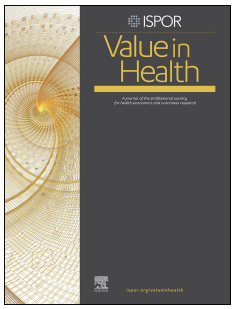

Practicality, Validity and Responsiveness of Using the Proxy Version of the CHU-9D with Children Aged 2 to 5 Years

Tracey H. Sach, PhD, Hywel C. Williams, DSc, on behalf of the BEEP study team, Hilary Allen, Robert Boyle, Maeve Kelleher, Sara Brown, Mike Cork, Carsten Flohr, Nicola Jay, Stella Lartey, Charlotte Davies, Sandra Lawton, Michael Perkin, Matthew Ridd, Tracey Sach, Joanne Brooks, Rachel Haines, Eleanor Mitchell, Alan Montgomery, Richard Swinden, Stella Tarr, Laura Wyatt, Kim Thomas, Hywel Williams, Joanne Chalmers, Susan Davies-Jones

PII: S1098-3015(24)02861-4

DOI: https://doi.org/10.1016/j.jval.2024.08.010

Reference: JVAL 4129

To appear in: *Value in Health*

Received Date: 4 October 2023

Revised Date: 12 August 2024

Accepted Date: 19 August 2024

Please cite this article as: Sach TH, Williams HC, on behalf of the BEEP study team, Allen H, Boyle R, Kelleher M, Brown S, Cork M, Flohr C, Jay N, Lartey S, Davies C, Lawton S, Perkin M, Ridd M, Sach T, Brooks J, Haines R, Mitchell E, Montgomery A, Swinden R, Tarr S, Wyatt L, Thomas K, Williams H, Chalmers J, Davies-Jones S, Practicality, Validity and Responsiveness of Using the Proxy Version of the CHU-9D with Children Aged 2 to 5 Years, *Value in Health* (2024), doi: https://doi.org/10.1016/j.jval.2024.08.010.

Target Journal: Value in Health

Title: Practicality, Validity and Responsiveness of Using the Proxy Version of the CHU-9D with Children Aged 2 to 5 Years

Authors: Tracey H Sach, PhD,[1,2] Hywel C Williams, DSc[3] on behalf of the BEEP study team*.

Author Affiliations: [1] School of Primary Care, Population Sciences and Medical Education, , University of Southampton, Southampton, United Kingdom; [2] Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, United Kingdom. [3] Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, United Kingdom

Corresponding Author Information:

Tracey H Sach, PhD,

School of Primary Care, Population Sciences and Medical Education,

Faculty of Medicine, University of Southampton,

Aldermoor Health Centre, Southampton SO16 5ST.

Email: t.sach@soton.ac.uk.

ORCID: 0000-0002-8098-9220.

Precis: Valuing the health of children aged 2 to 5 years using the CHU-9D shows potential according to the three measurement properties (practicality, validity, responsiveness) tested. (25 words)

Word Count: 4,533

Number of Pages: 24

Number of Figures: 0

Number of Tables: 4

Supplementary materials:

Pages: 2

Figures: 0

Tables: 3

Author Contributions:

Concept and design: Sach, Williams

Acquisition of data: Williams

Analysis and interpretation of data: Sach

Drafting of the manuscript: Sach

Critical revision of the paper for important intellectual content: Sach, Williams

Statistical Analysis: Sach

Provision of study materials or patients: Williams

Obtaining funding: Williams, Sach

Administrative, technical, or logistic support: Sach, Williams

Supervision: Sach

Other (if applicable, please specify): N/A

Funding/support: This study presents independent research funded by the National Institute for Health and Care Research (NIHR) under its Health Technology Assessment programme (12/67/12). Research nurse support was provided by the NIHR Clinical Research Networks. The trial was developed with and supported by the UK Dermatology Clinical Trials Network (UK DCTN). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

This trial was registered prospectively with the ISRCTN registry (ISRCTN21528841). The study was sponsored by the University of Nottingham, coordinated by the Nottingham Clinical Trials Unit (CTU)

The role of the funder/sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements: We would like to thank the parents and infants who took time to participate in this trial, and the patients who contributed to trial design by providing helpful feedback at different stages of trial development. We would like to thank Daniel Simpkins from the Nottingham CTU for providing the trial database. We acknowledge the contribution of the whole BEEP study team* for making this paper possible. Particular thanks go to Bradshaw, Brown, Davies, and Flohr for reviewing the manuscript and making useful suggestions. The corresponding author attests that listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Sach is the guarantor.

*The BEEP trial team members are:

Hilary Allen, Robert Boyle, and Maeve Kelleher, National Heart and Lung Institute, Imperial College London, London, UK

Sara Brown. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Scotland, UK;

Mike Cork, Sheffield Dermatology Research, Department of Infection and Immunity, University of Sheffield, Sheffield, UK

Carsten Flohr, Unit for Paediatric & Population-Based Dermatology Research, St John's Institute of Dermatology, Guy's & St Thomas' NHS Foundation Trust and King's College London, UK;

Nicola Jay, Sheffield Children's Hospital, Sheffield, UK;

Stella Lartey and Charlotte Davies, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK;

Sandra Lawton, Rotherham NHS Foundation Trust, UK;

Michael Perkin, Population Health Research Institute, St. George's, University of London, London, UK;

Matthew Ridd, Population Health Sciences, University of Bristol, Bristol, UK;

Tracey Sach, School of Primary Care, Population Sciences and Medical Education, University of Southampton, UK; and Health Economics Group, University of East Anglia, UK.

Joanne Brooks, Rachel Haines, Eleanor Mitchell, Alan Montgomery, Richard Swinden, Stella Tarr, and Laura Wyatt, Nottingham Clinical Trials Unit, University of Nottingham, Nottingham, UK;

Kim Thomas, Hywel Williams, Joanne Chalmers, and Susan Davies-Jones, Centre of Evidence Based Dermatology, School of Medicine, University of Nottingham, Nottingham, UK;

Ethics Statement: Informed consent was obtained from mothers during pregnancy, or from the mother, father, or guardian after delivery. The trial was overseen by an independent Trial Steering Committee and approved by the West Midlands Ethics Committee, UK (14/WM/0162).

ABSTRACT

Objectives: To assess the practicality, validity and responsiveness of proxy CHU-9D in children aged 2-5 years.

Methods: We used data from BEEP, a UK randomised controlled trial testing whether daily emollients in infancy could prevent eczema in high-risk infants. The main parent/carer completed the proxy CHU-9D using developers' additional guidance for completion in under-5's and the Patient-Orientated Eczema Measure (POEM) at ages 2, 3, 4 and 5. Practicality was assessed by completion rates. Construct validity assessed if CHU-9D could discriminate between those with/without eczema and between eczema severity levels on POEM. Responsiveness was determined by ability to discriminate between three groups: those whose POEM score, i) deteriorated ≥3 points, ii) change not clinically important (-2.9 to 2.9 points), and iii) improved ≥3 points. Analysis was conducted in STATA 17.

Results: Of 1,394 children participating in BEEP, study questionnaires were completed by 1,212 (87%), 981 (70%), 990 (71%), and 976 (70%) at 2, 3, 4 and 5-years. Of these the CHU-9D was completed by 1,066 (88.0%), 685 (69.8%), 925 (93.4%) and 923 (94.6%) respectively. Mean utility at all timepoints was around 0.934 (range 0.443-1). For construct validity, very small differences on the CHU-9D between known groups were observed($p < 0.01$). 801 participants had responsiveness data: 13% deteriorated, 72% had non-clinically important change, and 15% improved. Mean utility change (standardised response mean) for these groups was -0.0198 (0.21), 0.0041 (0.05), and 0.0175 (0.21)showing small change and small responsiveness.

Conclusions: Proxy CHU-9D in 2-5 year old children shows potential but further research is needed.

Keywords: CHU-9D, Paediatric, psychometric properties, Proxy

Highlights:

- Measuring child utility in health economic evaluations is challenging. The Child Health Utility – 9 dimension (CHU-9D) is a generic preference-based measure with 9 dimensions each with 5 levels that has been used with children aged ≥5 years. Few studies have examined the psychometric properties of CHU-9D in the under 5's.
- This paper explores the practicality, validity (construct and convergent), and responsiveness of the proxy CHU-9D in children aged 2 to 5 years using data collected as part of a previously reported clinical trial.
- The practicality of the proxy CHU-9D improved with age. In terms of validity and responsiveness only small changes and responsiveness were observed for the relatively healthy children aged under 5 years in this study. A small proportion found the 'School Work/Homework' question difficult particularly at the lower age range despite additional guidance. Further research is needed to corroborate these findings, examine other measurement properties, and consider the appropriateness of the value set for younger children.

Introduction

Economic evaluations inform resource allocation decisions in many countries. This is often undertaken using cost-utility analysis but questions remain on how best to measure child utility for use in such studies. Measuring child utility in economic evaluations is challenging,[1-5] especially in younger children and in studies where the age of child participants spans wide age/developmental ranges. Despite an increasing range of measures[6-22] and interest in the area of child outcome measurement, there is still a lack of guidance about how to measure child health utility. For instance, the National Institute for Health and Care Excellence (NICE) in the UK do not currently recommend a specific measure of health-related quality of life (HRQL) in children or young people although they do state a generic measure with good psychometric properties be used and that details of who completed the questionnaire should be reported.[23]

Research has shown that no single childhood measure performs better than others on all psychometric properties, and that further testing of measures in terms of their psychometric properties is needed particularly for measures in preschool children and for the CHU-9D.[24,25] There is little evidence about the responsiveness of the CHU-9D and what is available is mixed.[25] Whilst there is positive evidence of construct validity for the CHU-9D, this evidence comes from studies of children aged over 5 years.[25]

When designing the economic evaluation to conduct alongside the BEEP clinical trial[26-29] (study start date June 2014), a UK multicentre, pragmatic randomised controlled trial, designed to estimate the effectiveness and cost-effectiveness of advice to apply emollients at least once daily all over the infants body for the first year of life in order to prevent eczema (also known as atopic eczema or atopic dermatitis) compared to best practice skincare advice

only (control group), it was necessary to consider how best to measure outcomes. A review of the literature at that time offered little guidance and there were no available, validated preference-based instruments for this age group. For this reason it was decided the primary economic study would be a cost-effectiveness analysis using the primary clinical outcome from the trial. However, in making this decision we recognised the importance of developing and/or testing utility instruments for younger children if interventions aimed at this group are to be compared for the purpose of informing resource allocation decisions. Whilst cost-effectiveness analyses using clinical outcomes can be performed it is often unclear how much a decision maker would be willing to pay for a unit of clinical outcome (e.g. to prevent a case, or reduce the severity, of eczema) or how that willingness to pay would compare across clinical outcomes. Therefore, we saw the BEEP study as an opportunity to undertake research to contribute evidence to this under-researched area.

At the time of the study design there were few options in terms of preference-based instruments for this age group, although there is now other instruments for the first year of life[18-20] and 0-3 years[21-22] which had they existed at the time we might have considered. Having had experience of using the CHU-9D in a previous trial[30] with children aged ≥5 years, we knew from the developer that they had produced guidance to help parents of pre-school age children complete the CHU-9D but this had not been tested. The BEEP trial seemed a good opportunity to contribute evidence towards testing the suitability of the instrument for this age group and the trial team, which includes patient and public members, were supportive having considered the additional guidance and questionnaire. There is now evidence that the CHU-9D is beginning to be used more widely in the under 5's[31, 32] and therefore it is important a body of evidence is built to understand and support or otherwise the use of the instrument with pre-school children.

The aim of this study was therefore to explore the practicality, validity (construct) and responsiveness of the proxy CHU-9D completed by the main parent/carer on behalf of their child aged 2 to 5 years in order to identify if the proxy version is potentially suitable for use with younger children.

Methods

Participants

The data for this study comes from participants taking part on the Barrier Enhancement for Eczema Prevention (BEEP) trial.[26-29] In this trial Infants were randomised (1:1) to receive either emollient and best practice skin-care advice (emollient group) or best practice skincare advice only (control group) within a maximum of 21 days from delivery. Families were recruited to the study through their contact with antenatal or postnatal services, by invitation letters from their general practitioners, or through posters describing the study in hospitals and the community. The infants had to be born at term (at least 37 weeks' gestation) and be at high risk of developing eczema (defined by the presence of at least one first-degree relative with parent/carer-reported eczema, allergic rhinitis, or asthma diagnosed by a doctor). Infants were excluded if they had a severe widespread skin condition making it hard to detect or assess eczema or if they had a serious health issue. Ethical approval was granted by West Midlands Ethics Committee, UK (14/WM/0162).

Outcomes measures

The primary outcome in the BEEP trial was a first diagnosis of eczema using validated diagnostic criteria (UK working party refinement of the Hanifin and Rajka diagnostic criteria[33]) assessed by research nurses masked to treatment allocation at age 2 years. The trial

also collected information from the main parent/carer annually from 2 years about (i) any report of a clinical diagnosis of eczema, (ii) the presence of eczema using parental completion of the UKWP diagnostic criteria for eczema, and (iii) parental report of the child suffering from eczema in the last year (not asked at 2 years). To ensure consistency the main parent/carer was also asked to complete two outcome questionnaire measures – Patient Oriented Outcome Measure (POEM) and the Child Health Utility – nine dimensions instrument (CHU-9D) at 2, 3, 4 and 5 years of age for their child. These were given at a face to face (either at home or clinic according to parental preference) visit at 2 years for self-completion by the parent/carer and via online questionnaires at 3, 4 and 5 years unless a preference for postal paper-based questionnaires was expressed.

The Patient Oriented Outcome Measure (POEM) consists of seven questions about eczema symptoms (itch, sleep disturbance, skin bleeding, skin weeping, skin cracking, skin flaking and skin dryness) rated "Over the last week" as "no days (0), 1-2 days (1), 3-4 days (2), 5-6 days (3), everyday (4)". The scores to individual questions are added together to give an overall score that ranges between 0 (no eczema) to 28 (very severe eczema).[34,35] Severity of eczema, as assessed by the main carer, can be grouped by severity as follows: score of 0 to 2 = clear or almost clear; 3 to 7 = mild eczema; 8 to 16 = moderate eczema, 17 to 24 = severe eczema; 25 to 28 = very severe eczema.[34] The POEM has been used in relevant NICE guidelines and is recommended as a core outcome measure by the Harmonising Outcome Measure for Eczema (HOME, https://www.homeforeczema.org/) initiative. Research has suggested that a 3-point change or more on the POEM is likely to represent a clinically important difference.[36] The POEM was always before the CHU-9D in questionnaires.

The Child Health Utility – nine dimensions (CHU-9D), is a generic preference based instrument, consisting of nine dimensions (worried, sad, pain, tired, annoyed, schoolwork/homework, sleep, daily routine, able to join in activities) rated across 5 levels (as either doesn't feel/have, a little bit, a bit, quite or very or as no problems/a few problems/some problems/ many problems/can't do). The self-complete version asks the child to choose one option for each question which best describes themself today whereas the proxy version asks the parent to make this choice. [12,13] Valuation interviews were undertaken with 300 members of the UK adult general population to obtain preference weights for a sample of the health states in the CHU-9D descriptive system using standard gamble and ordinary least squares (OLS) with utility ranging from 0.337 to 1.[37] The original version was developed with children, specifically for children aged 7 to 11 years. Work has since been published validating the CHU-9D in 11 to 17 year olds[38-40] and a proxy version for 5 and 6 year olds.[41] However, as yet there is limited evidence to support the use of the proxy CHU-9D in pre-school aged children, despite the developer having additional guidance designed to help parents of preschool aged children complete the instrument. This additional guidance consists of extra text contextualising questions for children not yet in school but who may or may not be in nursery. To illustrate this, consider question 6 as an example. This question asks about "school work/homework (such as reading, writing, doing lessons)", the additional guidance adds the following text to advise how to answer if the child does not go to school or preschool:

"If your child is at preschool/nursery/kindergarten then please think about that. If your child didn't go today because of their health and they usually would have, please tick the last option "My child can't do their schoolwork/homework today". If today is not a day they usually would have gone, then please think about how you think they would have been had they gone. If your child does not go to preschool/nursery/kindergarten, then please think

about whether they have had any problems with activities such as colouring, looking at books/reading, and concentrating, as appropriate for their age". The CHU-9D was used under a license applied for in 2016 via https://licensing.sheffield.ac.uk/product/CHU-9D for the BEEP trial.

The primary economic evaluation[29] undertaken for the BEEP trial was a cost-effectiveness analysis using the clinical outcome as the main outcome in recognition that the use of the CHU-9D in this age group is experimental as acknowledged in the limitations of that work.

Assessing the performance of the CHU-9D

Practicality

Practicality (sometimes referred to as feasibility or acceptability) was assessed by measuring completion rates[42] for the CHU-9D at different timepoints. All nine questions have to be complete in order to calculate a utility.[37] We also report the number (%) who scored full health or the worst health state possible in addition to the number of unique health states reported. Practicality could also be assessed using a range of qualitative methods. In this study we included one open-ended free-text question after the CHU-9D to ask participants how they found completing the CHU-9D.

Validity

Construct validity was assessed by whether the CHU-9D could discriminate between (a) individuals who had eczema according to established diagnostic criteria, or otherwise (here defined using UKWPdiagnostic criteria[33]); (b) any parental report of a clinical diagnosis of eczema, or otherwise; (c) presence of eczema using parental completion of UKWP diagnostic criteria for AD; (d) parent reported child suffered from eczema in the last year; and (e) five

eczema severity levels on POEM (i) clear/Almost clear (score 0-2), (ii) mild (3-7), (iii) moderate (8-16), (iv) severe (17-24) and (v) very severe (25-28).[34] Significance was tested using t-tests for comparisons between (a) to (d), and a one-way analysis of variance (ANOVA) for (e). (a) was only possible at the 2-year timepoint as this was the only follow-up point with a blinded assessment of eczema. (b) to (e) were repeated for data at 2, 3, 4 and 5 years (with the exception of (d) which was not collected at 2 years) in order to see if construct validity improved with age. .

Convergent validity, which measures the strength of association between the measure of interest and other measures of the same (HRQL) or similar (e.g.disease severity) construct,[25] was tested using the Spearman rank test to see the degree to which scores on the CHU-9D were correlated with POEM scores at each of the four timepoints. Convergent validity is found if correlation coefficients lie in the moderate (0.41 to 0.6) or good (0.61 to 0.8) range or stronger.[25]

Responsiveness

Responsiveness explores whether the CHU-9D has the ability to detect meaningful or clinically important changes by examining whether the instrument can discriminate between those who change a lot compared to those who change little.[43,44] Previous research has suggested that an improvement/deterioration of ≥3 points on the POEM is likely to be clinically important.[36] Therefore, we examined whether the CHU-9D could discriminate between three groups: those whose POEM score, i) deteriorated ≥3 points, ii) change was not clinically important (-2.9 to 2.9 points), and iii) improved ≥3 points, where the change was estimated by taking the difference between the year-2 and year-5 POEM score. The mean change scores for the CHU-9D were estimated for these three groups, with the expectation

that the direction of the mean change in utility would be negative for group i) and positive for group iii) and the direction for group ii) being dependent on the number seeing small (non-clinically important) differences in either direction. Wilcoxon signed-rank tests were undertaken to detect any significant changes in scores within each group, together with effect size (ES) calculated as mean change divided by the standard deviation at baseline. The magnitude of utility change over time was assessed using standardised response mean (SRM), which is estimated by dividing the mean change in utility by the standard deviation of the change in utility, and enables comparison between studies.[45]

In all analyses a p-value of <0.05 was deemed significant and all analyses were undertaken using STATA 17.

Results

Participants

A total of 1,394 were randomised to the emollient group (n=693) or control group (n=701). The two groups were comparable in terms of characteristics, with the mean age of mothers at randomisation of 31.7/31.5 years respectively, 85%/86% of mothers of white ethnicity, and roughly 50% of infants male and female.[27] The trial found no evidence that daily emollient use during the first year of life prevents eczema in high-risk children as measured at 2 and 5 years of age.[27,28] Quantitatively there was no difference between the groups and as a consequence the analysis in this paper ignores treatment group, treating the sample as one group.

The characteristics of participants are reported in Table S1. This shows that for those completing the CHU-9D they were comparable to the full sample. The responsiveness sample is slightly less ethnically diverse.

Assessing the performance of the CHU-9D

Practicality

Of those returning the study questionnaire, the completion rate for the CHU-9D ranged from 88.0% at 2-years to 94.6% at 5-years (excluding completion at 3-years). Completion rates to the 3-year questionnaire (see Table 1) were lower due to the CHU-9D being inadvertently left out of the study questionnaire initially and as such do not fully reflect the ease of completing the CHU-9D in this age group. The results suggest that acceptability of the CHU-9D improves with age and that at all ages tested there was more than 5% missing data, although not considerably at 4-years and 5-years.[24] Missing data patterns across the four timepoints are available in supplementary material (Table S2 and S3).

The number of participants who reported being in full health (scoring 1 for each domain) ranged from 23.8% in year-3 to 32.8% at year-5. No participants at any timepoint reported being in the worst health state on the CHU-9D. The number of unique health states reported ranged from 149 (age-3) to 200 (age-4), although a small proportion of the potential 1,953,125 possible health states on the CHU-9D this perhaps reflects the study selected infants at high risk of atopic disease and excluded infants with serious health issues.

Table 2 provides detail about missing data for those who returned the study questionnaire. Whilst the percentage missing all nine questions on the CHU-9D remained fairly constant over the different ages (with the exception of year-3 due to an error omitting the CHU-9D initially), at 2-years the majority of additional missingness compared to other timepoints is explained by partial completion of the CHU-9D. For the majority of partial completers of the CHU-9D it was mostly one question (question 6 about schoolwork and homework) that

caused difficulties. Eighty-seven parent/carer respondents could not complete this item representing 90.6% of the 96 respondents who partially completed the CHU-9D at 2-years. The additional guidance given for question 6 is lengthy and it is the second half of the guidance that is particularly relevant for parents of children who do not go to nursery, preschool or kindergarten. As such parents may not take in all the guidance and think the question irrelevant for their child. Completion of question 6 improved with age because older children are more likely to be in childcare settings.

Respondents were given opportunity to provide comments on completing the CHU-9D, most chose not to or chose to use the space to explain why they had rated their child's health how they had in terms of their current conditions/symptoms. Only a couple of feedback points related to the content or appropriateness of the questions. One respondent commented; "*this questionnaire not age appropriate and very difficult to answer. For example, as typical for a 3-year-old, he has temper tantrums and so his emotions and feelings vary throughout the day*". Another respondent indicated that question 6 didn't seem relevant because: "*it is half term so no school work*" which supports concerns about question 6 discussed in the previous paragraph.

The mean utility score was 0.934 (sd 0.067) at 2-years, indicating participants were of good health, and remained similar at all four timepoints (see Table 1).

Validity

With respect to construct validity Table 3 shows that at 2-years those with eczema tended to have lower mean utility scores than those without eczema (0.923 versus 0.938), according to established diagnostic criteria. Using any parental report of a clinical diagnosis of eczema

also identified that the mean utility was lower for those with eczema compared to those without at all four timepoints. Although this was not statistically significant at 2-years it was at 3, 4 and 5-years but the mean differences were very small. Presence of eczema using parental completion of the UKWP diagnostic criteria for eczema identified mean utility which was lower for those with eczema at all four timepoints, although the difference between mean utility for those without eczema was not significant at 3 years. Likewise a parent report of a child suffering from eczema in the last year also found lower mean utility scores at 3, 4 and 5-years and this was statistically significant ($p<0.01$). Mean utility scores were found to be lower where eczema severity was higher according to the POEM severity levels and this was significant at all four timepoints. However, it can be seen that whilst CHU-9D scores could differentiate between those with and without eczema and between eczema disease severity, the mean differences are very small. The mean utility values elicited in the BEEP trial were higher than has been reported elsewhere in the literature for childhood eczema. For instance in the ECO parent/carer trial, testing an online self-management intervention the mean utility was around 0.863 at baseline with mean POEM score in the moderate range (score around 12.8).[46] Whilst in the CLOTHES trial, testing silk garments in the management of eczema, mean utility at baseline was around 0.834[47]. In both studies utility was measured using the CHU-9D but all participants had eczema. In the BEEP trial no evidence of a preventative effect was found for the intervention and the majority of children were reasonably healthy (70.5% did not develop eczema) which may explain the small differences found.

In terms of convergent validity, scores on the CHU-9D were correlated with scores on the POEM ($r = -0.116$ year-2, $r = -0.061$ year-3, $r = -0.172$ year-4 and $r = -0.167$ year-5) each $p<0.001$ but the size of the correlation coefficients suggests a weak relationship perhaps in

line with measuring the strength of association with disease severity rather than another measure of HRQL.

Responsiveness

Change scores were estimated for 801 respondents with complete data on both the POEM and CHU-9D at 2 and 5 years. At 5 years the POEM score deteriorated by ≥3 points for 103 (12.9%) respondents, for 580 (72.4%) respondents the POEM score did not change or improved/deteriorated by <3 points, and for 118 (14.7%) respondents the POEM score improved by ≥3 points. The mean change in utility (between year-2 and year-5) for each of these groups was -0.0198, 0.0041 and 0.0175 respectively on the CHU-9D) (see Table 4). The mean change in utility was in the direction expected but the size of the SRM was 0.21 for group i and iii (indicating small change and responsiveness), and 0.05 for group ii.

Discussion

The practicality of the CHU-9D appears to improve with age based on questionnaire completeness. For respondents who returned a study questionnaire at 2-years (n=1,212) CHU-9D questions were completely missing for 4.1% of respondents and partially missing for 7.9%. Amongst those partially completing the CHU-9D one question (question 6) caused most difficulty such that it may be possible to improve completion to levels similar to that of older age timepoints if the wording of the additional guidance for this question could be improved. It might, for instance help to move the final sentence in the additional guidance about ‘what to do if the child does not go to preschool/nursery/kindergarten’ to the beginning. At 2-years the main parent/carer was given the CHU-9D along with other questionnaires to

self-complete at a face-to-face visit and this may have encouraged completion at that timepoint.

In terms of construct validity, the CHU-9D was able to discriminate between those with and without eczema and between those with different levels of eczema severity, albeit the mean differences were small. In terms of convergent validity although the CHU-9D scores were significantly correlated to POEM scores the magnitude of the correlation was weak. This is likely because the POEM is a measure of disease severity. It would have been stronger to test the strength of association of the CHU-9D against other measures of HRQL. Recently the PedsQL has been used in this context, one example of this is an Australian sample comparing the proxy version of the CHU-9D to the Pediatric Quality of Life Inventory™ version 4.0 (PedsQL) completed by parents and carers of children aged 2-4 years which found a strong correlation.[48]

In terms of responsiveness, the SRM estimates of 0.21 are considered small. However, these estimates are within the range found for a general population sample where the CHU-9D was used in 2-4 year olds[48] and larger than those reported for the CHU-9D for children aged 2 to 4 years with eczema.[45]

These findings on practicality, validity and responsiveness of the proxy CHU-9D for children aged <5 years at high-risk of eczema contributes to an emerging body of research[48] seeking to assess whether the proxy CHU-9D can be used with children aged <5 years. This research agenda is important to ensure greater inclusion of younger child participants in cost-utility studies since the alternative might be to exclude those considered too young for the

instrument[30] or to have to conduct a separate analysis for the subset of younger children using a different utility instrument to the older children.

The strengths of this study are that the dataset is large and collected prospectively as part of a well conducted randomised controlled trial following the same children from birth to 5-years. However, there are limitations. Firstly, the study is limited to data collected alongside a trial which was designed and started recruiting over 10 years ago. The BEEP trial population was a select sample, although at high-risk of atopic disease broadly a healthy sample, limiting the range of HRQL values we might observe and the generalisability of results to other contexts. The participants did not have close contact with the trial team or clinical researchers as if the preventative intervention been found (cost) effective this would not have happened in practice. We did not therefore conduct in-depth qualitative work to understand how participants found answering the CHU-9D. Secondly, the inadvertent leaving out of the CHU-9D in a proportion of the first questionnaires sent out at year-3 means the data at this timepoint is not as fully reflective as the other timepoints. This limits the conclusions that can be reached about practicality at this age. Thirdly, around 30% of respondents did not complete any of the BEEP trial study questionnaire at each of the 3-, 4-, and 5-year timepoints (compared to 13% at 2-years), due to loss to follow-up or the withdrawal of consent, which means there is potential for bias in our findings as they focus on those that chose to complete the study questionnaire. As a consequence, it is unknown whether the results would generalise to the part of the study population with missing data.[49]

Another potential limitation is that because parents were proxies completing the questionnaires over multiple timepoints for their child, it is impossible to disentangle whether completion rates were slightly better at older age groups due to the questionnaire being more

appropriate for older aged children and/or whether there may have been a learning curve effect such that completion rates improved through increased familiarity with the questionnaire. Focus-group research has shown parent proxies found the CHU-9D offered a more comprehensive assessment of HRQL than the EQ-5D-Y. Although they also felt that the higher number of questions might increase difficulty for proxy completers particularly around aspects on emotional well-being (although unclear if the CHU-9D version with additional guidance for younger children was used).[50] A review of literature comparing self and proxy reported utility in childhood also found inter-rater agreement was lower for more subjective aspects of health for other preference-based measures for which this evidence exists.[51]

Whilst use of the CHU-9D in young children shows potential in terms of practicality, it is unclear whether the small differences and small responsiveness observed when looking at validity and responsiveness is due to the relatively healthy sample with small number of participants in the different disease severity groups (other than (almost)/clear) or due to the use of the CHU-9D in this young age group. Therefore, further research is needed to validate the CHU-9D in <5's including children for whom HRQL is expected to vary[44] and to examine other measurement properties, such as reliability. Qualitative research with individual participants to explore completion of the CHU-9D in preschool children would be valuable, as this could broaden consideration to whether we should be using the CHU-9D in this age group. The appropriateness of the value set ought to also be considered when applying it to young children given the respondents in the valuation study were asked to imagine being a 10-year-old child in valuation tasks.[37]

Conclusions

Our work has assessed the proxy CHU-9D in terms of practicality, validity, and responsiveness for children aged <5 years at high-risk of eczema. It contributes to an emerging body of research which seeks to assess the psychometric properties of the proxy CHU-9D with children aged <5 years. A small proportion found the 'School Work/Homework' question difficult particularly at the lower age range despite additional guidance, such that it might be useful to assess changes to the wording of the guidance for this question to improve this. Further research is needed using datasets from other studies for other conditions in pre-school aged children and to examine other measurement properties of the CHU-9D in this age group.

References

1. Bailey C, Howell M, Raghunandan R, Salisbury A, Chen G, Coast J, et al. Preference Elicitation Techniques Used in Valuing Children's Health-Related Quality-of-Life: A Systematic Review. *Pharmacoeconomics.* 2022;40(7):663-698. doi: 10.1007/s40273-022-01149-3.
2. Petrou S. Methodological challenges surrounding QALY estimation for paediatric economic evaluation. *Cost Eff Resour Alloc.* 2022;20(1):10. Published 2022 Mar 3. doi:10.1186/s12962-022-00345-4.
3. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ.* 2003;12(8):697-702. doi: 10.1002/hec.775.
4. Chen G, Ratcliffe J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics.* 2015;33(10):1013-1028. doi: 10.1007/s40273-015-0286-7.

5. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics.* 2005;115(5):e600-e614. doi:10.1542/peds.2004-2127Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *Pharmacoeconomics.* 2011 Aug;29(8):641-52. doi: 10.2165/11591570-000000000-00000.
6. Apajasalo M, Sintonen H, Holmberg C, Sinkkonen J, Aalberg V, Pihko H, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). *Qual Life Res.* 1996;5(2):205-211. doi:10.1007/BF00434742.
7. Apajasalo M, Rautonen J, Holmberg C, Sinkkonen J, Aalberg V, Pihko H, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res.* 1996;5(6):532-538. doi:10.1007/BF00439227.
8. Beusterien KM, Yeung JE, Pang F, Brazier J. Development of the multi-attribute Adolescent Health Utility Measure (AHUM). *Health Qual Life Outcomes.* 2012;10:102. Published 2012 Aug 28. doi:10.1186/1477-7525-10-102.
9. Richardson, J., Atherton Day, N., Peacock, S. and Iezzi, A. Measurement of the Quality of Life for Economic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. *Aust Econ Rev.* 2004;37(1):62-88. https://doi.org/10.1111/j.1467-8462.2004.00308.x.
10. Kang E. Validity of Child Health-6 Dimension(Ch-6d) for Adolescents. *Value Health*. 2016;19(7):A854. Published 2016 Nov. https://doi.org/10.1016/j.jval.2016.08.458.
11. Saigal S, Rosenbaum P, Stoskopf B, Hoult L, Furlong W, Feeny D, et al. Development, reliability and validity of a new measure of overall health for pre-school children. *Qual Life Res.* 2005;14(1):243-257. doi:10.1007/s11136-004-4228-7.

12. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res.* 2009;18(8):1105-1113. doi:10.1007/s11136-009-9524-9.

13. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qual Health Res.* 2010;20(3):340-351. doi:10.1177/1049732309358328.

14. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res.* 2010;19(6):875-886. doi:10.1007/s11136-010-9648-y

15. Kreimeier S, Åström M, Burström K, Egmar AC, Gusi N, Herdman M, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res.* 2019;28(7):1951-1961. doi:10.1007/s11136-019-02115-x

16. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care.* 1996;34(7):702-722. doi:10.1097/00005650-199607000-00004.

17. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med.* 2001;33(5):375-384. doi:10.3109/07853890109002092.

18. Jabrayilov R, Vermeulen KM, Detzel P, Dainelli L, van Asselt ADI, Krabbe PFM. Valuing Health Status in the First Year of Life: The Infant Health-Related Quality of Life Instrument. *Value Health.* 2019;22(6):721-727. doi:10.1016/j.jval.2018.12.009.

19. Jabrayilov R, van Asselt ADI, Vermeulen KM, et al. A descriptive system for the Infant health-related Quality of life Instrument (IQI): Measuring health with a mobile app. *PLoS One.* 2018;13(8):e0203276. Published 2018 Aug 31. doi:10.1371/journal.pone.0203276.

20. Krabbe PFM, Jabrayilov R, Detzel P, Dainelli L, Vermeulen KM, van Asselt ADI. A two-step procedure to generate utilities for the Infant health-related Quality of life Instrument (IQI). *PLoS One.* 2020;15(4):e0230852. Published 2020 Apr 3. doi:10.1371/journal.pone.0230852.

21. Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. *J Patient Rep Outcomes.* 2020;4(1):94. Published 2020 Nov 9. doi:10.1186/s41687-020-00251-4.

22. Verstraete J, Ramma L, Jelsma J. Item generation for a proxy health related quality of life measure in very young children. *Health Qual Life Outcomes*. 2020;18(1):11. Published 2020 Jan 14. doi:10.1186/s12955-020-1271-1.

23. NICE. *NICE health technology evaluations: the manual Process and methods [PMG36]. Section 4.3.13-4.3.15*. Published January 31, 2022. Accessed Sept 6, 2023. https://www.nice.org.uk/process/pmg36/chapter/economic-evaluation.

24. Kwon J, Smith S, Raghunandan R, Howell M, Huynh E, Kim S, et al. Systematic Review of the Psychometric Performance of Generic Childhood Multi-attribute Utility Instruments. *Appl Health Econ Health Policy.* 2023;21(4):559-584. doi:10.1007/s40258-023-00806-8.

25. Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A Review of the Psychometric Performance of Selected Child and Adolescent Preference-Based Measures Used to Produce Utilities for Child and Adolescent Health. *Value Health*. 2021;24(3):443-460. doi:10.1016/j.jval.2020.09.012.

26. Chalmers JR, Haines RH, Mitchell EJ, Thomas KS, Brown SJ, Ridd M, et al. Effectiveness and cost-effectiveness of daily all-over-body application of emollient during the first year of life for preventing atopic eczema in high-risk children (The BEEP trial):

protocol for a randomised controlled trial. *Trials.* 2017;18(1):343. Published 2017 Jul 21. doi:10.1186/s13063-017-2031-3.

27. Chalmers JR, Haines RH, Bradshaw LE, Montgomery AA, Thomas KS, Brown SJ, et al. Daily emollient during infancy for prevention of eczema: the BEEP randomised controlled trial. *Lancet.* 2020;395(10228):962-972. doi:10.1016/S0140-6736(19)32984-8.
28. Bradshaw LE, Wyatt LA, Brown SJ, Haines RH, Montgomery AA, Perkin MR, et al. Emollients for prevention of atopic dermatitis: 5-year findings from the BEEP randomized trial. *Allergy.* 2023;78(4):995-1006. doi:10.1111/all.15555.
29. Sach TH, Lartey ST, Davies C, Chalmers JR, Haines RH, Bradshaw LE, et al. Emollients for preventing atopic eczema: Cost-effectiveness analysis of the BEEP trial. *Clin Exp Allergy.* 2023;53(10):1011-1019. doi:10.1111/cea.14381.
30. Thomas KS, Bradshaw LE, Sach TH, et al. Randomised controlled trial of silk therapeutic garments for the management of atopic eczema in children: the CLOTHES trial. *Health Technol Assess.* 2017;21(16):1-260. doi:10.3310/hta21160
31. Weerasuriya SR, Hettiarachchi RM, Kularatna S, et al. Comparison of the Early Childhood Oral Health Impact Scale (ECOHIS-4D) and Child Health Utility Index (CHU-9D) in children with oral diseases. Community Dent Oral Epidemiol. 2024;52(2):224-231. doi:10.1111/cdoe.12917.
32. Frizelle P, Mckean C, O'Shea A, Horgan A, Murphy A. Economic evaluation of the Happy Talk pilot effectiveness trial: A targeted selective speech, language and communication intervention for children from areas of social disadvantage. Int J Speech Lang Pathol. 2022;24(2):200-211. doi:10.1080/17549507.2021.1975815
33. Williams HC, Burney PG, Pembroke AC, Hay RJ. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol.* 1994;131(3):406-416. doi:10.1111/j.1365-2133.1994.tb08532.x.

34. Charman CR, Venn AJ, Ravenscroft JC, Williams HC. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. *Br J Dermatol*. 2013;169(6):1326-1332. doi:10.1111/bjd.12590.

35. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective [published correction appears in *Arch Dermatol*. 2005 Mar;141(3):381]. *Arch Dermatol.* 2004;140(12):1513-1519.

36. Howells L, Ratib S, Chalmers JR, Bradshaw L, Thomas KS; CLOTHES trial team. How should minimally important change scores for the Patient-Oriented Eczema Measure be interpreted? A validation using varied methods. *Br J Dermatol*. 2018;178(5):1135-1142. doi:10.1111/bjd.16367.

37. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics.* 2012;30(8):729-747. doi:10.2165/11599120-000000000-00000.

38. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual Life Res*. 2012;21(4):717-725. doi:10.1007/s11136-011-9971-y

39. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. *Value Health*. 2012;15(8):1092-1099. doi:10.1016/j.jval.2012.07.011.

40. Chen G, Flynn T, Stevens K, Brazier J, Huynh E, Sawyer M, et al. Assessing the Health-Related Quality of Life of Australian Adolescents: An Empirical Comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. *Value Health*. 2015;18(4):432-438. doi:10.1016/j.jval.2015.02.014.

41. Canaway AG, Frew EJ. Measuring preference-based quality of life in children aged 6-7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y--the WAVES pilot study. *Qual Life Res*. 2013;22(1):173-183. doi:10.1007/s11136-012-0119-5.
42. Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P. EQ-5D versus SF-6D in an older, chronically Ill patient group. *Appl Health Econ Health Policy*. 2004;3(2):91-102. doi:10.2165/00148365-200403020-00005.
43. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide To Their Development And Use*. Oxford; 2003.
44. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care.* 2000;38(9 Suppl):II84-II90.
45. Xiong X, Dalziel K, Huang L, Mulhern B, Carvalho N. How do common conditions impact health-related quality of life for children? Providing guidance for validating pediatric preference-based measures. *Health Qual Life Outcomes*. 2023;21(1):8. Published 2023 Jan 25. doi:10.1186/s12955-023-02091-4
46. Sach TH, Onoja M, Clarke H, et al. Cost-effectiveness of two online interventions supporting self-care for eczema for parents/carers and young people. *Eur J Health Econ.* Published online January 9, 2024. doi:10.1007/s10198-023-01649-9
47. Thomas KS, Bradshaw LE, Sach TH, et al. Silk garments plus standard care compared with standard care for treating eczema in children: A randomised, controlled, observer-blind, pragmatic trial (CLOTHES Trial). *PLoS Med*. 2017;14(4):e1002280. Published 2017 Apr 11. doi:10.1371/journal.pmed.1002280
48. Xiong X, Carvalho N, Huang L, et al. Psychometric Properties of Child Health Utility 9D (CHU9D) Proxy Version Administered to Parents and Caregivers of Children Aged 2-4 Years Compared with Pediatric Quality of Life Inventory™ (PedsQL).

Pharmacoeconomics. Published online January 27, 2024. doi:10.1007/s40273-024-01355-1.

49. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011.
50. Wolstenholme JL, Bargo D, Wang K, Harnden A, Räisänen U, Abel L. Preference-based measures to obtain health state utility values for use in economic evaluations with child-based populations: a review and UK-based focus group assessment of patient and parent choices. *Qual Life Res*. 2018;27(7):1769-1780. doi:10.1007/s11136-018-1831-6.
51. Khadka J, Kwon J, Petrou S, Lancsar E, Ratcliffe J. Mind the (inter-rater) gap. An investigation of self-reported versus proxy-reported assessments in the derivation of childhood utility values for economic evaluation: A systematic review. *Soc Sci Med*. 2019;240:112543. doi:10.1016/j.socscimed.2019.112543

Table 1: Questionnaire completion and summary scores for the CHU-9D and POEM at 2, 3, 4 and 5 years

	2-year	**3-year**	**4-year**	**5-year**
Study questionnaires returned (n=1394)	1212 (87%)	981 (70%)	990 (71%)	976 (70%)
Number (%) completed for CHU-9D^	1066 76.5%/88.0%	685* 49.1%/69.8%	925 66.4%/93.4%	923 66.2%/94.6%
Number completed for POEM^	1171 84.5%/97.2%	946 67.9%/96.4%	958 68.7%/96.8%	954 68.4%/97.7%
CHU-9D				
Number of respondents	1,066	685	925	923
Mean score	0.934	0.926	0.929	0.937
St Dev	0.067	0.069	0.074	0.068
Median score	0.952	0.931	0.952	0.952
25-75 percentile	0.903 to 1	0.894 to 0.979	0.900 to 1	0.900 to 1
Range	0.479 to 1	0.443 to 1	0.518 to 1	0.533 to 1
Skewness	-1.510	-1.617	-1.553	-1.472
Number (%) in Full Health	304 (28.5)	163 (23.8)	262 (28.3)	303 (32.8)
Number (%) in the worst health state possible	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Number of unique health states reported	188 (0.18)	149 (0.22)	200 (0.22)	179 (0.19)

POEM				
Number of respondents	1171	946	958	954
Mean score	1.887	1.580	1.646	1.585
St Dev	3.948	3.441	3.602	3.405
Poem severity				
(Almost)/Clear	917 (78.31)	752 (79.49)	766 (79.96)	758 (79.45)
Mild	145 (12.38)	127 (13.42)	115 (12.00)	115 (12.05)
Moderate	92 (7.86)	58 (6.13)	66 (6.89)	77 (8.07)
Severe	16 (1.37)	9 (0.95)	10 (1.04)	3 (0.31)
Very severe	1 (0.09)	0 (0.00)	1 (0.10)	1 (0.10)
Median score	0.000	0.000	0.000	0.000
25-75 percentile	0 to 2	0 to 1	0 to 1	0 to 1
Range	0 to 26	0 to 24	0 to 28	0 to 27
Skewness	2.689	2.909	2.890	2.764

^First % is completion as a percentage of the total 1394 sample and the second % is the percentage completing from those completing the overall questionnaire at the time point.

*CHU-9D completion rates at the 3-year timepoint were lower due to the CHU-9D being inadvertently left out of the questionnaire initially.

CHU-9D indicates Child Health Utility – Nine Dimensions and POEM indicates Patient Oriented Eczema Measure.

Table 2: Nature of missingness for CHU-9D for respondents completing the study question at each timepoint

	Year 2	**Year 3**	**Year 4**	**Year 5**
Completely missing	50 (4.1%)	256 (26.1%)*	38 (3.8%)	42 (4.3%)
Partially missing	96 (7.9%)	40 (4.1%)	27 (2.7%)	11 (1.1%)
Number of Questions missing per respondent:				
1	90	37	21	10
2	3	1	1	0
3	3	0	0	0
4	0	0	2	0
5	0	0	2	1
6	0	2	1	0
Number of respondents missing each question:				

1 (worried)	2	2	5	2
2 (sad)	4	2	6	1
3 (pain)	4	3	5	1
4 (tired)	2	2	5	2
5 (annoyed)	1	3	5	3
6 (work)	87	38	13	2
7 (sleep)	1	1	2	0
8 (daily routine)	0	0	4	0
9 (activities)	4	0	2	0

CHU-9D indicates Child Health Utility – Nine Dimensions

*the year 3 study questionnaire inadvertently left out the CHU-9D when it was first distributed until this was noticed and it added in. Therefore, this number largely reflects this error.

Table 3: Construct Validity: Mean (standard deviation, sd) CHU-9D utility scores for presence of eczema and each eczema severity levels

	CHU-9D score year 2	CHU-9D score year 3	CHU-9D score year 4	CHU-9D score year 5
(a) Diagnosis of eczema according to established diagnostic criteria over the past year: Mean utility (SD) (n=number of participants)				
No Eczema	0.938 (0.065) (n=812)			
Eczema	0.923 (0.073) (n=253)*			
(b) Any parental report of a clinical diagnosis of eczema in the previous year: Utility (SD) (number of participants)#				
No Eczema	0.937 (0.064) (n=594)	0.928 (0.068) (n=587)	0.932 (0.071) (n=830)	0.939 (0.064) (n=838)
Eczema	0.932 (0.072) (n=472)	0.908 (0.077) (n=85)*	0.905 (0.097) (n=93) ‡	0.916 (0.094) (n=80)*
(c) Presence of eczema using parental completion of UKWP diagnostic criteria for eczema: Utility (SD) (number of participants)				
No Eczema	0.938 (0.064) (n=730)	0.929 (0.063)(n=527)	0.933 (0.69)(n=671)	0.944 (0.063)(n=654)

Eczema	0.925 (0.074) (n=336)*	0.917 (0.085) (n=155)	0.916 (0.085)(n=242)*	0.924 (0.075)(n=256)‡
(d) Parent reported child suffered from eczema in the last year: Utility (SD) (number of participants)				
No Eczema		0.931 (0.063)(n=439)	0.935 (0.068)(n=612)	0.941 (0.064)(n=601)
Eczema		0.917 (0.079)(n=236)*	0.917 (0.085)(n=297)*	0.927 (0.074)(n=306)*
(e) POEM severity: Utility (SD) (number of participants)				
(Almost)/Clear	0.939 (0.065)(n=833)	0.929 (0.065)(n=532)	0.935 (0.068)(n=717)	0.941 (0.066)(n=718)
Mild	0.927 (0.063)(n=136)	0.928 (0.069)(n=91)	0.921 (0.085)(n=110)	0.926 (0.060)(n=109)
Moderate	0.916 (0.076)(n=80)	0.897 (0.073)(n=39)	0.891 (0.086)(n=65)	0.923 (0.075)(n=76)
Severe	0.860 (0.089)(n=15)	0.872 (0.126)(n=4)	0.827 (0.126)(n=10)	0.792 (0.120)(n=3)
Very severe	0.679 (0.000)(n=1) ‡	n/a (n=0) *	n/a (n=0) ‡	0.933 (0.000) (n=1) ^

Results of the T-Tests and ANOVA are also noted (‡ $p < 0.001$, * $p<0.01$, ^ $p<0.05$). CHU-9D indicates Child Health Utility – Nine Dimensions and POEM indicates Patient Oriented Eczema Measure. # For (b) any parental report of a clinical diagnosis of eczema was asked since birth at 2 years and in the previous year at 3, 4 and 5 years.

Table 4 Responsiveness of the CHU-9D between 2 years and 5 years for POEM change

	N	2-year CHU-9D (mean)	5-year CHU-9D (mean)	Mean change	SD at 2 years	SD of change	ES	SRM [b]	P-Value [a]
POEM declined ≥ -3*	103	0.9305	0.9107	-0.0198	0.0632	0.093	-0.313	0.21*	**0.049**
POEM improved by <3 points, did not improve or declined by <3 points	580	0.9378	0.9419	0.0041	0.065	0.079	0.063	0.05	0.227
POEM Improved by ≥ 3*	118	0.9150	0.9325	0.0175	0.075	0.083	0.233	0.21*	**0.035**

Note: ES = effect size = (mean change/SD at baseline)

a Wilcoxon signed-rank test conducted and p-values in **bold** are statistically significant at the 5% level

b SRM = standardized response mean = (mean change/SD of change). If SRM=0.2 to 0.50 equals small, 0.50 to 0.80 equals moderate and 0.80 and above equals large. *Small change, small responsiveness