REVIEW

Strategies used for measuring long-term control in atopic dermatitis trials: A systematic review

Sebastien Barbarot, PhD,[a,b] Natasha K. Rogers, PhD,[a] Katrina Abuabara, MD,[c] Helene Aubert, MD,[b] Joanne Chalmers, PhD,[a] Carsten Flohr, PhD,[d] Jon Hanifin, MD,[e] Luigi Naldi, MD,[f] David J. Margolis, MD,[g] Carle Paul, MD,[h] Matthew J. Ridd, PhD,[i] Marie-Louise Anna Schuttelaar, PhD,[j] Eric Simpson, MD,[e] Marie Tauber, MD,[h] Annika Volke, MD,[k] Stephan Weidinger, MD,[l] Sally R. Wilkes, PhD,[a] Andreas Wollenberg, PhD,[m] and Kim S. Thomas, PhD[a]

Nottingham, London, and Bristol, United Kingdom; Nantes, France; San Francisco, California; Portland, Oregon; Bergamo, Italy; Philadelphia, Pennsylvania; Toulouse, France; Groningen, The Netherlands; Tartu, Estonia; and Kiel and Munich, Germany

Background: Atopic dermatitis (AD) is a chronic inflammatory skin disease. There are no standardized methods for capturing long-term control of AD.

Objective: We sought to identify how long-term control has been captured in published randomized controlled trials (RCTs). Results will initiate consensus discussions on how best to measure long-term control in the core outcome set for AD.

Methods: We conducted a systematic review of RCTs of AD treatments published between 2000 and 2013, with a follow-up period of 3 months or longer, at least 1 outcome measure recorded at 3 or more time points, full article available, and published in English.

Results: In all, 101 of 353 RCTs were eligible. Methods to capture long-term control included: repeated measurement of AD outcomes (92 RCTs; 91%), use of AD medication (29 RCTs; 28.7%), and AD flares/remissions (26 RCTs; 25.7%). Repeated measurements of AD outcomes were typically collected 3 to 5 times during a trial, but analysis methods often failed to make best use of the data. Time to first flare was most commonly used for trials including flare data (21/52). Medication use was recorded based on quantity, potency, and frequency of application.

Limitations: We included RCT data only.

Conclusion: This review illustrates the difficulties in measuring long-term control, and points to the need for improved harmonization of outcomes. (J Am Acad Dermatol http://dx.doi.org/10.1016/j.jaad.2016.05.043.)

Key words: atopic dermatitis; atopic eczema; flares; long-term control; outcome measures; randomized controlled trials; systematic review.

From the Center of Evidence-based Dermatology, University of Nottingham[a]; Department of Dermatology, Centre Hospitalier Universitaire (CHU) de Nantes[b]; Department of Dermatology, University of California, San Francisco[c]; Population-based Dermatology Research Unit, St John's Institute of Dermatology, Guy's and St Thomas' National Health Service Foundation Trust and King's College London[d]; Department of Dermatology, Oregon Health and Science University[e]; Study Center Italian Group for Epidemiologic Research in Dermatology and Department of Dermatology, Azienda Ospedaliera Papa Giovanni XXIII[f]; Departments of Dermatology and of Biostatistics and Epidemiology, University of Pennsylvania, Perelman School of Medicine[g]; Dermatology, Paul Sabatier University and Hôpital Larrey, Toulouse[h]; School of Social and Community Medicine, University of Bristol[i]; Department of Dermatology, University Medical Center Groningen, University of Groningen[j]; Department of Dermatology, University of Tartu[k]; Department of Dermatology, Venereology, and Allergy, University Hospital Schleswig-Holstein, Campus Kiel[l]; and Department of Dermatology and Allergy, Ludwig-Maximilian University Munich.[m]

Dr Ridd is funded by National Institute for Health Research (NIHR) Post Doctoral Fellowship (PDF-2014-07-013). The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

Conflicts of interest: None declared.

Accepted for publication May 31, 2016.

Reprint requests: Kim S. Thomas, PhD, University of Nottingham, King's Meadow Campus, Lenton Lane Nottingham, NG7 2NR United Kingdom. E-mail: kim.thomas@nottingham.ac.uk.

Published online August 10, 2016.

0190-9622/$36.00

http://dx.doi.org/10.1016/j.jaad.2016.05.043

Atopic dermatitis (AD) (atopic eczema) is a highly prevalent, itchy, inflammatory skin condition that affects children and adults. As with other chronic inflammatory diseases, AD severity tends to wax and wane over time, with periods of relative remission, interspersed with periods of increased disease activity or "flare."[1] AD treatments aim to reduce disease intensity, minimize the number of flares, and increase the duration of remissions. The ability to measure long-term control of AD over time is an important outcome when evaluating effectiveness of treatments, as this reflects patients' experiences of living with the condition, and long-term control has been identified as a core outcome to be included in future AD clinical trials.[2]

CAPSULE SUMMARY

- There is no consensus over how best to measure long-term control of atopic dermatitis in clinical trials.
- To date, repeated measurement of eczema severity, assessment of flares, and use of atopic dermatitis medications have all been used.
- Consensus agreement of core outcome sets for atopic dermatitis will improve evidence-based practice.

To date, there is little consensus over how best to capture long-term control in AD. Two systematic reviews have demonstrated the variability in AD flare definitions used in published studies,[3,4] and have highlighted the methodological challenges in capturing AD flares. Other approaches to capture long-term control include measurement of anti-inflammatory medication use over time, or the repeated measurement of AD severity and other health outcomes.

The Harmonizing Outcome Measures for Eczema (HOME) initiative (www.homeforeczema.org) identified long-term control as 1 of 4 key domains to measure in all clinical trials in AD. The current systematic review has been conducted to inform the HOME initiative's consensus discussions on how long-term control has been captured in previously published randomized controlled trials (RCTs). It represents stage 1 on the HOME Roadmap,[5] namely to identify available outcome instruments for capturing the domain of interest.

METHODS

This systematic review was performed according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) recommendations.[6] The protocol was agreed to before starting the review, and registered online (October 6, 2014) (http://nottingham.ac.uk/research/groups/cebd/documents/researchdocs/ltc-protocol-final.pdf).

Eligibility criteria and search strategy

We searched for RCTs with at least a 3-month follow-up period[7] that included adults or children with AD, and were published between January 1, 2000, and March 12, 2013. This period was chosen as, before 2000, most AE trials were of relatively short duration.[8] Eligible studies were identified using the Global Resource of Eczema Trials (GREAT) database (www.greatdatabase.org.uk). This freely available online database contains records of RCTs for AD treatments found within MEDLINE, EMBASE, CINAHL, AMED, LILACS, the Cochrane Library, and the Skin Group Specialized Register databases.

The search strategy used to identify RCTs in the GREAT database and validation of the GREAT database have been published elsewhere.[9] Observational studies were not included in this review because of time and resource limitations.

Study selection and data extraction

Inclusion criteria were predefined. Studies were included if the duration of patient follow-up was 3 months or longer, and a clinician- or patient-reported outcome measure was recorded at 3 or more time points. We excluded studies published in abstract form only, that did not include clinical outcomes (eg, studies only containing data pertaining to biomarkers or skin barrier function tests), and not published in English. Titles of studies were retrieved and the full text was then obtained and screened against the inclusion criteria by 2 authors (N. K. R. and S. B.). Responses were compared and discrepancies resolved by consensus (N. K. R. and S. B.).

Studies that met the inclusion criteria were divided between author pairs, who independently extracted data using a standardized data extraction form. Details were extracted for: (1) trial attributes (size of trial, age of participants); (2) repeated measurement of clinician- or patient-reported AD outcomes over time; (3) use of AD medication—defined as any treatment used to control AD symptoms other than the randomly allocated intervention; and (4) AD flares/relapse—defined as a decline in condition (worsening of symptoms) that met 1 of the recommended descriptions of flare,[3] regardless of whether "flare," "relapse," or "remission" was specifically used within the text. For all long-term control outcomes, details of how the outcomes were recorded, analyzed, and presented in the article were recorded. Data extraction forms

Abbreviations used:

AD: atopic dermatitis
GREAT: Global Resource of Eczema Trials
HOME: Harmonizing Outcome Measures for Eczema
IGA: Investigator Global Assessment
RCT: randomized controlled trial

were reviewed by another 2 authors (N. K. R. and S. R. W.), who checked for completeness and resolved any discrepancies by referring to the original trial publications.

Results were summarized qualitatively, and the statistical techniques used in the original trial reports were reviewed by a medical statistician to ascertain the appropriateness of the analysis techniques used. The analyses techniques described in the trial reports were categorized into "efficient analysis techniques" (best use of all available date); "inefficient analysis techniques" (statistically correct, but potentially inefficient use of available data); "inappropriate analysis techniques" (analysis of multiple time points individually without adjustment for multiple testing); or "unclear."

RESULTS

A search of the GREAT database for studies published between January 1, 2000, and March 12, 2013, yielded a total of 353 RCTs (Fig 1). Overall, 101 trials were included in the review (67% included either children or adults, 31% included both children and adults, 1 trial did not state the ages of the participants involved). Nearly all trials were conducted in a secondary or tertiary care setting.

Types of long-term control outcomes used

Long-term control outcomes were measured in a variety of ways, and 72 trials (71.2%) measured long-term control in 2 or more ways. In 92 trials (91%), repeated measurements of clinical or patient-reported outcomes were reported; in 26 trials (25.7%), AD flares were captured as an outcome measure; and in 29 trials (28.7%), the use of AD medication was used to measure long-term control. In all cases there was considerable heterogeneity in how the outcomes were defined and captured.

Of the studies assessed, 68 of 101 (67.3%) had at least 1 graphic representation of long-term data.

Repeated measurement of AE outcomes. A total of 196 outcomes were used in the 92 trials that reported repeated measurement of AD outcomes (median 1.9 per trial) (Fig 2). The most commonly used outcomes were: Scoring Atopic Dermatitis (SCORAD) or objective SCORAD (25%), quality-of-life scales (14%), pruritus scales (10%), body surface area (8%), Eczema Area and Severity Index or modified Eczema Area and Severity Index (8%), and Investigator Global Assessment (IGA) (7%). As previously shown, there was large variability in IGA definitions between studies.[10] The breakdown of clinician- and patient-reported outcomes is summarized in Fig 3.

Outcomes were most often collected on a monthly basis (40% monthly, 27% more than a month apart, 25% irregular intervals, 6% weekly, 0.5% daily). Most trials (66/92, 71%) collected the outcomes between 3 and 5 times over the duration of the trial, with 11 trials including 11 or more data collection points.

Medication use. The use of AD medications as an indicator of disease control (rather than adherence with study medications), was collected by less than a third of included trials (29/101), and only 4 reported this information as a primary outcome. Topical corticosteroid use was assessed in all 29 of these trials, but some trials also monitored other types of medication, including: antibiotics (n = 5); antihistamines (n = 5); calcineurin inhibitors (n = 4); emollients (n = 2); and systemic therapy (n = 2). Information was documented solely during visits for just over half of the studies (15/29, 52%), with a minority collecting data on medication use from participant diaries (4/29, 14%), or a combination of clinic visits and participant diaries (3/29, 10%). The remaining studies did not give any details about the collection method (7/29, 24%). None of the included trials that provided details of data collection gathered information from medical notes. The manner in which medication use was captured varied considerably and included measurement of frequency of application, amount of medication used, and potency (Fig 4).

AD flares. For 26 of 101 (25%) included trials, the concept of disease flares (including relapse/remission) was captured, and for 15 (58%) of these, flares were the primary outcome. In line with previously suggested categorizations for flare outcomes,[3] 9 of 26 (35%) used an arbitrary cut-off such as a change in score from a baseline measurement (eg, IGA >4 or SCORAD >75% of baseline), 6 of 26 (23%) used a behavioral measure such as the need for stepping up topical steroid treatment (rescue medication) according to the patient or the physician, 9 of 26 (35%) used a composite measure (eg, IGA >4 AND the need for recue medication), and 2 of 26 (7%) were classed as other/unknown. Data on flares were most commonly collected during clinic visits (14/26, 53%), with only 6 of 26 (23%) being collected from participants at home.

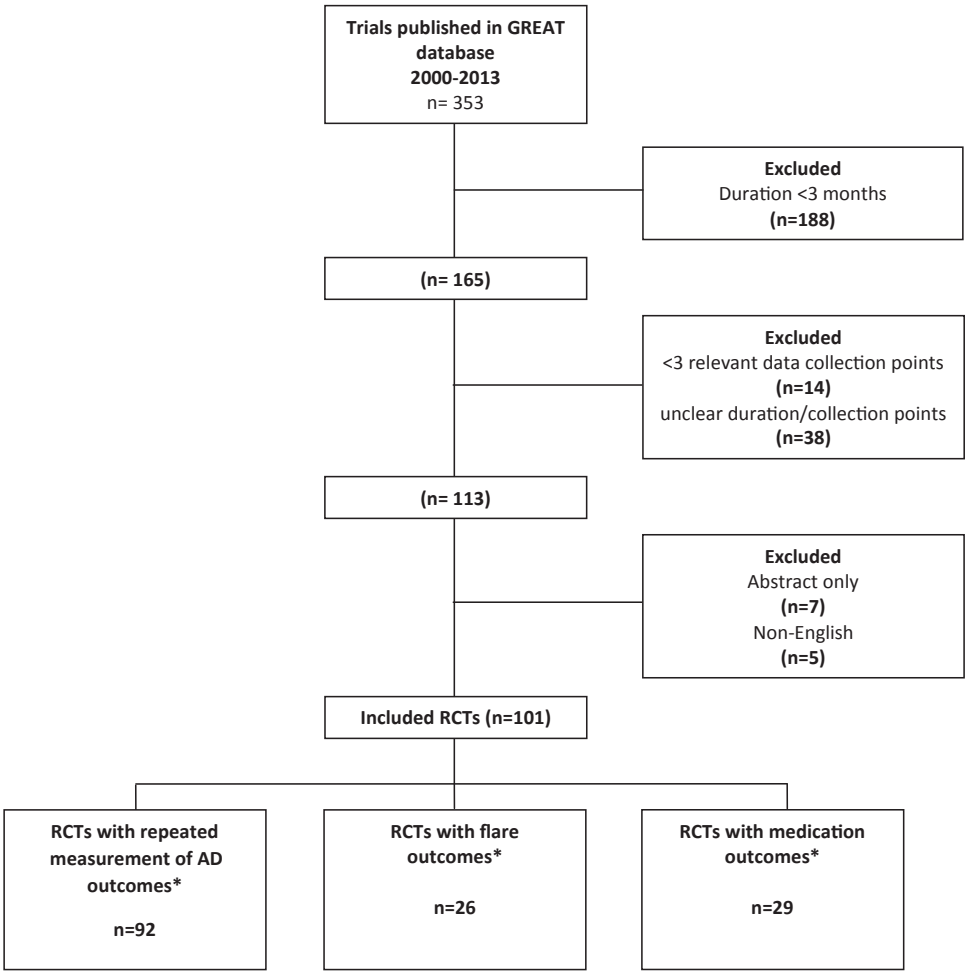

Fig 1. LTC flow diagram. *AD*, Atopic dermatitis; *GREAT*, Global Resource of Eczema Trials; *LTC*, long term control systematic review flow; *RCT*, randomized controlled trial.

Most trials analyzed flares in multiple ways, with a total 52 analyses performed (Table I). Time to first flare was the most commonly used summary measure (21/52 analyses), followed by number of flares (17/52 analyses).

Data analysis techniques used. Despite considerable efforts having been taken to collect long-term control outcome data throughout these trials, only 72 of 196 (37%) of the reported analyses made best use of the available data and included all time points in the analysis (Table II). Analyses considered to be best use of the data included: analysis of variance (n = 35 analyses), linear mixed model (n = 13 analyses), analysis of covariance (n = 12 analyses), nonlinear mixed model (n = 2 analyses), nonparametric repeated measures (n = 2 analyses), area under the curve (n = 1 analysis), log-rank test (n = 1 analysis), McNemar (n = 1 analysis), and other (n = 5 analyses).

DISCUSSION

Main findings

This review shows how previous researchers have tackled the measurement of long-term control in published RCTs of AD treatments, and serves to highlight some of the complexities of measuring disease control over time.

Because almost all of the trials used repeated measurement of clinician- or patient-reported outcomes over time, it would appear that such an approach is both feasible and acceptable. However, appropriate analysis of these data is challenging, and few trials reported their results in the most appropriate and efficient manner. The analysis of repeated measures requires the use of specific statistical tests (eg, analysis of variance, analysis of covariance, or mixed models). Using multiple tests to compare data between groups at each time points leads to increased risk of identifying a significant difference by chance. The fact that 39.7% of the

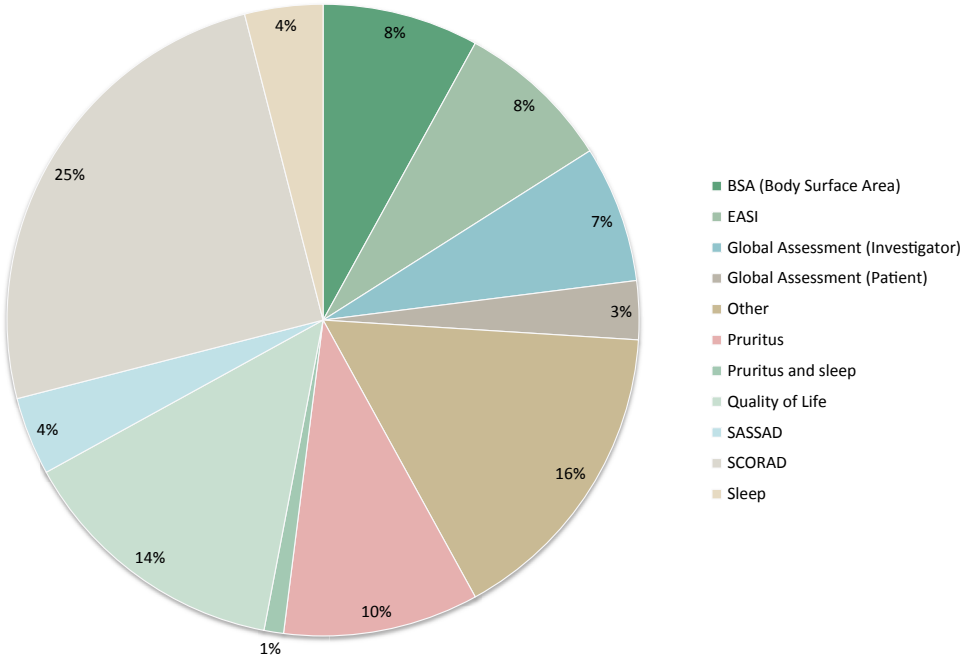

Fig 2. Distribution of the 196 outcomes used in 92 trials that reported repeated measurement of atopic dermatitis outcomes. *EASI*, Eczema Area and Severity Index, *SASSAD*, Six Area, Six Sign Atopic Dermatitis; *SCORAD*, Scoring Atopic Dermatitis.

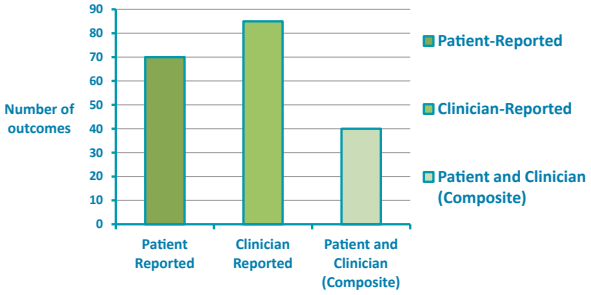

Fig 3. Number of patient- and clinician-reported outcomes used in the included trials.

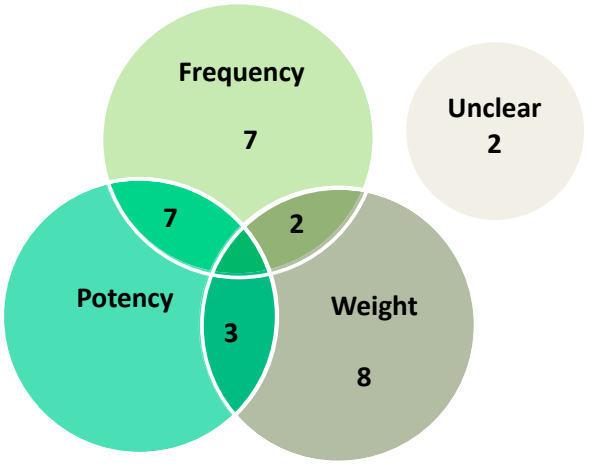

Fig 4. Methods of collection for medication use.

reported analyses described in this review were performed using inappropriate statistical techniques, such as repeated significance testing at multiple time points (without adjustment for multiple testing),[11] is something that the dermatology research community and academic journals could do more to address.

We chose to report medication usage and analysis of flares separately. However, these concepts are often linked, as incidence of flares may be inversely related to the amount of anti-inflammatory medication used, and flare definitions commonly rely on the concept of escalation of therapy as an indicator of worsening disease.[12] Similarly, worsening disease severity as captured by validated severity scales used repeatedly over time are likely to be capturing disease flares as experienced at specific time points. Further work is required to establish whether choosing one option over another is likely to miss a fundamental aspect of disease control that is important to patients.

In considering the suitability of different methods for capturing long-term control, several issues are relevant. The need for chosen outcomes to be feasible in all trial settings is crucial when selecting measurement instruments for a core outcome set, and this can be a particular challenge when evaluating long-term control, which can be resource intensive and difficult to interpret.[12]

Table I. Summary of methods to analyze flare outcomes

Analysis	No.
Time to first flare	21
No. of flares	17
Duration of remission	5
Duration of flare	4
"Totally controlled weeks" and "well-controlled weeks"	1
Other	4

Table II. Summary of methods of analysis for repeated measures data

Appropriateness of analysis	Category	No. (%)
Best use of data	Took into account all time points in single analysis	73 (37.2)
Inefficient analysis	Only compared baseline and end point	30 (15.3)
Inefficient analysis	Only data at a single time point are assessed	4 (2.0)
Inappropriate analysis	Compared each time point to baseline individually	71 (36.2)
Inappropriate analysis	Compared groups at each individual time point	7 (3.5)
Not analyzable	Unclear	11 (5.6)

Equally important is the concept that outcomes should be relevant to patients with all severities of disease and health care settings. Most patients with AD are treated in primary care and have relatively mild disease. As such, many patients are controlled with emollients only and rarely experience severe flares. In this setting, judging treatment response based on the amount of topical corticosteroid used, or the number of flares experienced over periods of a few months, is unlikely to be an efficient trial design because of low event rates. Similarly, for patients with very severe disease who require systemic medication, or who experience fewer fluctuations in their disease severity, the concept of disease control defined by topical corticosteroid use or number of flares may be less useful.

The optimum frequency of outcome assessments (eg, daily, weekly, monthly, or bi-monthly) has yet to be established, and will no doubt be determined by the feasibility of outcome assessments. For patient-reported outcomes, more frequent data collection may be possible through the use of apps or other online data collection tools,[13] thus facilitating data collection between clinic visits. By contrast, long-term control measured by independent observers during clinic visits or at participants' homes will, by necessity, limit the number and timing of outcome assessments.

As a chronic, relapsing condition, AD has many similarities with other inflammatory conditions such as asthma and rheumatoid arthritis, where considerable efforts are now being made to establish working definitions for disease flares.[14-17] An agreed upon definition of disease flare (or remission) as part of the outcome domain for long-term control would be a helpful step forward, and consistency in assessing AD long-term control in RCTs and observational studies will improve the comparability of research, thus benefitting patients and health care providers. It is also disappointing that over half of the identified trials had to be excluded from this review as they were of less than 3 months' duration, making assessment of long-term control impossible.

Strengths and limitations

This review sought to summarize the current approaches used in previously published AD RCTs to capture long-term control of AD. However, this approach means that more recent trends in data collection may have been missed as the included trials will all have been conceived and designed several years ago. Similarly, by excluding observational studies, it is possible that alternative means of capturing long-term control of AD have been missed. This review was also unable to comment directly on the feasibility of different approaches, or on the practical difficulties encountered from the methods used.

What does this mean for the HOME initiative and for future research?

This systematic review has been conducted on behalf of the Long-Term Control Working Group for the HOME initiative and represents the first step in defining how best to measure long-term control in clinical trials as part of the core outcome set for AD. A review of validation studies that have evaluated outcomes for long-term control will be conducted, along with a suite of studies to address known research gaps, including validity and responsiveness of different approaches to capturing long-term control, and the optimum timing of outcome assessments.

The HOME initiative has already achieved international consensus that clinical signs should be captured using the Eczema Area and Severity

Index[7,18] and that patient-reported symptoms should be captured using the Patient-Oriented Eczema Measure. As such, in the absence of an agreed upon instrument for capturing long-term control, we recommend an interim solution of using at least 1 of these scales at multiple time points (preferably at least monthly for a minimum of 3 months). The analysis of the data should be done using appropriate statistical techniques that take into account all time points in a single analysis. If possible, it would be ideal to use the HOME core outcome instruments for signs and symptoms alongside measures of disease flare or topical medication use, as this would provide additional data to inform future consensus agreement over the best way to measure long-term control.

This study has been conducted in support of the Harmonizing Outcome Measures for Eczema (HOME) initiative, and we thank HOME members who helped to inform the development and concepts described in this study. For full details of the HOME initiative see: www.homeforeczema.org.

REFERENCES

1. Weidinger S, Novak N. Atopic dermatitis. *Lancet*. 2016;387:1109-1122.
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;67:1111-1117.
3. Langan SM, Schmitt J, Williams HC, Smith S, Thomas KS. How are eczema 'flares' defined? A systematic review and recommendation for future studies. *Br J Dermatol*. 2014;170:548-556.
4. Langan SM, Silcocks P, Williams HC. What causes flares of eczema in children? *Br J Dermatol*. 2009;161:640-646.
5. Schmitt J, Apfelbacher C, Spuls PI, et al. The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol*. 2015;135:24-30.
6. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
7. Chalmers JR, Schmitt J, Apfelbacher C, et al. Report from the third international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME). *Br J Dermatol*. 2014;171:1318-1325.
8. Hoare C, Li Wan Po A, Williams H. Systematic review of treatments for atopic eczema. *Health Technol Assess*. 2000;4:1-191.
9. Nankervis H, Maplethorpe A, Williams HC. Mapping randomized controlled trials of treatments for eczema—the GREAT database (the Global Resource of EczemA Trials: a collection of key data on randomized controlled trials of treatments for eczema from 2000 to 2010). *BMC Dermatol*. 2011;11:10.
10. Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: many options, no standards. *J Am Acad Dermatol*. 2016;74:288-294.
11. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ*. 1990;300:230-235.
12. Thomas KS, Stuart B, O'Leary CJ, et al. Validation of treatment escalation as a definition of atopic eczema flares. *PLoS One*. 2015;10:e0124770.
13. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012;21:1305-1314.
14. Bartlett SJ, Hewlett S, Bingham CO III, et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis*. 2012;71:1855-1860.
15. van der Maas A, den Broeder AA. Measuring flares in rheumatoid arthritis. (Why) do we need validated criteria? *J Rheumatol*. 2014;41:189-191.
16. Bartlett SJ, Bykerk VP, Cooksey R, et al. Feasibility and domain validation of rheumatoid arthritis (RA) flare core domain set: report of the OMERACT 2014 RA Flare Group Plenary. *J Rheumatol*. 2015;42:2185-2189.
17. Loymans RJ, Ter Riet G, Sterk PJ. Definitions of asthma exacerbations. *Curr Opin Allergy Clin Immunol*. 2011;11:181-186.
18. Schmitt J, Spuls PI, Thomas KS, et al. The Harmonizing Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. *J Allergy Clin Immunol*. 2014;134:800-807.