability - SDC	**T_2 after 4-12 weeks (at home)** Responsiveness, Interpretability – MIC
Completed by participants • Demographics - Age - Age of onset - Sex • RECAP • Disease severity of atopic dermatitis - POEM - PtGA of AD severity • Skin-specific HRQoL - DLQI - Skindex-29 • Generic HRQoL - EQ-5D-5L • Patient-reported symptoms - NRS for peak itch - NRS for eczema-related sleep disturbance • Anchor question - PtGA of AD control **Completed by physicians** • Eczema Area and Severity Index (EASI) • Validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD)	**Completed by participants** • RECAP • Global Rating of Change Scale • Anchor question - PtGA of AD control	**Completed by participants** • RECAP • Disease severity of atopic dermatitis - POEM - PtGA of AD severity • Skin-specific HRQoL - DLQI - Skindex-29 • Generic HRQoL - EQ-5D-5L • Patient-reported symptoms - NRS for peak itch - NRS for eczema-related sleep disturbance • Anchor question - PtGA of AD control • Global Rating of Change Scale

Abbreviation: POEM, Patient-Oriented Eczema Measure; PtGA, Patient Global Assessment; AD, atopic dermatitis; HRQoL, Health-related Quality of Life; DLQI, Dermatology Life Quality Index; EQ-5D-5L, quality-of-life questionnaire of the EuroQol Group; NRS, numeric rating scale; EASI, Eczema Area and Severity Index; vIGA, validated Investigator Global Assessment; SDC, smallest detectable change; MIC, minimally important change.

1 **Table 2.** Basic characteristics of study population stratified by sex at T_0.

	Total, n (%) N=200	Male, n (%) N=115	Female, n (%) N=85	P-value[†]
Age, years, mean (SD)	38.5 (14.5)	40.6 (13.7)	35.6 (15.0)	**0.01**
Missing, n	*0*	*0*	*0*	
Age of onset				
Early onset (0-2y)	124 (62.3)	67 (58.8)	57 (67.1)	0.23
Childhood onset (3-11y)	42 (21.1)	25 (21.9)	17 (20.0)	0.74
Adolescent onset (12-17y)	12 (6.0)	7 (6.1)	5 (5.9)	0.94
Adult onset (18-50y)	18 (9.0)	14 (12.3)	4 (4.7)	0.07
Late onset (>50y)	3 (1.5)	1 (0.9)	2 (2.4)	0.40
Missing, n	*1*	*1*	*0*	
EASI, Mean (SD)	9.9 (9.8)	10.2 (10.9)	9.5 (8.2)	0.73
Clear (0)	4 (2.1)	2 (1.8)	2 (2.6)	0.71
Mild (0.1-5.9)	83 (44.1)	54 (48.6)	29 (37.7)	0.14
Moderate (6.0-22.9)	79 (42.0)	38 (34.2)	41 (53.2)	**0.01**
Severe (23.0-72)	22 (11.7)	17 (15.3)	5 (6.5)	0.06
Missing, n	*12*	*4*	*8*	
vIGA				
Clear/almost clear	41 (23.3)	27 (26.2)	14 (19.2)	0.28
Mild	41 (23.3)	26 (25.2)	15 (20.5)	0.47
Moderate	54 (30.7)	26 (25.2)	28 (38.4)	0.06
Severe	40 (22.7)	24 (23.3)	16 (21.9)	0.83
Missing, n	*24*	*12*	*12*	
PtGA of AD severity				
Clear	8 (4.0)	6 (5.2)	2 (2.4)	0.31
Mild	74 (37.2)	49 (42.6)	25 (29.8)	0.06
Moderate	52 (26.1)	33 (28.7)	19 (22.6)	0.34
Severe	52 (26.1)	21 (18.3)	31 (36.9)	**<0.01**
Very severe	13 (6.5)	6 (5.2)	7 (8.3)	0.38
Missing, n	*1*	*0*	*1*	
POEM, Mean (SD)	13.8 (8.0)	12.5 (7.9)	15.6 (7.9)	**0.01**
Clear/almost clear	16 (8.0)	12 (10.4)	4 (4.7)	0.14
Mild	39 (19.5)	25 (21.7)	14 (16.5)	0.35
Moderate	68 (34.0)	42 (36.5)	26 (30.6)	0.38
Severe	54 (27.0)	25 (21.7)	29 (34.1)	**0.05**
Very severe	23 (11.5)	11 (9.6)	12 (14.1)	0.32
Missing, n	*0*	*0*	*0*	
PtGA of AD control				
Not at all controlled	43 (21.6)	17 (14.8)	26 (31.0)	**0.01**
A little controlled	38 (19.1)	19 (16.5)	19 (22.6)	0.28
Moderately controlled	47 (23.6)	29 (25.2)	18 (21.4)	0.53
Mostly controlled	53 (26.6)	37 (32.2)	16 (19.0)	**0.04**
Completely controlled	18 (9.0)	13 (11.3)	5 (6.0)	0.19
Missing, n	*1*	0	*1*	
RECAP				
Median (IQR)	11.0 (14.0)	8.0 (11.0)	13.0 (12.0)	**<0.001**
Mean (SD)	11.5 (8.0)	9.6 (7.6)	14.1 (7.8)	**<0.001**
Missing, n	*0*	0	0	
DLQI, Mean (SD)	6.0 (10.0)	6.5 (6.8)	9.5 (7.3)	**<0.001**
0-1 (no impact)	39 (19.5)	29 (25.2)	10 (11.8)	**0.02**

2-5 (small impact)	60 (30.0)	38 (33.0)	22 (25.9)	0.28
6-10 (moderate impact)	44 (22.0)	25 (21.7)	19 (22.4)	0.92
11-20 (very large impact)	42 (21.0)	15 (13.0)	27 (31.8)	**<0.01**
21-30 (extremely large impact)	15 (7.5)	8 (7.0)	7 (8.2)	0.73
Missing, n	*0*	*0*	*0*	
Skindex-29, Mean (SD)	41.3 (22.6)	36.3 (21.8)	48.1 (21.9)	**<0.001**
Missing, n	*1*	*0*	*1*	
EQ-5D-5L, Mean (SD)				
Value score	65.7 (18.9)	68.2 (17.9)	62.3 (19.7)	**0.02**
VAS score	0.7 (0.2)	0.8 (0.2)	0.7 (0.3)	**0.01**
Missing, n	*2*	*1*	*1*	
NRS Peak itch, Mean (SD)	4.9 (3.1)	4.3 (2.9)	5.7 (3.0)	**<0.01**
Missing, n	*8*	*5*	*3*	
NRS Sleep disturbance, Mean (SD)	2.8 (3.2)	2.2 (3.0)	3.6 (3.4)	**0.01**
Missing, n	*1*	*0*	*1*	

[†]Continuous variables according to a Mann Whitney U or median test, and categorical variables according to a Chi-Square test; significant P values (<0.05) are in bold.
Abbreviation: N, number; SD, standard deviation; y, year; EASI, Eczema Area and Severity Index; vIGA, validated Investigator Global Assessment; PtGA, Patient Global Assessment; AD, atopic dermatitis; POEM, Patient-Oriented Eczema Measure; DLQI, Dermatology Life Quality Index; EQ-5D-5L, quality-of-life questionnaire of the EuroQol Group; VAS, visual analogue scale; NRS, numeric rating scale.

Table 3. Single-score validity (at T_0) correlations between the RECAP and reference instruments.

Reference instruments	Correlation hypothesized[†]	Correlation found (r)	R^2	Hypotheses confirmed?
EASI	++	0.67	0.39	Yes
vIGA	++	0.68	0.44	Yes
PtGA of AD severity	+++	0.84	0.71	Yes
PtGA of AD control	+++	-0.82[ǂ]	0.65	Yes
POEM	+++	0.89	0.79	Yes
DLQI	+++	0.89	0.77	Yes
Skindex-29	+++	0.86	0.76	Yes
EQ-5D-5L (Value score)	+	-0.54[ǂ]	0.30	No
EQ-5D-5L (VAS score)	+	-0.52[ǂ]	0.38	No
NRS Peak itch	+++	0.89	0.76	Yes
NRS Sleep disturbance	+++	0.78	0.68	Yes
Total amount of hypotheses that were rejected	2/11 (18%)			

[†]Strong correlation (+++) is defined as r > 0.7; moderate correlation (++) as 0.4 < r <0.7; and weak correlation (+) as 0.2 < r <0.4, using Spearman's rho (r).
[ǂ]Negative value due to both the PtGA of AD control and EQ-5D-5L being scored inversely to the RECAP.
Abbreviation: EASI, Eczema Area and Severity Index; vIGA, validated Investigator Global Assessment; PtGA, Patient Global Assessment; AD, atopic dermatitis; POEM, Patient-Oriented Eczema Measure; DLQI, Dermatology Life Quality Index; VAS, visual analogue scale; EQ-5D-5L, quality-of-life questionnaire of the EuroQol Group; NRS, numeric rating scale.

Table 4. Responsiveness between T_0 and T_2.

	Correlations found	**Hypotheses confirmed?**
Hypothesis on correlations[*]		
Change RECAP – Change PtGA of AD control > Change POEM - Change PtGA of AD control	-0.67[†] vs. -0.63[†]	No

Change RECAP – Change PtGA of AD control > Change PtGA of AD severity - Change PtGA of AD control	-0.67[†] vs. -0.60[†]	No
Change RECAP – Change PtGA of AD control > Change DLQI - Change PtGA of AD control	-0.67[†] vs. -0.60[†]	No
Change RECAP – Change PtGA of AD control > Change Skindex-29 - Change PtGA of AD control	-0.67[†] vs. -0.60[†]	No
Change RECAP – Change PtGA of AD control > Change EQ-5D Value - Change PtGA of AD control	-0.67[†] vs. 0.41	Yes
Change RECAP – Change PtGA of AD control > Change EQ-5D VAS - Change PtGA of AD control	-0.67[†] vs. 0.42	Yes
Change RECAP – Change PtGA of AD control > Change NRS Peak itch - Change PtGA of AD control	-0.67[†] vs. -0.59[†]	No
Change RECAP – Change PtGA of AD control > Change NRS Sleep disturbance - Change PtGA of AD control	-0.67[†] vs. -0.53[†]	Yes
Hypothesis according to COSMIN		
Instruments measuring similar constructs (≥0.50)		
Change RECAP – Change PtGA of AD control	-0.67[†]	Yes
Change RECAP – Change NRS Peak itch	0.71	Yes
Change RECAP – Change NRS Sleep disturbance	0.65	Yes
Change RECAP – Change POEM	0.74	Yes
Change RECAP – Change PtGA of AD severity	0.69	Yes
Instruments measuring related, but dissimilar constructs (0.30-0.50)		
Change RECAP – Change Skindex-29	0.74	No
Change RECAP – Change EQ-5D Value	-0.44[‡]	Yes
Change RECAP – Change EQ-5D VAS	-0.44[‡]	Yes
Change RECAP – Change DLQI	0.78	No
Total amount of hypotheses that were rejected	7/17 (41%)	

[*]A correlation difference of ⩾0.1 was deemed relevant and thus hypothesis confirmed.

[†]Negative value due to the PtGA of AD control being scored inversely to the RECAP and other reference instruments except EQ-5D-5L.

[‡]Negative value due to the EQ-5D-5L being scored inversely to the RECAP.

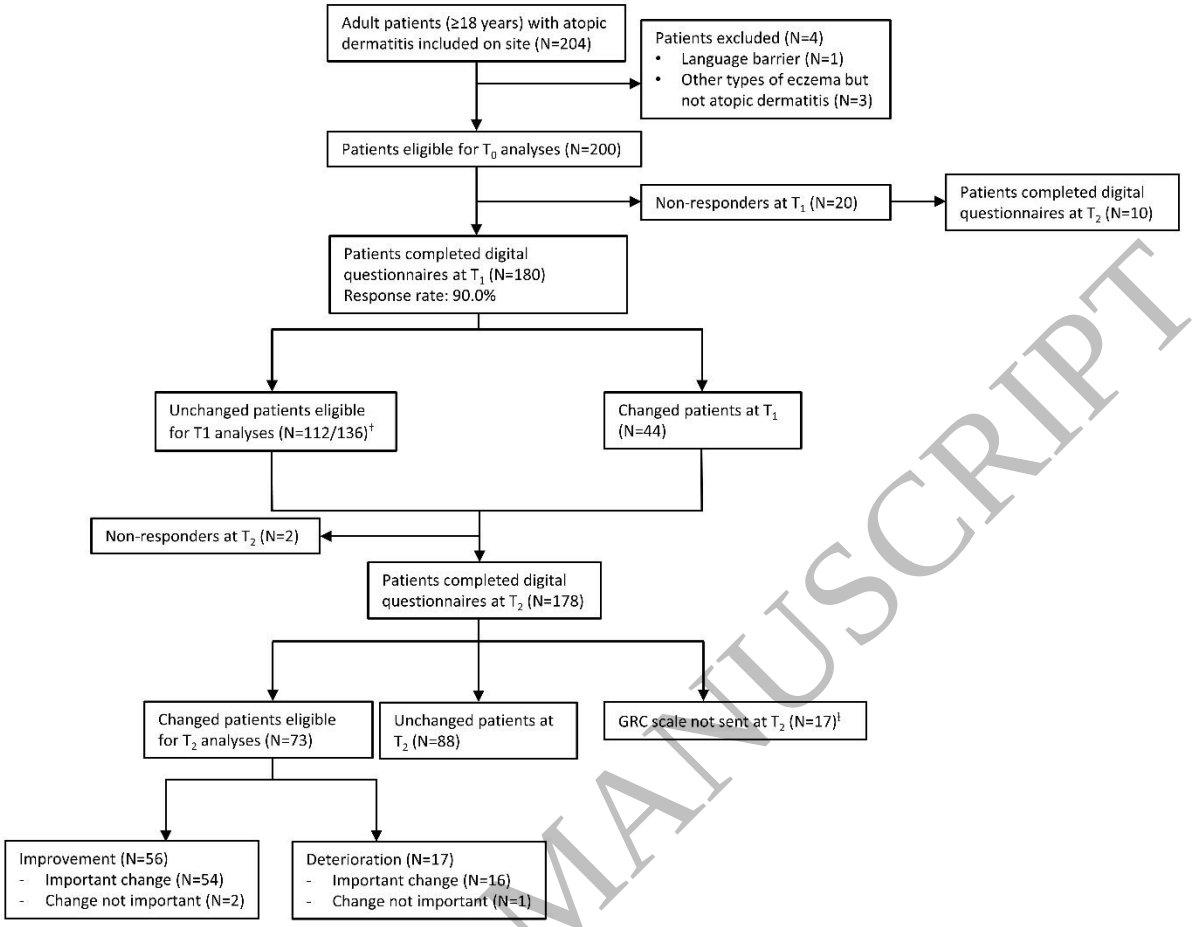

Figure 1
221x174 mm (x DPI)

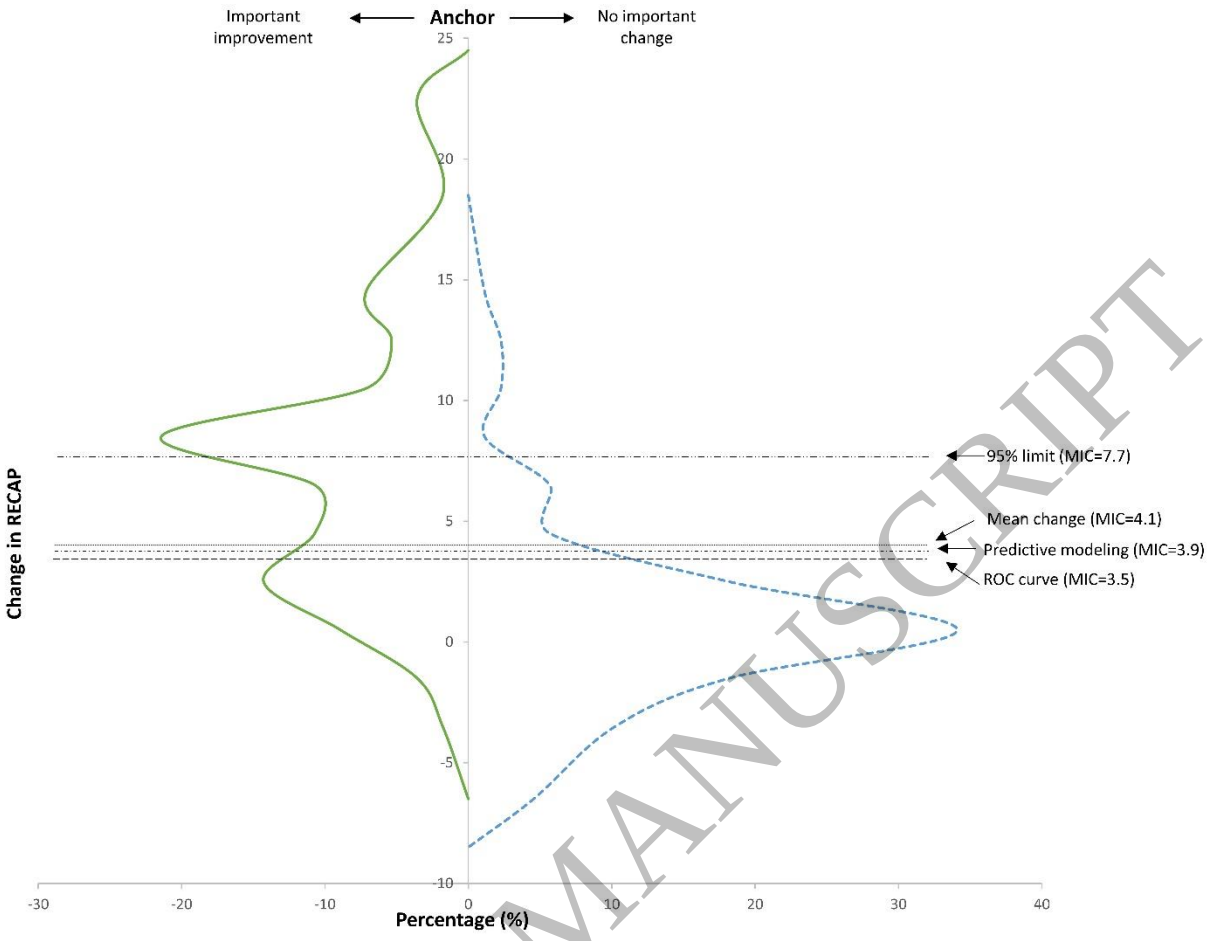

Figure 2
238x183 mm (x DPI)