Report from the seventh international consensus meeting of the harmonising outcome measures for eczema (HOME) initiative

Funding

Grants were obtained by the local hosts from the following organisations to support the meeting: Japanese Dermatological Association, Kyoto Prefectural University of Medicine and Kyushu University, The Uehara Memorial Foundation, The Nakatomi Foundation and Lydia O'Leary Memorial Pias Dermatological Foundation. All attendees other than patient representatives paid a meeting fee. Travel and accommodation costs for Maarten Boers, independent facilitator were covered by Hywel Williams, using funds from his National Institute for Health Research Senior Investigator award. Laura Howells was supported by a PhD studentship from the British Skin Foundation (reference 8016).

Conflicts of interest

Kim Thomas, Joanne Chalmers, Christian Apfelbacher, Laura Howells, Phyllis Spuls, Lynita Howie and Tim Burton were involved in development of RECAP. Abhijit Gadkari, Laurent Eckert and Eric Simpson were involved in development of ADCT.

Acknowledgments

We would like to thank all attendees who contributed to the HOME VII meeting, with a particular thank you to our patient contributors: Jennifer Austin, Julie Block, Alix Bullock, Tim Burton, Lynita Howie, Rosemary Humphreys, Yuko Ikegami, Henrique Ishii, Michael Lanigan, Kyoko Maru, Eri Maruyama, Miwako Ogino, Vanessa Sultana and Akane Yasui.

Thank you to the local organising committee: Norito Katoh, Masutaka Furue, Hidehisa Saeki and Takeshi Nakahara and Convention Academia, to Masaki Futumura for supporting our patient partners during the meeting and assisting with running the patient session. Yukiyasu Arakawa and Susumu Ichiyama translated for Japanese patient representatives

Assistance with pre-meeting preparations and reviews, information specialist support and manuscript preparation: Natasha Rogers, Douglas Grindlay, Rachel Pattinson and Barbara Maston.

Key words

Consensus meeting, core outcome set, quality of life, long-term control, itch, atopic eczema, atopic dermatitis

Introduction and aims

This report details the seventh meeting of the Harmonising Outcome Measures for Eczema (HOME) initiative, HOME VII. This face-to-face consensus meeting was held in Tokyo, Japan from 8th to 10th April 2019. The meeting was hosted by Professor Norito Katoh and colleagues. The meeting was chaired by Hywel Williams, chair of the HOME initiative, and independently moderated by Maarten Boers of OMERACT (http://www.omeract.org).

HOME has an open, free-of-charge membership policy. Approximately 400 members represent all key stakeholders from around the world. The purpose of HOME is to develop a core outcome set for atopic eczema (syn. atopic dermatitis, eczema) clinical trials and a clinical practice outcome set.

With regards to clinical trials, the four core outcome domains of clinical signs, patient-reported symptoms, quality of life and long term control were agreed at the first HOME consensus meeting in Amsterdam, 2011[1]. Subsequent biennial consensus meetings have taken place to recommend core outcome instruments to measure these domains. Consensus to recommend EASI for measuring clinician-reported signs and POEM for patient-reported symptoms was achieved at the HOME III[2] and HOME IV[3] meetings respectively, and the domain of long-term control was further defined at HOME V[4]. No agreement was reached at previous meetings on recommended instrument(s) for quality of life[4]. Summaries and outputs from previous HOME meetings are available on the HOME website (www.homeforeczema.org).

The aims of this HOME VII meeting were i) to recommend instruments for the remaining domains of the core outcome set for eczema clinical trials (quality of life and long-term control), ii) to determine whether itch intensity should be added as a distinct subdomain of symptoms and if yes, recommend an instrument, and iii) to establish when and how often core outcomes should be measured.

Methods

Recruitment of participants

All HOME members were invited to participate and the meeting was advertised via relevant mailing lists. Additional invitations were sent to potential attendees local to the meeting. Registration was via the HOME website and the charge was 100 Euros (standard rate) and 500 Euros (industry rate). Patient representatives attended free-of-charge. All participants covered their own travel and accommodation costs with the exception of patients whose costs were covered in part by bursaries provided by the local hosts with additional contributions from members of the HOME executive group.

Pre-meetings

Two introductory pre-meeting sessions were held to ensure all participants were sufficiently equipped to fully contribute to the main meeting. The first was held on the 7th April 2019 and was only open to patients, carers and patient representatives. The second, held on the

morning of the 8th April 2019, was entitled “Unpacking the mysteries of HOME” and was open to all participants. However, it was particularly aimed at non-patient stakeholders who had not previously been involved in HOME or who required a refresher session.

Both meetings covered what HOME is and why HOME is needed, the role of different stakeholders, what has been achieved already in HOME, methods to be used in the current meeting and the aim of the next 3 days. The sessions also provided an opportunity for participants to get to know each another ahead of the meeting.

Evidence and background information

Two weeks prior to the meeting, all participants were sent an introductory video, a summary of the available validation data for candidate instruments, and copies of the candidate instruments. Participants also had copies of these materials to refer to during the meeting. As a pre-meeting task, participants were asked to consider the feasibility and face-validity of candidate instruments for the Quality of Life and Long-term Control domains and personal preferences of the available instruments.

Meeting methods

As with previous HOME evidence-based consensus meetings, the process for arriving at consensus on recommending instruments was guided by the HOME roadmap and COSMIN guidance[5 6]. Evidence from recent systematic reviews of validation studies of the measurement properties for candidate instruments was presented to all participants with subsequent whole group discussions and clarifications. For each session, a modified nominal group technique was then used, in which participants split into small groups for discussion and (non-binding) voting within each group. Participants were pre-assigned to one of six small groups of 12 or 13 participants to ensure a geographical spread and stakeholder representation in each group, and these allocations remained the same for the whole meeting (see supplementary material). Each small group had a facilitator and a note taker. The decisions and reasons were fed back to the whole group, followed by further whole group discussions and voting.

Voting and consensus rule

Voting was conducted using anonymous handsets and the results were presented to the group in real time (with some exceptions, as detailed in the results). Each participant had one vote, but did not vote where there was potential conflict of interest such as having been involved in developing an instrument, and all votes had equal weighting. Observers and translators did not vote. The HOME consensus rule remained unchanged and was applied at this meeting; consensus is reached when less than 30% of voters disagree with the statement[1].

Results

Participants

A total of seventy-five people from 16 countries representing Asia, North and South America, Australasia and Europe attended the HOME VII meeting as active participants, plus 5 translators/observers (see supplementary material). All stakeholder groups were well represented (see Figure 1 for breakdown per stakeholder group) and a maximum of two employees of any pharma company were permitted to attend.

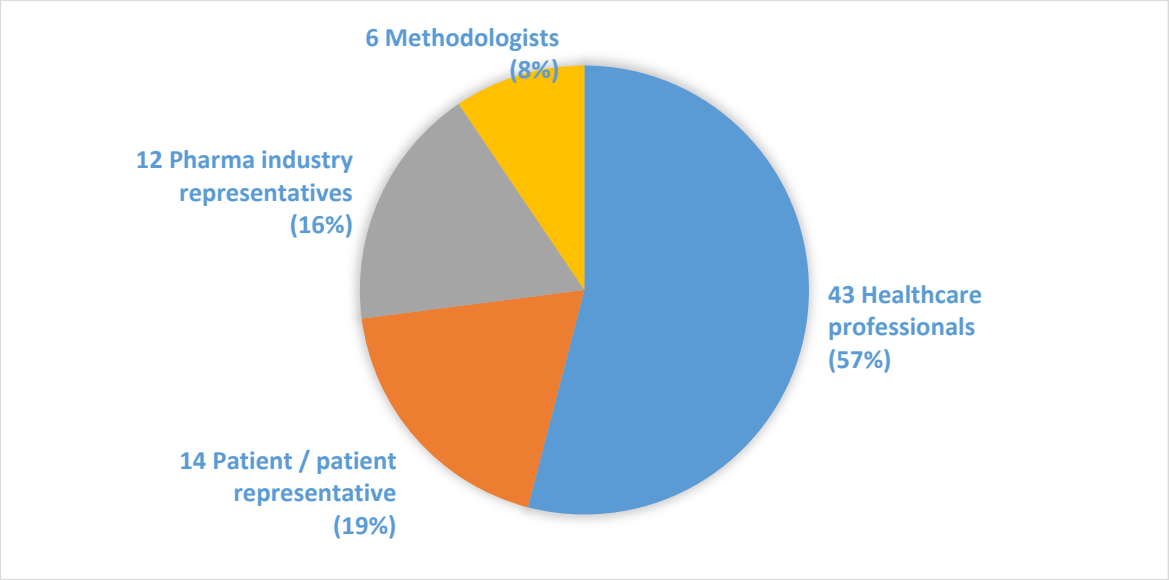

Figure 1– attendees by stakeholder group

Day 1

Quality of Life domain

Introduction and background

Christian Apfelbacher, lead for the QoL working group, opened the session by reminding the group that QoL is a complex and often multidimensional construct, usually measured by multi-item questionnaires that have different versions for adults, adolescents, children and infants. HOME had previously agreed that Qol included three key subdomains; physical, social and psychological. Although QoL in adults, adolescents, children and infants had been discussed at previous HOME meetings, it had not been possible to achieve consensus on a preferred instrument[3 4]. This meeting would use the pragmatic approach of the COSMIN guidance on instrument selection in which content validity is prioritised above other measurement properties, and allows instruments to be considered based on a minimum set of criteria pending missing validation data[6]. The aim of this session was to achieve consensus on which candidate skin-specific instruments for measuring quality of life (QoL) in adults, adolescents and children should be recommended for inclusion in the COS for eczema trials.

At previously HOME meetings, the Dermatology Life Quality Index (DLQI), the Children's Dermatology Life Quality Index (CDLQI) and the Infant's Dermatitis Quality of Life Index (IDQoL) had been the preferred instruments for the COS based on face validity and feasibility. However, concerns about content and structural validity meant that no consensus had been reached, even for a preliminary recommendation of these instruments[3 4]. There was, however, recognition that the DLQI, CDLQI and IDQoL are very widely used in clinical trials, routine practice, and reimbursement. A recent qualitative study on the difference in content between DLQI and Skindex[7] and a study using exploratory factor analysis to develop a short-form CADIS[8] were also presented.

The content validity of one (or two related) candidate instruments using the COSMIN criteria and rating system (see supplemental material) was conducted by each break-out group.[9] As had been noted, content validity was previously lacking and this was an opportunity to add to the data available. A summary of each small group discussion was presented to the whole group (summarised in Table 1) and further whole group discussion took place, summarised in Table 1.

Group	Instrument(s)	Group feedback
Children / infants instruments		
1	CADIS & Short-form CADIS	• Comprehensiveness; CADIS was considered comprehensive but too long (45 questions). Short form feasible but sleep missing. • Some redundancy in items in full CADIS • Geared towards parent perception of disease rather than the child's quality of life so not considered to reflect the construct of interest and degree of parental knowledge of disease may affect responses. • Recall period; 1 month considered problematic and may compromise accuracy.
2	InToDermQoL	• Poorly presented; no subheadings. • Difficult to understand and complete. • Response options had overlap. • Recall period; 1 week considered appropriate • Appropriateness of content differed depending on age of child being considered • Only appropriate for children aged 0-4 and the use of different instruments for different age groups could be an issue in clinical trials
4	CDLQI IDQoL	• Both instruments generally liked overall • The negative response options and recall period were only issues • Comprehensiveness; Some aspects specifically relevant to teenagers may be missing • Important to consider impact on the family when assessing QoL in children • Unclear if young children (4 to 6 year olds) are able answer some of the questions themselves such as how much has your sleep been disturbed by eczema
5	DISABKIDS	• Comprehensiveness; missing some key elements; effect of treatment on life (e.g. time spent on treatment), family impact and focussed on frequency rather than intensity. • Some overlap in questions on appearance. • Parents may not always know the impact of the eczema on a child's sleep • Wording considered to be generally negative and some considered difficult for most 8 year olds. • Language too direct for some cultures e.g. Japanese, and "bother" may be a problematic to translate in some languages.
Adult instruments		
3	ABS-A	• Generally liked the instrument overall • Comprehensiveness; lacked psychosocial, mood, anxiety and embarrassment • Significant number of questions with 'not applicable' as an option could make interpretation in trials difficult with high levels of missing data • Likely to be susceptible to floor effects
6	DLQI & Skindex	• Comprehensiveness; Skindex considered more comprehensive than DLQI as it includes missing sleep and emotions such as anxiety and mood (more focus on activity than emotion) • Comprehensiveness; Skindex missing treatment burden but not considered as important as sleep and emotions

Table 1: Summary of assessment of content validity by small breakout groups

General points • It is important to view the COS as a whole. Key elements missing from one domain may be captured adequately within others such as sleep missing from Skindex but included in the symptoms core instrument, POEM. Redundancy where items are repeated across domains is less of a problem than lack of comprehensiveness. • QoL instruments ideally only capture the effect of symptoms on quality of life. However, some instruments also include the presence or absence of symptoms themselves. HOME has not stated that the core outcome instrument for QoL should be an impact only instrument. • Given that one of the main goals of a COS is to enable the synthesis of data, the inclusion of "legacy" instruments that are widely used in trials should be considered.
Instruments for adults • Despite some lack of comprehensiveness in some areas, the candidate instruments all measure many useful aspects. • Of the 3 key subdomains of QoL (physical, social and psychological), DLQI was considered to not adequately capture the psychological element. This is not measured elsewhere in the COS.
Instruments for children • CADIS includes parent QoL which, although important, is not part of the QoL construct as defined by HOME which is the QoL of the person with eczema (i.e. the child). It is possible to report the family and the child domain scores separately but it is unlikely that trials would routinely report in this way. • The short form CADIS was developed to reduce the item pool to produce a more feasible candidate instrument. The process of item reduction had inevitably led to a loss of some concepts, and general consensus was that removal of the loss of sleep subdomain was not appropriate.

Table 2: Summary of whole group discussions following feedback from small group content validity assessments

The updated systematic review of the measurement properties of candidate QoL instruments was presented by Michaela Gabes (Table 3)[10]. This update included instruments from the previous reviews[11 12] only if they had been agreed as feasible for the COS at previous HOME meetings[4 13]. These were IDQoL, CDLQI, Disabkids, and Childhood Atopic Dermatitis Impact Scale (CADIS) for children, and DLQI and Skindex for adults. All new instruments with any validation studies were also included as these had not been previously assessed for feasibility; short-form (SF) CADIS and the Infants and Toddlers Dermatology Quality of Life (IntoDermQoL) for children, and Atopic Dermatitis Burden Scale for Adults (ABS-A) and ADerm-IS for adults. The review was conducted using the revised COSMIN risk of bias checklist and rating categories (A-C) to assess the methodological quality of measurement properties.[14] The quality of evidence per measurement property was graded as high, moderate, low, or very low and downgraded for risk of bias, inconsistency, imprecision, and/or indirectness. A summary of the recommendations based on the updated review was presented (Table 3).

	Category A		**Category C**	
Instrument	Evidence for sufficient content validity (any level)	At least low quality evidence for sufficient internal consistency	High quality evidence for an insufficient measurement property	Recommendation
IDQoL	✓	X	X	B

CADIS	✓	X	X	B
DISABKIDS	✓	X	X	B
CDLQI	✓	X	X	B
InToDermQoL	✓	X	X	B
DLQI	✓	X	✓	C
Skindex	X	✓	X	B
ABS-A	✓	X	X	B
ADerm-IS	✓	X	X	B

Table 3: Degrees of recommendation of candidate instruments

In the whole group discussion that followed, general opinion was that a recommendation for QoL instrument(s) needed to be made very soon because several years have elapsed since the COS domains were agreed. However, despite an additional ten validation studies included in the updated review, it was clear that no instrument could be easily recommended over the others. Furthermore, it was felt that due to its established widespread use in trials, not including DLQI (and CDLQI/IDQoL) in the COS would decrease the ability to synthesise QoL data with a very high number of previously published trials. It was acknowledged that these factors should be taken into account in the decision making process regarding DLQI. The need to choose an instrument led to discussions about potential new QoL instruments that are likely to be available in the near future. Some instruments currently in development are using up to date methodology with good patient input, but there was concern that they may take longer than planned to develop or may not perform as well as anticipated. Therefore, the general agreement was that the group should move towards with a decision considering the currently available instruments. However, there was a firm commitment to review in future any promising new instruments that emerge.

It was noted that DLQI is rated as category C because of high quality evidence for insufficient internal consistency. However, there are no studies on other instruments and it was pointed out that it is entirely possible that other instruments may perform equally poorly in this respect. There was a discussion around the relative importance of internal consistency for these instruments and that confirmatory factor analysis (CFA) is required in order to fully assess internal consistency. Although responsiveness to change is not included in this COSMIN initial recommendation table, it is important for trials and should be considered with regards to the COS.

It was also discussed that although it is important to consider the requirements and preferences of regulators such as the FDA, the COS should focus primarily on what is important for trials overall, rather than requirements for license and labelling claims.

Day 2

After recapping the results of the content validity study carried out by the small groups on day one, and the data from the review on content validity, Christian Apfelbacher reiterated that no strong recommendation can be made for any currently available skin-specific QoL instruments at present. Joanne Chalmers then presented data from a recent study assessing the uptake of the COS registered in a clinical trials database,[15] highlighting only 30% of trials measured of QoL compared to approximately 80% for signs and symptoms. This strongly supported the need to recommend an instrument(s) for the QoL domain. Of the trials that did measure QoL, 63% used DLQI, CDLQI or IDQoL and only 4% used another instrument (the remaining trials were unclear).

Subsequent discussions were largely around comparisons between the stronger contenders for the adult instruments (DLQI and Skindex) and the potential for new and better instruments and summarised in Table 4.

Validity • Although the review found no published evidence for good content validity for Skindex, the results of the breakout group exercise indicated that content validity was good • The psychosocial domain would be lost if DLQI is recommended, and patients expressed the opinion that this is an important element of eczema-related QoL. • Responsiveness is important in trials so it was felt this should be considered.
Existing use of instruments • DLQI is clearly the most commonly used QoL instrument, whilst Skindex has been rarely used in trials • Many current industry phase 2 and 3 programmes are using DLQI. • Any recommendation for an alternative instrument to DLQI would eventually be taken up by trialists. • A recommendation for DLQI would maximise the ability to synthesise data from new and existing trials.
Feasibility • DLQI was considered easier to complete than Skindex by some patients. • Since the COS has a global scope, cross-cultural validation should be considered and differential Item functioning (DIF) has been investigated for DLQI across many languages, but not for Skindex. • Neither instrument can take the place of generic QoL measures for cost-effectiveness analyses.
Age groups • Coverage of all age groups with the DLQI, CDLQI and IDQoL was considered convenient for conducting clinical trials • Although they include similar domains, these instruments are not different versions of the same instrument so data cannot automatically be combined across instruments/age groups
New instruments on the horizon • Instruments are being developed using modern methods are likely to be available in the next 5-10 years. • There was concern that asking trialists to switch now from using DLQI to using another less than perfect instrument, followed by a second switch to a new instrument in a few years' time would not be acceptable or beneficial in terms of data synthesis.

• Recommending existing instrument(s) may reduce the impetus to develop new instruments.

Table 4: Summary of discussion

The voting then commenced. The group voted overwhelmingly to progress to voting on an instrument for each age groups (Questions 1-3 in Table 5). All potential instruments were then subject to an individual vote regarding suitability for recommendation as a core outcome instrument.

Consensus was reached that IDQoL (infants and very young children), CDLQI (children), and Skindex and DLQI (both for adults) were suitable for recommendation for the COS (Questions 4, 9, 11 and 12 in Table 5). InToDerm, short-form CADIS and long-form CADIS (infants and very young children) Disabkids (children), and ABS-A (adults) received *insufficient* votes to be suitable for recommendation for the COS. (Questions 5-8 and 10 in Table 5Table 5).

Because only one instrument for infants/very young children and children (IDQoL and CDLQI respectively) were voted as being acceptable for the COS, no further voting was required and these instruments would be recommended for the COS. However, for adults both DLQI and Skindex were considered acceptable by consensus vote, and therefore, a vote to choose between these instruments was required in order to adhere to the HOME roadmap of selecting one core instrument per domain[5]. In this final consensus vote DLQI was recommended as a core outcome instrument for QoL in adults, but Skindex was not (Questions 13 and 14, Table 5).

Question	Voting question	Yes n (%)	No n (%)	Unsure n (%)
1	It is clear there is not sufficient evidence to clearly recommend one instrument for measuring quality of life in infants / very young children. Should we still move to a vote on candidate instruments for measuring quality of life in infants / very young children?	87	13	-
2	It is clear there is not sufficient evidence to clearly recommend one instrument for measuring quality of life in children. Should we still move to a vote on candidate instruments for measuring quality of life in children?	86	14	-
3	It is clear there is not sufficient evidence to clearly recommend one instrument for measuring quality of life in adults. Should we still move to a vote on candidate instruments for measuring quality of life in adults?	83	17	-
Instruments for infants / very young children				
4	Do you agree that IDQoL can be recommended as a core outcome instrument for measuring quality of life in infants / very young children?	58 (83)	8 (11)	4 (6)
5	Do you agree that InToDermQoL can be recommended as a core outcome instrument for measuring quality of life in infants / very young children?	16	68	16
6	Do you agree that Short form CADIS can be recommended as a core outcome instrument for measuring quality of life in infants / very young children?	32	50	18
7	Do you agree that Long form CADIS can be recommended as a core outcome instrument for measuring quality of life in infants / very young children?	9	78	13
Instruments for children				

8	Do you agree that DISABKIDS can be recommended as a core outcome instrument for measuring quality of life in children?	11	76	13
9	Do you agree that CDLQI can be recommended as a core outcome instrument for measuring quality of life in children?	60 (87)	6 (9)	3 (4)
Instruments for adults (first vote)				
10	Do you agree that ABS-A can be recommended as **a** core outcome instrument for measuring quality of life in adults?	9	78	13
11	Do you agree that Skindex can be recommended as **a** core outcome instrument for measuring quality of life in adults?	46 (66)	16 (23)	8 (11)
12	Do you agree that DLQI can be recommended as **a** core outcome instrument for measuring quality of life in adults?	60 (83)	9 (13)	3 (4)
Instruments for adults (second vote – agree to recommend)				
13	Do you agree that DLQI can be recommended as **the** core outcome instrument for measuring quality of life in adults?	58 (81)	11 (15)	3 (4)
14	Do you agree that Skindex can be recommended as **the** core outcome instrument for measuring quality of life in adults?	24 (34)	42 (59)	5 (7)

Table 5: Results of voting on candidate QoL instruments.

In summary, the following instruments were recommended for inclusion in the core outcome set for measuring QoL: infants & very young children - IDQoL, children - CDLQI, and adults – DLQI.

When and how often should the core outcomes be measured?

Eric Simpson presented background to why in addition to harmonising outcome domains and instruments it is crucial to also address the timing and frequency of outcome measurement as variation prevents data synthesis between trials. HOME has previously recommended presenting the mean and SD or median and quartile ranges for EASI and POEM data, at baseline and end of treatment for each group.[16] The aim of this session was to determine whether further recommendations should be made regarding at what time points trial outcomes should be measured.

A pre-meeting survey of HOME members was conducted and the results were presented; 106 responses were received (50 clinicians, 33 patients/patient representatives, 14 Pharma industry, 7 methodologists, 2 not stated), representing 23 countries (see supplementary material).

The survey showed there was little consensus on appropriate time points for measuring initial and medium response to interventions (see supplementary material). Differences in the mode of action of different interventions, the expected time to onset of response and different disease severity were common reasons given as to why a specific time point may not be appropriate or feasible across all trials.

There followed a whole group discussion on this topic, summarized in Table 6.

Mandatory versus guidance
• There was overall agreement that it is *not* appropriate or feasible for HOME to *mandate* time points but should instead provide *guidance* to be followed where possible.

• The role of HOME is not to dictate trial design but rather to optimize the ability to synthesise data across trials
Multiple factors influence the choice of time points • Many factors including the type/phase of trial, the intervention e.g. mode of action, severity of disease, regulatory requirements, and payer requirements all influence the trial design and required time points for assessing outcomes. • Domains differ in when they are likely to respond e.g. itch may respond quickly to an intervention, whereas sleep may take longer, and QoL longer again.
Appropriate short and medium term time points • Short-term: 1 -2 weeks may be more relevant than 4 weeks for patients so they have an idea of whether or not they should be expecting a rapid response. • Medium-term: 12-16 weeks generally considered to be reasonable. • Decisions on time points should be driven by patients.
All outcomes or only PROs? • Although ideally, recommended time points would apply to both patient and clinician reported outcomes, it was acknowledged that the cost and burden of study clinic visits needed for clinician reported outcomes may restrict the ability of trials to adhere to guidance for clinician reported outcomes. • HOME could consider recommending time points for PROs only as they can be collected more easily and without study clinic visits. However, HOME should recommend that when study visits are taking place, clinician reported outcomes should be reported. • More guidance on timing of clinic visits would support researchers in designing trials, rather than each group having to work through the same set of decisions.
Reporting of time point data • Restrictions in journal word counts may affect the ability to include all relevant data. • Confirmed that the data simply need to be available in the public domain, and not necessarily in the results publication e.g. trial registry or online supplementary material. Pharma will not post data to open access websites.
Synthesising data • Work is needed to determine which time points are sufficiently comparable to combine when synthesising data. For instance, 12 and 16 weeks may be combined but perhaps 12 and 20 weeks are too dissimilar. • This question can be drug-dependent e.g. for some eczema drugs, the response at 12 and 16 weeks can be very different. • The reality is that systematic reviewers are constantly making decisions on combining data from different time points anyway, so it would be helpful to have a recommendation.

Table 6: Summary of discussion around recommending time points for the COS

The group then moved on to voting. A number of points were clarified ahead of votes being cast:

- Q1: the primary endpoint is usually the end of treatment but where these differ, measurement should be at the primary endpoint.
- Q2: voting relates to recommendations as a principle rather than specifically short and/or medium term time points.

- Q2: voting refers to timepoints *and* windows although shortened to time points in the voting wording for brevity.

Voting results:

Question	Voting question	Yes n (%)	No n (%)	Unsure n (%)
1	As a minimum, the core outcome set should be reported at baseline and at the time of the primary endpoint	72 (97)	0 (0)	2 (3)
2	Should HOME make recommendations (not mandatory) about additional time points for reporting the core outcome instruments?	60 (83)	9 (13)	3 (4)

In summary, the group agreed that the COS should be reported at baseline and the primary endpoint (which is usually end of treatment). Further work is required to determine; i) appropriate time points, ii) what time points are important to patients iii) whether it is possible to use time windows rather than specific time points, iv) what duration of time window is appropriate and v) what time points can be combined in systematic reviews. Further decisions on appropriate time points / windows will be made once there is more evidence on which to base a decision.

Long term control

Kim Thomas, lead for the long-term control domain working group, introduced the domain and the concept of long-term control being eczema control measured repeatedly over time throughout a trial. The group were reminded of the previous consensus that long-term control included signs, symptoms, QoL, all of which are already captured in the core set, and a global measure of eczema control[4]. There was also a reminder that although feasible for some trials, it had previously been agreed in HOME that flares, well-controlled weeks or treatment use are not feasible options for the core set[4]. The conceptual model of eczema control that had been developed by the working group was presented.[17] The aim of this session was to build on this previous agreement and work towards consensus on which instrument(s) should be recommended to measure eczema control in the core set bearing in mind feasibility, validity and interpretability.

All conflicts of interest were declared as any involvement in the development or validation of the candidate instruments and it was agreed they would not vote.

The development and validation of individual candidate instruments were presented (Table 7). Eric Simpson then summarised the measurement properties that have been assessed for each of the five candidate items and the quality of the evidence for each measurement property (Table 8). Eczema control could be measured by a single or a multi item instrument. With regards to multi-item instruments, ADCT and RECAP had both been developed according to FDA and COSMIN guidance and are well validated with moderate or high quality evidence for several measurement properties including content validity. Much

less evidence was available for PBI and ADS7 and where data were available, these performed less well. Although many different patient global assessment (PGA) instruments are available, validation data was available for only one.[18] There was high quality evidence for sufficient content validity and other key measurement properties for this PGA.

	Instrument	**Presented by :**
Multi-item instruments	Atopic Dermatitis Control Tool (ADCT)[19]	Laurent Eckert and Abhijit Gadkari
	Recap for atopic eczema (RECAP)[20]	Laura Howells
	The Patient Benefit Index 2.0 (PBI)[21-23]	Jochen Schmitt
	Atopic Dermatitis Score 7 (ADS7)[18 24 25]	Jochen Schmitt
Single-item instruments	Patient Global Assessment (PtGA) [18 25]	Eric Simpson

Table 7: Candidate instruments for measuring eczema control

		RECAP	ADCT	PBI	ADS7
Content validity	Overall rating	Sufficient (+)	Sufficient (+)	Inconsistent (±)	Insufficient (-)
	Quality of evidence	Moderate	Moderate	Moderate	Very low
Reliability	Overall rating	Sufficient (+)	TBC		
	Quality of evidence	Moderate			
Hypotheses Testing	Overall rating	Sufficient (+)	Sufficient (+)	Indeterminant (?)	
	Quality of evidence	High	High	High	
Responsiveness	Overall rating	Sufficient (+)	TBC	Indeterminant (?)	
	Quality of evidence	High		High	
Category of recommendation		A	A	B	

Table 8: measurement properties of candidate instruments

The subsequent discussions are summarised in Table 9.

Identification of candidate instruments The robust process for identifying candidate instruments was clarified: • The literature was systematically reviewed for HOME V and updated for this meeting, but this did not identify any multi-item eczema control scales. • Two instruments had been developed to fill this gap (ADCT and RECAP) but these had not yet been published so were not picked up by the review. • A survey of the HOME membership was also conducted which highlighted the PBI and ADS7 which could be considered candidate instruments, although they hadn't met the definition for an eczema control instrument in the review.
Concept of the domain of LTC • Because the domain of LTC was previously defined as including signs, symptoms, QoL and a "global measure", there was debate around whether this domain should only focus on the "global measure" or whether it should aim to capture all four elements of the domain. By including all four, redundancy with the rest of the COS will be present. • Although severity is part of control, it was felt that a global severity instrument is unlikely to truly capture eczema control.
V*alidation* • Clarification that ADCT and Recap are moderate for content validity because the content validity study was conducted in the same population as the development of the instruments. • It was agreed that HOME would not include all the reverse validation for the PGAs • Proxy version available for Recap and may be better understood for children • PGA may be better cross-culturally • ADCT and RECAP both have some overlap with POEM

Table 9: *Summary of discussions following presentation of the measurement properties of candidate instruments for measuring long-term control*

After near-neighbour discussions, delegates moved into their small groups to decide whether a single- or multi-item instrument is preferable and to determine their preferred instrument(s) based on face validity, feasibility and validity.

Group	Preferred single-item or multi-item global measure	Reasons why preferred single or multi-item instrument	Which single-item global measure was preferred?	Which multi-item global measure was preferred?	Reasons for preference choice
1	No consensus (50:50 split)		Single item to be developed	RECAP	PBI was excluded immediately Very little difference between RECAP and ADCT (4 voted for each) but more aspects of RECAP preferred; general linguistic reasons, didn't like bother question in ADCT, felt "acceptable" captured control well, liked redundancy of itch question (regular itch and flare)
2	-	Didn't discuss single item instruments as there was not a suitable validated instrument available.	Single item to be developed	RECAP	PBI good instrument but not for purpose of control RECAP and ADCT very similar – could consider this as a replication study 2 itch questions are an advantage on the RECAP

					Children Q6 feelings would be better understood than mood. Bother may not be easily translated.
3	Single item	Provided further research showed sufficiently good measurement properties	Single item to be developed	RECAP	Didn't like bother question in ADCT
4	Single item		Single item to be developed	RECAP	Conceptual framework covers everything important for control. But single-item better instrument because less prescriptive. Multi-item instrument adds to burden in trials and questionnaire fatigue. Patients wonder why they being asked the same questions again, can feel lit a"lie detector".
5	Single item		Single item to be developed	Didn't vote because preferred single item instruments.	Noted that it is possible to be in control but still have disease activity. Control should be a summary of the other domains but not overlap and a single-item instrument would fill this gap. Considered ADCT and RECAP to be very similar instruments.
6	Multi item	Single item lets patients decide for themselves what is important to them.	Considered one existing PGA to be acceptable but preferred a single item to be developed	RECAP	Discarded ADS7 quickly Liked concept of PBI but not designed for control RECAP has strong validation and patient involvement and most closely reflects model, but validation has not yet been peer reviewed whereas ADCT validation has. Mixed views about having 2 itch questions in RECAP. Like the inclusion of intensity and frequency questions. Acceptable question not considered good by all. Treatment is missing Bother is an issue ADCT has no proxy for children yet

Table 10: results of small group discussions and voting. RECAP - Recap of eczema control instrument; ADCT - atopic dermatitis control test instrument; PBI - Patient Benefit Index; ADS7 - Atopic Dermatitis Score 7

There was a strong push from the small group feedback that a single-item instrument may be a good option for this domain, but that a suitable instrument with appropriate validation

was not currently available. Most small groups favoured either single-item (to be developed) or RECAP, but there was a strong and widely held view from all groups that both ADCT and RECAP were good instruments and that choosing between them felt like "splitting hairs", and only chose because they were asked to do so as part of the small group task.

A whole group discussion then followed and a summary of the key points is in Table 11.

Treatment; why not adequately covered in RECAP or ADCT) given it's importance to patients? • Feasibility is an issue; the treatment regimen is often fixed in a clinical trial and during RECAP development it was not possible to develop a question that worked well. • Treatment is on the pathway to getting control, rather than a measure of control and a treatment goal may change over time.
Patient burden increased with a multi-item instrument (ADCT or RECAP) • Inclusion of a single-item instrument would minmise patient burden and overlap with other domains. • Some of the reluctance towards a multi-item instruments may be influenced by the earlier inclusion of QoL instruments and therefore caution of increased burden of including another multi-item instrument. • The core set should be viewed overall patient burden and repetitiveness should be considered, but currently the burden of the HOME COS is acceptable.
Advantages and disadvantages of a single item instrument • Control is multi-dimensional and varies between individuals and a single-item instrument allows for this individuality • Patients' individual priorities are unlikely to change over the duration of the average clinical trial. • These apparently simple questions may be more difficult to answer then they appear.

Table 11: Summary of discussion

The situation ahead of voting was summarised. Although there was significant support for a single-item instrument, an acceptable instrument has not yet been validated so is currently unavailable as an option for recommendation at present. However, two good multi-item instruments are available for measuring eczema control, ADCT and RECAP and therefore offer an option for recommendation. As a way forward, it was suggested that one or more multi-item instruments (ADCT and/or RECAP) could be recommended and the group could work on a single-item scale in readiness for assessment at a subsequent HOME meeting. One option could be to validate the global control questions contained within the multi-item instruments.

The HOME roadmap advocates the recommendation of just one instrument per domain/subdomain. However, in this case, both ADCT and RECAP are very similar and perform equally well, but are both lacking in evidence from real-life clinical trial settings. Choosing between the two instruments was considered to be premature and largely arbitrary. The group were urged to consider whether both could be recommended as a temporary measure in the first instance. Mapping may support the longer term inclusion of both instruments, or further validation work may demonstrate that one performs better than the other.

A series of indicative votes then took place to guide subsequent discussions but were not binding (Table 12). Results showed mixed opinion regarding preference for a single or multi item instrument (Q1). ADCT, RECAP and a new, as yet undeveloped, single item instrument

were all considered to be potential instruments for inclusion, but there was no support for inclusion of PBI or the current validated single-item instrument (Q2-6).

	Voting question (indicative ONLY votes)	Multi-item	Single-item	Unsure
		n (%)	n (%)	n (%)
Q1*	Do you prefer a multi-item instrument or a single-item instrument for capturing eczema control?	20 (37%)	32 (59%)	2 (4%)
		Yes	No	Unsure
Q2*	Do you agree that ADCT can be recommended as a core instrument to measure long-term control?	26 (51%)	19 (37%)	6 (12%)
Q3*	Do you agree that Recap can be recommended as a core instrument to measure long-term control?	31 (59%)	17 (32%)	5 (9%)
Q4*	Do you agree that PBI can be recommended as a core instrument to measure long-term control?	0 (26%)	49 (70%)	4 (4%)
Q5*	Do you agree that the validated single-item global can be recommended as a core instrument to measure long-term control?	14 (65%)	38 (29%)	2 (6%)
Q6*	Do you agree that a non-validated single-item global should be progressed as the core outcome instrument - but requires more research?	34 (65%)	15 (%)	3 (6%)

* *All conflicted participants did not vote*

Table 12: Results of indicative votes. Results of Q2-6 were not presented until voting on Q2-6 was complete

DAY 3

Long-term control (continued)

The long-term control session continued with a summary of the previous day's discussions, highlighting that the indicative voting shows the group has not yet reached consensus. There followed whole group discussion about different options of how to progress, detailed in Table 13Table 12.

OPTION 1	Single-item instrument preferred	Recommend a single-item global (new) but not able to specify an instrument. Research on preferred single-item to continue. In the meantime, measure eczema control using <u>either</u> of the multi-item instruments if you wish to (optional). *In effect this means no recommendation on a core instrument.*
OPTION 2		Recommend the single-item global instrument "How has your eczema been?" Research on preferred single-item to continue. If possible, measure eczema control using <u>either</u> of the multi-item instruments if you wish to (optional).
OPTION 3	Multi-item instrument preferred	Recommend both RECAP and ADCT as both are so similar that mapping and data synthesis is likely to be possible. If possible, measure eczema control using single-item global if you wish to (optional).
OPTION 4		Recommend one multi-item instrument (either RECAP <u>or</u> ADCT) as the core instrument. If possible, measure eczema control using single-item global if you wish to (optional).

Table 13: options for how to progress with the LTC domain

There followed a lengthy discussion about how to proceed with voting, appropriate wording of the voting questions and the implications of the proposed different options. Rather than voting for the proposed options, it was decided that a straightforward definitive vote would be held to determine whether each existing instrument could be recommended for inclusion in the core set. It was reiterated that because two well-developed, high-quality instruments are available, there is an expectation that the group should be able to move forward towards recommending a core outcome instrument for this domain.

Before voting, some points of clarification were made. Both RECAP and ADCT validation studies included patients of all eczema severities, but were not performed within the context of clinical trials. Two translations are available for the ADCT. RECAP is validated in children and adults, but ADCT is currently only validated in adults, although validation work in children is ongoing. Although ADCT was developed by industry it was confirmed use will be free of charge.

Voting results were not revealed until all voting had been completed. Consensus was reached that both ADCT and RECAP could be recommended as core outcome instruments for measuring eczema control in the core outcome set (Table 14). However, more than 30% voted against the inclusion of PBI and the validated single-item instrument and so these were not recommended for the core set.

	Voting question (definitive votes)	Yes	No	Unsure
		n (%)	n (%)	n (%)
Q1*	Do you agree that ADCT can be recommended as a core instrument to measure long-term control?	36 (66%)	15 (28%)	3 (6%)
Q2*	Do you agree that Recap can be recommended as a core instrument to measure long-term control?	34 (63%)	16 (30%)**	4 (7%)
Q3*	Do you agree that PBI can be recommended as a core instrument to measure long-term control?	1 (2%)	52 (96%)	1 (2%)
Q4*	Do you agree that the validated single-item global can be recommended as a core instrument to measure long-term control?	16 (29%)	35 (64%)	4 (7%)

* *12 conflicted participants did not vote*

*** When rounded up, 30% voted "no" but it is 29.62% unrounded, therefore RECAP was recommended.*

Table 14: Results of voting on candidate LTC instruments. Results of Q1-4 were not presented until voting on Q1-4 was complete

Although the current validated single-item instrument was not considered suitable for the core set, there was a strong desire for a single-item instrument. Developing and validating a single-item control instrument will be included in the future research agenda for the HOME long-term control working group and evaluated at future HOME meeting(s). The control questions in ADCT and RECAP should be considered as candidate instruments, although a new instrument may need to be developed and validated.

Because both ADCT and RECAP were both voted as acceptable for the core set, there was discussion around whether both should be recommended at least for the time being, or whether the group should choose between the two and recommend only one instrument. To recommend more than one would mean deviating from the HOME roadmap and the principle of recommending one instrument per domain, the purpose of which is to ensure trials become more comparable. Although previously, HOME has made difficult choices in order to recommend just one instrument for a domain[2], it was discussed that this situation may be different. ADCT and RECAP are both new instruments, very similar in their content and developed to high standards by two independent groups. Although they both show good psychometric properties in testing, real world evidence for their use is lacking and more experience is needed with both in order to make an informed and evidence-based choice between them. The intention is to move to one instrument in the long term.

The group decided to vote on the principle of including more than one instrument for the time being for the long-term control domain on this occasion. All 62 participants were permitted to vote on this issue, and consensus was reached that this was acceptable with only 19% voting against. This was followed by a confirmatory vote that RECAP and ADCT should be accepted as the options for measuring eczema control for the time being. Conflicted participants did not vote. Only 23% disagreed with this so consensus was reached.

In summary, both the ADCT and RECAP were recommended for the core outcome set for the time being. Researchers can choose to use one or both instruments. Future research will focus on developing and testing a single-item instrument, as well as more validation data on ADCT and RECAP. The long-term control domain will be revisited at a future meeting when additional data and/or instruments are available.

Itch Intensity

The session was opened by Phyllis Spuls, lead for the symptoms working group. The aim of the session was to confirm whether itch intensity should be included as additional essential subdomain of symptoms in the COS and if so, recommend an instrument to measure itch intensity.

Toshiya Ebata presented results of a published systematic review[26] of the measurement properties of itch instruments together with an unpublished recently completed update which included an additional 14 studies. It should be noted that although this review provided valuable evidence for discussion, the data were from studies of patients with pruritus, and therefore, final decisions were based on data from a separate review of data from studies of atopic eczema patients. The presentation focussed on unidimensional (single question) instruments and showed that numerical rating scales (NRS) generally performed better than visual analogue scales (VAS) or verbal rating scales (VRS).

Louise Gerbens then presented further background and evidence. A global patient survey had shown itch is important to patients[27], and HOME had previously agreed itch is an

essential subdomain of symptoms[3]. However, the current recommended core outcome instrument for symptoms (POEM) includes only frequency of itch. It was agreed at HOME V that a measure of itch intensity should be included in the core set, and that a single item instrument should be recommended as multi-item instruments cover more than itch intensity[4]. Therefore, multi-item instruments were not included in the discussion (Table 15).

	Multi-item instruments (not discussed at this meeting)	**Single-item instruments (considered for inclusion in the COS)**
Category B recommendation	Paediatric ISS PO-SCORAD Adapted SA-EASI SA-EASI	
Category C recommendation	Eppendorf Itch Questionnaire (EIQ) Leuven Itch Scale (LIS) Adult Itch Severity Scale (ISS)	
Category D recommendation	Method 4 Subjective SCORAD	VAS (horizontal 0-10) past 24 hours[28] VRS-5 past 24 hours[28] NRS-11 peak itch past 24 hours[29]

Table 15: Instruments with studies assessing one or more measurement properties in patients with atopic eczema

Evidence was then presented from the systematic review of the measurement properties of symptoms instruments in patients with atopic eczema, carried out by the symptoms working group.[30] Category recommendations were based on the original COSMIN rating guidance, in line with the previous review. Two instruments were included (VAS and VRS, both with a recall period of the past 24 hours) which had some limited evidence only for construct validity.[28] An update to the review identified a third single-item itch intensity instrument. The study tested multiple measurement properties including content validity of an eleven point NRS peak itch over the past 24 hours in adults[29]. This study also showed patients found that worst (peak) itch easier to remember and judge than average itch, and fluctuating itch intensity even over a short period can make it difficult for patients to estimate an average itch. Reducing worst itch is important to patients. In summary, only one instrument (NRS-11 peak itch) had any meaningful degree of validation in patients with atopic eczema.

The results of a survey of the HOME membership prior to the meeting were presented (Table 16). A total of 106 responses were received (50 clinicians, 33 patients, 14 Pharma industry, 7 methodologists, 2 not stated). Results showed no clear steer on preferred options. The NRS was preferred by almost half of respondents, the number choosing peak and average itch were equal, and the number choosing a recall period of a week and 24 hours were similar.

Response options preferred:	
Numerical rating scale (NRS)	50 (47%)
Verbal rating scale (VRS)	22 (21%)
Visual analogue scale (VAS)	24 (22%)
Unsure	7 (7%)

Not stated	3 (3%)
Rating of itch intensity preferred:	
Average itch	45 (42.5%)
Peak itch	45 (42.5%)
Unsure	9 (8%)
Not stated	7 (7%)
Timescales for recalling itch intensity preferred:	
Past 24 hours	41 (40%)
Past week	48 (45%)
None of these	9 (8%)
Not stated	8 (7%)

Table 16: Results pre-meeting survey of HOME membership on preferences regarding itch intensity instruments

It was clear throughout the subsequent whole group discussion that the intensity of itch is important to patients with eczema. In fact, it was a widely held view amongst the patients present that if only one outcome could be measured in a trial they would choose itch intensity. Therefore, although itch intensity is captured in the disease control instruments (ADCT and RECAP), and frequency of itch in the symptoms instrument (POEM), patients felt strongly that a separate instrument on itch intensity should be included in the COS.

A summary of the key points raised in the discussion is in Table 17.

• Although validation data is only available for adults, any recommendation would also encourage use in clinical trials involving children where they are able to self complete, in order to generate relevant data for validation in this population. Proxy report of itch is not considered appropriate.
• It is crucial for the pharma industry that any recommended instrument for itch intensity meets FDA PRO development guidance as it may be used for label applications. It was noted that an FDA approved instrument is in use. It was agreed that HOME should be cognisant of this, but also that COS developers can advise the FDA on instrument suitability based on evidence.
• Three pharma companies reported having done conducted qualitative studies on measuring itch (not yet published) all concluding that patients preferred peak (worst) itch over average. All patients present agreed that it is difficult to give an average score for itch due to fluctuations for example between day and night time. This was considered good evidence in support of peak itch.
• There was no strong opinion from patients on whether VAS or NRS was preferred. The VAS is often completed incorrectly, so NRS was considered preferable from a data collection perspective.

Table 17: summary of whole group discussion on itch intensity

It was agreed that pharma industry representatives are potentially conflicted for itch intensity because data on itch intensity can be used for label claims and marketing, so may favour instruments that most closely matches their own. It was therefore agreed that all pharma industry stakeholders would not vote on type of instrument and individual instruments.

The group agreed to proceed with indicative (not binding) votes on the type of instrument that should be included (recall period and peak versus average itch) to provide an indication of the multi-stakeholder group views. This raised some concern about differences in approach to other domains and meetings. It was agreed that there was no need to vote on response options as it was clear from the evidence presented and the discussions that NRS was preferred over VAS or VRS.

These indicative votes clearly showed that itch intensity should be included as a separate subdomain in the symptoms domain of the COS (Table 18). Peak itch was preferred to average itch, and a recall period of 24 hours was favoured over a period of a week.

	Indicative question	Yes n (%)	No n (%)	Unsure n (%)
Q1	Is itch intensity something you always want separate information on in a clinical trial?	52 (81%)	6 (9%)	6 (9%)
Q2*	Would you prefer itch intensity to be measured using average itch?	10 (20%)	34 (68%)	6 (12%)
Q3*	Would you prefer itch intensity to be measured using worst (peak) itch intensity?	40 (80%)	3 (6%)	7 (14%)
Q4*	Do you consider the last 24 hours to be the most appropriate timescale for capturing worst (peak) itch?	40 (80%)	4 (8%)	6 (12%)
Q5*	Do you consider the last week to be the most appropriate timescale for capturing worst (peak) itch?	10 (20%)	28 (56%)	12 (24%)

* *Pharma industry participants did not vote due to potential conflict*

Table 18: results of indicative voting on itch intensity

In order to move forward towards recommending an instrument, Christian Apfelbacher recapped the data on the quality and findings of the eight validation studies on NRS-11 peak itch in pruritus patients.[26] He showed that the only data available were on reliability and hypothesis testing and that none had undergone content validity testing. He then presented the validation data on content validity, test-retest reliability, known groups validity, sensitivity to change and construct validity of the NRS-11 peak itch instrument tested in adults with atopic eczema published[29] after the review update. There followed a further whole group discussion; the key points are summarised in Table 19

• Content validity data for NRS-11 peak itch is strong, through patient opinion at this meeting, the recent validation study of NRS-11 peak itch and unpublished pharma studies in which very similar instruments have been developed.
• There was some debate about the availability of the NRS-11 peak itch worded as per the Yosipovitch study and whether it was possible to copyright such a widely used question.
• Although multiple versions of NRS-11 peak itch are available that differ only slightly, it is key that for the COS, just one version is recommended to avoid unnecessary variability.
• It was noted that there had been little discussion about recall period, although the general patient opinion in the meeting was that any longer would be difficult. It was pointed out that the only instrument for which validation is available in atopic eczema patients, the NRS-11 peak itch, has a 24 hour recall period. So whilst other recall periods may be acceptable the instrument with 24 hour recall is the only one currently shown to be valid.

Table 19: summary of discussion

The group then agreed to move to a final vote on the recommendation of an instrument for measuring itch intensity in the COS (Table 20).

	Voting question	Yes n	No n	Unsure n

		(%)	(%)	(%)
Q6*	Do you think there is sufficient evidence to vote on the NRS peak itch?	40 (80%)	5 (10%)	5 (10%)
Q7*	Do you agree to include the peak NRS-11 over the past 24 hours**[29] as the core outcome instrument for measuring the subdomain of itch intensity in adults	42 (86%)	4 (8%)	3 (6%)

* *Pharma industry participants did not vote due to potential conflict*

** wording of the question: "On a scale of 0 to 10, with 0 being 'no itch' and 10 being 'worst itch imaginable', how would you rate your itch at the worst moment during the previous 24 hours?"

Table 20: final voting on itch intensity

In summary, since only 8% disagreed, peak NRS-11 over the past 24 hours as specified in Yosipovitch 2019[29] was voted as the recommended instrument for measuring itch intensity in adults. However, it was agreed that researchers should be encouraged to use this instrument for any age where self-completion is considered appropriate. Further validation in children is required.

Summary

This meeting successfully completes the core outcome set for clinical trials in eczema, the first in dermatology. There may be changes to the core set in the future, particularly around the long-term control domain if a single item instrument is developed and preferred. However, there are a clear set of instruments for trialists to include in all future eczema trials which will significantly increase the ability to synthesis data from future trials.

The following instruments were recommended during HOME VII for the COS:

Domain	**Recommended instruments for the COS**
Quality of Life	DLQI – adults CDLQI – children IDQoL – infants/young children
Long term control	ADCT and RECAP (only one required to be measured)
Itch intensity	Peak NRS-11 over the past 24 hours[29]

The core outcome set should be reported at baseline and the primary endpoint (usually end of treatment) as a minimum, but further work is required to facilitate discussions regarding appropriate interim time points.

The challenge now facing HOME is to ensure the recommendations are taken up by the wider research community. The completion of the core set is timely as there are several new therapeutic interventions currently being developed for treating eczema.

Strengths and weaknesses

A range of stakeholders from around the world participated in the meeting. There was increased representation from Asia compared to previous meetings, particularly from Japan and South Korea, demonstrating the need to hold consensus meetings in different areas of the world. There were notable exceptions though, including African countries and China, as has been the case for previous meetings. A representative from the Japanese regulatory authority was present for part of the meeting, but it remains a challenge to attract representatives from the EMEA or the FDA to HOME meetings. The meeting included an independent moderator (Maarten Boers) and had significant patient input.

Participants were well prepared for the meeting; they received the materials and evidence prior to the meeting, undertook pre-meeting tasks to become familiar with the instruments under discussion, and the two orientation sessions provided the opportunity to ask questions and get clarification.

The high proportion of participants who had been involved in the development of either ADCT or RECAP meant that the number available to vote was significantly diminished for the long-term domain. Similarly the decision taken to exclude all pharma representatives from voting on itch intensity instruments reduced the number available. However, all participants were able to input into the discussions prior to voting.

SUPPLEMENTARY MATERIAL

Meeting participants and stakeholder group

Maarten Boers: independent facilitator

1. Abhijit Gadkari	Pharmaceutical Industry	USA
2. Akane Yasui	Patient / patient representative	Japan
3. Åke Svensson	Clinician	Sweden
4. Alison Sears	Clinician	UK
5. Alix Bullock	Patient / patient representative	UK
6. Amy DeLozier	Pharmaceutical Industry	USA
7. Andreas Wollenberg	Clinician	Germany
8. Annika Volke	Clinician	Estonia
9. Ashish Bansal	Pharmaceutical Industry	USA
10. Beth Stuart	Methodologist	UK
11. Bo Bang	Pharmaceutical Industry	Denmark
12. Brian Calimlim	Pharmaceutical Industry	USA
13. Carl-Fredrik Wahlgren	Clinician	Sweden
14. Chanho Na	Clinician	South Korea
15. Christian Apfelbacher	Methodologist	Germany
16. Claire Feeney	Pharmaceutical Industry	UK
17. Dong Hun Lee	Clinician	South Korea
18. Dora Stölzl	Clinician	Germany
19. Eri Maruyama	Patient / patient representative	Japan
20. Eric Simpson	Clinician	USA
21. Fabio Nunes	Pharmaceutical Industry	USA
22. Henrique Ishii	Patient / patient representative	Brazil
23. Henrique Teixeira	Pharmaceutical Industry	USA
24. Hidehisa Saeki	Clinician	Japan
25. Hiroyuki Murota	Clinician	Japan
26. Hiroyuki Toyama	Pharmaceutical Industry	Japan
27. Hyejung Jung	Clinician	South Korea
28. Hywel Williams	Clinician	UK
29. Isabelle Guillemin	Pharmaceutical Industry	France
30. Jan Gutermuth	Clinician	Belgium
31. Jean-Francois Stalder	Clinician	France
32. Jennifer Austin	Patient / patient representative	USA
33. Jiyoung Ahn	Clinician	South Korea
34. Joanne Chalmers	Methodologist	UK
35. Jochen Schmitt	Clinician	Germany
36. Jooyoon Bae	Clinician	South Korea
37. Julie Block	Patient / patient representative	USA

38. Katrina Abuabara	Clinician	USA
39. Kazue Yoshida	Clinician	Japan
40. Ken Igawa	Clinician	Japan
41. Kim Thomas	Methodologist	UK
42. Kyoko Maru	Patient / patient representative	Japan
43. Louise Gerbens	Clinician	The Netherlands
44. Laura Howells	Methodologist	UK
45. Laura von Kobyletzki	Clinician	Sweden
46. Laurent Eckert	Clinician	France
47. Linda Wang	Pharmaceutical Industry	Sweden
48. Lynita Howie	Patient / patient representative	Australia
49. Magdalene Dohil	Clinician	USA
50. Mami Murakami	Pharmaceutical Industry	Japan
51. Maria Bradley	Clinician	Sweden
52. Masaki Futamura	Clinician	Japan
53. Masutaka Furue	Clinician	Japan
54. Michaela Gabes	Methodologist	Germany
55. Michael Lanigan	Patient / patient representative	Canada
56. Miwako Ogino	Patient / patient representative	Japan
57. Norito Katoh	Clinician	Japan
58. Phyllis Spuls	Clinician	The Netherlands
59. Rai Fujimoto	Clinician	Japan
60. Rosemary Humphreys	Patient / patient representative	UK
61. Sebastien Barbarot	Clinician	France
62. Stephan Weidinger	Clinician	Germany
63. Susumu Ichiyama	Clinician	Japan
64. Tae Young Han	Clinician	South Korea
65. Takeshi Nakahara	Clinician	Japan
66. Tatsuki Fukuie	Clinician	Japan
67. Teresa Berents	Clinician	Norway
68. Tim Burton	Patient / patient representative	UK
69. Toshiya Ebata	Clinician	Japan
70. Vanessa Sultana	Patient / patient representative	Australia
71. Yasutomo Imai	Clinician	Japan
72. Yik Weng Yew	Clinician	Singapore
73. Yoko Kataoka	Clinician	Japan
74. Yukihiro Ohya	Clinician	Japan
75. Yuko Ikegami	Patient / patient representative	Japan

Small group allocations

Group 1

1. Eric Simpson (facilitator)
2. Alix Bullock
3. Eri Maruyama
4. Henrique Ishii
5. Åke Svensson
6. Chanho Na
7. Takeshi Nakahara
8. Magdalene Dohil
9. Hyejung Jung
10. Yoko Kataoka
11. Abhijit Gadkari
12. Mami Murakami

Group 2

1. Christian Apfelbacher (facilitator LTC)
2. Joanne Chalmers (facilitator QoL)
3. Jennifer Austin
4. Vanessa Sultana
5. Alison Sears
6. Kazue Yoshida
7. Jiyoung Ahn
8. Tatsuki Fukuie
9. Jan Gutermuth
10. Katrina Abuabara
11. Masaki Futamura
12. Amy DeLozier
13. Henrique Teixeira

Group 3

1. Jochen Schmitt (facilitator)
2. Julie Block
3. Yuko Ikegami
4. Andreas Wollenberg
5. Dong Hun Lee
6. Hiroyuki Toyama
7. Laura Howells
8. Masutaka Furue
9. Norito Katoh
10. Teresa Berents
11. Fabio Nunes
12. Ashish Bansal

Group 4

1 Phyllis Spuls (facilitator)
2 Maarten Boers
3 Akane Yasui
4 Lynita Howie
5 Kyoko Maru
6 Rai Fujimoto
7 Sebastien Barbarot
8 Toshiya Ebata
9 Annika Volke
10 Dora Stölzl
11 Jooyoon Bae
12 Isabelle Guillemin
13 Bo Bang

Group 5

1 Louise Gerbens (facilitator LTC)
2 Kim Thomas (facilitator other domains)
3 Michael Lanigan
4 Rosemary Humphreys
5 Beth Stuart
6 Hidehisa Saeki
7 Jean-Francois Stalder
8 Ken Igawa
9 Stephan Weidinger
10 Susumu Ichiyama
11 Yasutomo Imai
12 Laurent Eckert
13 Brian Calimlim

Group 6

1 Hywel Williams (facilitator)
2 Miwako Ogino
3 Tim Burton
4 Carl-Fredrik Wahlgren
5 Hiroyuki Murota
6 Laura von Kobyletzki
7 Maria Bradley
8 Yik Weng Yew
9 Yukihiro Ohya
10 Michaela Gabes
11 Tae Young Han
12 Linda Wang
13 Claire Feeney

Survey of HOME membership on timing and frequency of outcome measurement

Breakdown of respondents by country

Country	Number of respondents
Australia	2
Belgium	1
Brazil	1
Canada	4
Denmark	6
Estonia	1
France	5
Germany	6
Ireland	1
Israel	1
Italy	2
Japan	19
Mexico	1
Netherlands	9
Norway	1
Not stated	4
Republic of Korea	3
Singapore	1
South Africa	1
Spain	2
Sweden	2
Switzerland	3
UK	18
USA	12
Total	**106**

Responses

If assessing <u>initial response</u> to atopic eczema treatments, at what time-point would you prefer assessments to be made?		
Options	**Number of responses**	
1 week	26	
2 weeks	18	
4 weeks	33	
8 weeks	1	
None of these - please explain:	11	Reasons: • Depends on mode of action and expected time to onset of response • Should be at the clinicians discretion • Depends on the domain • Depends on the baseline disease severity
Left blank	17	
If assessing <u>medium-term</u> response to atopic eczema treatment, at what time-point would you prefer assessments to be made?		
8 week	22	
12 weeks	29	
16 weeks	13	
24 weeks	9	
None of these - please explain:	15	Reasons: • Depends on the mode of action and expected time for efficacy • Prefer 4 weeks • Prefer 48 weeks • Defer to the clinician/patient • Depends on baseline disease severity • How relevant is tis to a patient with life-long disease? • May not need to specify
Left blank	18	

Instructions to the breakout groups for QoL content validity study, using the COMSIN criteria and rating system

Group instructions:

- Rate the content validity of the measurement instrument for the instrument you have been assigned using the COSMIN criteria and rating system above.
- Please rate each criterion as sufficient (+) or insufficient (-).
- The general rule to give a sufficient rating per criterion is:
- +: ≥85% of the items of the PROM fulfil the criterion
- –: <85% of the items of the PROM does fulfil the criteria

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

References

1. 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(9):1111-7. doi: 10.1111/j.1398-9995.2012.02874.x [published Online First: 2012/08/01]
2. Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. *J Allergy Clin Immunol* 2014;134(4):800-7. doi: 10.1016/j.jaci.2014.07.043 [published Online First: 2014/10/06]
3. Spuls PI, Gerbens LA, Simpson E, et al. Patient-Oriented Eczema Measure (POEM), a core instrument to measure symptoms in clinical trials: a Harmonising Outcome Measures for Eczema (HOME) statement. *Br J Dermatol* 2016 doi: 10.1111/bjd.15179 [published Online First: 2016/11/20]
4. Chalmers JR, Thomas KS, Apfelbacher C, et al. Report from the fifth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2018;178(5):e332-e41. doi: 10.1111/bjd.16543 [published Online First: 2018/04/20]
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(1):24-30. doi: 10.1038/jid.2014.320 [published Online First: 2014/09/05]
6. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17(1):449. doi: 10.1186/s13063-016-1555-2 [published Online First: 2016/09/14]
7. Paudyal P, Apfelbacher C, Jones C, et al. "DLQI Seems to be 'Action', and Skindex-29 Seems to be 'Emotion'": Qualitative Study of the Perceptions of Patients with Psoriasis or Eczema on Two Common Dermatology-specific Quality of Life Measures. *Acta Derm Venereol* 2020;100(8):adv00105. doi: 10.2340/00015555-3417 [published Online First: 2020/02/06]
8. Gabes M, Chamlin SL, Lai JS, et al. Development of a validated short-form of the Childhood Atopic Dermatitis Impact Scale, the CADIS-SF15. *J Eur Acad Dermatol Venereol* 2020;34(8):1773-78. doi: 10.1111/jdv.16362 [published Online First: 2020/03/17]
9. Gabes M, Apfelbacher C. IDQoL, CDLQI and the 45-item CADIS received a sufficient content validity rating during the HOME VII meeting in Japan: a group discussion study. *J Eur Acad Dermatol Venereol* 2021;35(2):458-63. doi: 10.1111/jdv.16848 [published Online First: 2020/08/04]
10. Gabes M, Tischer C, Apfelbacher C. Measurement properties of quality-of-life outcome measures for children and adults with eczema: An updated systematic review. *Pediatr Allergy Immunol* 2020;31(1):66-77. doi: 10.1111/pai.13120 [published Online First: 2019/09/11]
11. Heinl D, Prinsen CA, Deckert S, et al. Measurement properties of adult quality-of-life measurement instruments for eczema: a systematic review. *Allergy* 2016;71(3):358-70. doi: 10.1111/all.12806 [published Online First: 2015/11/14]

12. Heinl D, Prinsen CAC, Sach T, et al. Measurement properties of quality-of-life measurement instruments for infants, children and adolescents with eczema: a systematic review. *Br J Dermatol* 2017;176(4):878-89. doi: 10.1111/bjd.14966 [published Online First: 2016/08/21]
13. Chalmers JR, Simpson E, Apfelbacher CJ, et al. Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2016;175(1):69-79. doi: 10.1111/bjd.14773 [published Online First: 2016/07/21]
14. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3 [published Online First: 2018/02/13]
15. Vincent R, Chalmers JR, McWilliams C, et al. Assessing uptake of the Harmonising Outcome Measures for Eczema (HOME) Core Outcome Set and recommended instruments. *Br J Dermatol* 2020;183(3):566-68. doi: 10.1111/bjd.19030 [published Online First: 2020/03/14]
16. Grinich EE, Schmitt J, Küster D, et al. Standardized reporting of the Eczema Area and Severity Index (EASI) and the Patient-Oriented Eczema Measure (POEM): a recommendation by the Harmonising Outcome Measures for Eczema (HOME) Initiative. *Br J Dermatol* 2018;179(2):540-41. doi: 10.1111/bjd.16732 [published Online First: 2018/05/11]
17. Howells L, Thomas KS, Sears AV, et al. Defining and measuring 'eczema control': an international qualitative study to explore the views of those living with and treating atopic eczema. *J Eur Acad Dermatol Venereol* 2019;33(6):1124-32. doi: 10.1111/jdv.15475 [published Online First: 2019/02/06]
18. Silverberg JI, Chiesa Fuxench ZC, Gelfand JM, et al. Content and construct validity, predictors, and distribution of self-reported atopic dermatitis severity in US adults. *Ann Allergy Asthma Immunol* 2018;121(6):729-34.e4. doi: 10.1016/j.anai.2018.07.040 [published Online First: 2018/08/10]
19. 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. doi: 10.1080/03007995.2019.1699516 [published Online First: 2019/11/30]
20. Howells LM, Chalmers JR, Gran S, et al. Development and initial testing of a new instrument to measure the experience of eczema control in adults and children: Recap of atopic eczema (RECAP). *Br J Dermatol* 2020;183(3):524-36. doi: 10.1111/bjd.18780 [published Online First: 2019/12/04]
21. Blome C, von Usslar K, Augustin M. Feasibility of Using Qualitative Interviews to Explore Patients' Treatment Goals: Experience from Dermatology. *Patient* 2016;9(3):261-9. doi: 10.1007/s40271-015-0149-5 [published Online First: 2015/11/01]
22. Augustin M, Radtke MA, Zschocke I, et al. The patient benefit index: a novel approach in patient-defined outcomes measurement for skin diseases. *Arch Dermatol Res* 2009;301(8):561-71. doi: 10.1007/s00403-009-0928-8 [published Online First: 2009/03/28]
23. Topp J, Augustin M, von Usslar K, et al. Measuring Patient Needs and Benefits in Dermatology using the Patient Benefit Index 2.0: A Validation Study. *Acta Derm Venereol* 2019;99(2):211-17. doi: 10.2340/00015555-3063 [published Online First: 2018/10/12]

24. Darrigade AS, Colmant C, de Montjoye L, et al. Atopic Dermatitis Score 7 (ADS7): A promising tool for daily clinical assessment of atopic dermatitis. *Allergy* 2020;75(5):1264-66. doi: 10.1111/all.14104 [published Online First: 2019/11/15]
25. Vakharia PP, Chopra R, Sacotte R, et al. Validation of patient-reported global severity of atopic dermatitis in adults. *Allergy* 2018;73(2):451-58. doi: 10.1111/all.13309 [published Online First: 2017/09/15]
26. Schoch D, Sommer R, Augustin M, et al. Patient-Reported Outcome Measures in Pruritus: A Systematic Review of Measurement Properties. *J Invest Dermatol* 2017;137(10):2069-77. doi: 10.1016/j.jid.2017.05.020 [published Online First: 2017/09/02]
27. von Kobyletzki LB, Thomas KS, Schmitt J, et al. What Factors are Important to Patients when Assessing Treatment Response: An International Cross-sectional Survey. *Acta Derm Venereol* 2016 doi: 10.2340/00015555-2480 [published Online First: 2016/06/16]
28. Kido-Nakahara M, Katoh N, Saeki H, et al. Comparative cut-off value setting of pruritus intensity in visual analogue scale and verbal rating scale. *Acta Derm Venereol* 2015;95(3):345-6. doi: 10.2340/00015555-1972 [published Online First: 2014/09/18]
29. Yosipovitch G, Reaney M, Mastey V, et al. Peak Pruritus Numerical Rating Scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. *Br J Dermatol* 2019;181(4):761-69. doi: 10.1111/bjd.17744 [published Online First: 2019/02/08]
30. Gerbens LA, Prinsen CA, Chalmers JR, et al. Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. *Allergy* 2017;72(1):146-63. doi: 10.1111/all.12959 [published Online First: 2016/06/21]