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

Authors: Sebastien Barbarot PhD\textsuperscript{1,2}, Natasha K Rogers PhD\textsuperscript{1}, Katrina Abuabara MD\textsuperscript{3}, Helene Aubert MD\textsuperscript{2}, Joanne Chalmers PhD\textsuperscript{1}, Carsten Flohr PhD\textsuperscript{4}, Jon Hanifin\textsuperscript{5}, Luigi Naldi MD\textsuperscript{6}, David J Margolis MD\textsuperscript{7}, Carle Paul MD\textsuperscript{8}, Matthew J Ridd PhD\textsuperscript{9}, Marie-Louise Anna Schutteelaar PhD\textsuperscript{10}, Eric Simpson MD\textsuperscript{5}, Marie Tauber MD\textsuperscript{8}, Annika Volke MD\textsuperscript{11}, Stephan Weidinger MD\textsuperscript{12}, Sally R Wilkes PhD\textsuperscript{1}, Andreas Wollenberg PhD\textsuperscript{13}, Kim S Thomas PhD\textsuperscript{1}

1 Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, UK
2 Department of Dermatology, CHU de Nantes, Nantes, France.
3 Department of Dermatology, University of California San Francisco, USA
4 Population-Based Dermatology Research Unit, St John’s Institute of Dermatology, Guy’s & St Thomas’ NHS Foundation Trust and King’s College London, London, UK
5 Department of Dermatology, Oregon Health and Science University, Portland, Oregon, USA
6 Study Centre Italian Group for Epidemiologic Research in Dermatology (GISED) and Department of Dermatology, Azienda Ospedaliera papa Giovanni XXIII, Bergamo, Italy
7 Department of Dermatology, Department of Biostatistics and Epidemiology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, USA
8 Dermatology, Paul Sabatier University and Hopital Larrey, Toulouse, France
9 School of Social & Community Medicine, University of Bristol, Bristol, UK
10 Department of Dermatology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
11 Department of Dermatology, University of Tartu, Tartu, Estonia
12 Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein, Campus Kiel, Germany
13 Department of Dermatology and Allergy, Ludwig-Maximilian University Munich, Germany

*Corresponding author: Professor Kim Thomas, The University of Nottingham, King’s Meadow Campus, Lenton Lane Nottingham, NG7 2NR (kim.thomas@nottingham.ac.uk)

Reprint requests: See above

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

Conflicts of interest: None. This review has been conducted in support of the Harmonizing Outcome Measure for Eczema (HOME) initiative.

Word Count:
Abstract: 200
Capsule summary: 47
Text: 2498
Figure: 3
Tables: 2

IRB review – not required – secondary data synthesis

Key words: atopic eczema, atopic dermatitis, long-term control, outcome measures, RCTs, systematic review; flares.
• 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.
Background: Atopic dermatitis (AD) is a chronic inflammatory skin disease. There are no standardised methods for capturing long-term control of AD.

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

Methods: Systematic review of RCTs of AD treatments published between 2000 and 2013, with a follow-up period of ≥3 months, at least one outcome measure recorded at ≥3 time-points, full paper available, and published in English.

Results: 101/353 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 for trials including flare data (21/52).

Medication-use was recorded based on quantity, potency and frequency of application.

Limitations: 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.
<table>
<thead>
<tr>
<th></th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>AD, Atopic Dermatitis</td>
</tr>
<tr>
<td>17</td>
<td>ANOVA, Analysis of Variance</td>
</tr>
<tr>
<td>18</td>
<td>ANCOVA, Analysis of Covariance</td>
</tr>
<tr>
<td>19</td>
<td>BSA, Body Surface Area</td>
</tr>
<tr>
<td>20</td>
<td>EASI, Eczema Area and Severity Index)</td>
</tr>
<tr>
<td>21</td>
<td>IGA, Investigators Global Assessment</td>
</tr>
<tr>
<td>22</td>
<td>HOME, Harmonizing Outcome Measure for Eczema</td>
</tr>
<tr>
<td>23</td>
<td>POEM, Patient-Oriented Eczema Measure</td>
</tr>
<tr>
<td>24</td>
<td>RTC, Randomised Controlled Trial</td>
</tr>
</tbody>
</table>
INTRODUCTION

Atopic dermatitis (syn. 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 and minimise 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 trials².

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 Measure for Eczema (HOME) initiative (www.homeforeczema.org) identified long-term control as one of four key domains to measure in all clinical trials in AD. The current systematic review has been conducted in order 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 Roadmap⁵ namely to identify available outcome instruments for capturing the domain of interest.

METHODS

This systematic review was performed according to PRISMA recommendations⁶. The protocol was agreed prior to 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 period⁷ that included adults or children with AD, and which were published between January 1, 2000 and March 12, 2013. This period was chosen as prior to 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, CINHAL, AMED, LILACS, the Cochrane
Library and the Skin Group Specialised Register databases.

The search strategy used to identify RCTs in the GREAT database and validation of the GREAT database have
been published elsewhere. Observational studies were not included in this review due to 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,
and a clinician or patient-reported outcome measure was recorded at three or more time points. We excluded
studies that were published in abstract form only, those which did not include clinical outcomes (e.g. studies
only containing data pertaining to biomarkers or skin barrier function tests), and those that were not published
in English. Titles of studies were retrieved and the full-text was then obtained and screened against the
inclusion criteria by two authors (NR, SB). Responses were compared and discrepancies resolved by consensus
(NR, SB).

Studies which met the inclusion criteria were divided between author pairs, who independently extracted data
using a standardised data extraction form. Details were extracted for: (i) trial attributes (size of trial, age of
participants); (ii) repeated measurement of clinician or patient-reported AD outcomes over time (iii) use of AD
medication - defined as any treatment used to control AD symptoms other than the randomly allocated
intervention; (iv) AD flares / relapse - defined as a decline in condition (worsening of symptoms) which met
one of the recommended descriptions of flare, 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, analysed and presented in the paper were recorded. Data extraction forms were reviewed by
another two authors (NR, SW), who checked for completeness and resolved any discrepancies by referring to
the original trial publications.

Results were summarised 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 categorised 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 1 January 2000 and 12 March 2013 yielded a total of 353 RCTs (Fig1 and Appendix). Overall, 101 trials were included in the review (67% included either children or adults, 31% included both children and adults, one 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 (71.2%) trials measured long-term control in two 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, 68/101 (67.3%) had at least one graphical 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) (Figure 2). The most commonly used outcomes were: SCORAD or objective SCORAD (25%), quality of life scales (14%), pruritus scales (10%), Body Surface Area (BSA) (8%), EASI or modified EASI (8%) and Investigator Global Assessment (IGA) (7%). As previously shown, there was large variability in IGA definitions between studies\textsuperscript{10}. The breakdown of clinician-reported and patient-reported outcomes is summarised in Figure 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 four 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 (Figure 4).

AD flares

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

Most trials analysed flares in multiple ways, with a total 52 analyses performed (Table 1). 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 72/196 (37%) of the reported analyses made best use of the available data and included all time points in the analysis (Table 2). Analyses that were considered to be best use of the data included: ANOVA (n = 35 analyses), linear mixed model (n=13 analyses), ANCOVA (n=12 analyses), nonlinear mixed model (n=2 analyses), non-parametric repeated measures (n=2 analyses), area under the curve (n=1 analysis), log-rank test (n=1 analysis), McNemar (n=1 analysis), 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.

Since 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 (such as ANOVA, ANCOVA or mixed models). Using multiple tests to compare data between groups at each time point leads to increased risk of identifying a significant difference by chance. The fact that 39.7% of the 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), 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. 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.
Equally important is the concept that outcomes should be relevant to patients with all severities of disease and healthcare settings. Most AD patients 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 due to 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 flares defined by topical corticosteroid use or flares may be less useful.

The optimum frequency of outcome assessments (e.g. 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 on-line data collection tools, 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. An agreed 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 salutatory 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 summarise 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 (EASI)\textsuperscript{7,18} and that patient-reported symptoms should be captured using the Patient Oriented Eczema Measure (POEM). As such, in the absence of an agreed instrument for capturing long-term control, we recommend an ‘interim solution’ of using at least one 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.

Acknowledgements

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

Table 1: Summary of methods to analyse flare outcomes

Table 2: Summary of methods of analysis for repeated measures data

FIGURE LEGENDS

Figure 1: LTC Flow diagram

Figure 2: Distribution of the 196 outcomes used in 92 trials that reported repeated measurement of AD outcomes

Figure 3: Number of patient-reported and clinician-reported outcomes used in the included trials

Figure 4: Methods of collection for medication use

APPENDIX

List of included studies


<table>
<thead>
<tr>
<th>Analysis</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to first flare</td>
<td>21</td>
</tr>
<tr>
<td>Number of flares</td>
<td>17</td>
</tr>
<tr>
<td>Duration of remission</td>
<td>5</td>
</tr>
<tr>
<td>Duration of flare</td>
<td>4</td>
</tr>
<tr>
<td>“Totally controlled weeks” and “well controlled weeks”</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
</tr>
<tr>
<td>Appropriateness of analysis</td>
<td>Category</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>Best use of data</td>
<td>Took into account all time points in single analysis</td>
</tr>
<tr>
<td>Inefficient analysis</td>
<td>Only compared baseline and end point</td>
</tr>
<tr>
<td>Inefficient analysis</td>
<td>Only data at a single time point is assessed</td>
</tr>
<tr>
<td>Inappropriate analysis</td>
<td>Compared each time point to baseline individually</td>
</tr>
<tr>
<td>Inappropriate analysis</td>
<td>Compared groups at each individual time point</td>
</tr>
<tr>
<td>Not analysable</td>
<td>Unclear</td>
</tr>
</tbody>
</table>
Figure 1: LTC Flow diagram

*Multiple categories are possible

Trials published in GREAT database 2000-2013  
\( n = 353 \)

Excluded  
Duration <3 months  
\( n = 188 \)

Excluded  
<3 relevant data collection points  
\( n = 14 \)  
unclear duration/collection points  
\( n = 38 \)

Excluded  
Abstract only  
\( n = 7 \)  
Non-English  
\( n = 5 \)

Included RCTs  
\( n = 101 \)

- RCTs with repeated measurement of AE outcomes*  
  \( n = 92 \)
- RCTs with flare outcomes*  
  \( n = 26 \)
- RCTs with medication outcomes*  
  \( n = 29 \)
Figure 2: Number of patient-reported and clinician-reported outcomes used in the included trials
Figure 3: Methods of collection for medication use

- Frequency: 7
- Potency: 7
- Weight: 8
- Unclear: 2
APPENDIX

Publications meeting inclusion criteria


101. Trookman NS, Rizer RL. Randomized Controlled Trial of Desonide Hydrogel 0.05% versus Desonide Ointment 0.05% in the Treatment of Mild-to-moderate Atopic Dermatitis. J Clin Aesthet Dermatol. 2011;4(11):34-8.


