Developing a patient-reported outcome measure of eczema control

Data analysis plan

Version 1.2
Additions made following expert panel discussion on 11th December 2018 (following impact analysis, prior to regression analysis) are highlighted in green.
Changes made on 4th December 2018 (V1.1) (prior to all analysis) are highlighted in yellow.

~~**Version 1.1**~~
~~**Amended 4.12.2018**~~
~~**(all changes highlighted)**~~

~~**Version 1.0**~~
~~**5.11.2018**~~

Contents

A. Objectives:

STAGE 1 *– To reduce the number of items included in the final RECAP measure.*

STAGE 2 *– To assess measurement properties of the final RECAP measure.*

(Iterative process: May return to stage 1 if required.)

B. Variables in dataset

Age
Sex
Ethnicity
RECAP items (15 in total)
Frequency of RECAP items in past year
Importance of RECAP items
Bother scale
Global severity
POEM items (7 in total)

C. Data cleaning

Aim: *to ensure data is in useable format.*

ACTIONS:

- Data will be exported from the survey software to STATA SE 15.
- All variables will be named and labelled and a data codebook will be set up.
- Score each RECAP item from 0 – 4. Will use stakeholder engagement to assess if higher scores should indicate more control or less control as this will affect the usability and interpretability of the final questionnaire.
- Score frequency of RECAP items in past year as 1 for yes and 0 for no.
  Use data entry assumption that if have answered no to this question, but had answered something other than 'no days/not at all' to RECAP items to score as 1 (yes) as would suggest a misunderstanding.
- Score importance of RECAP items from 1 to 5 (5 indicates the most important).
- Score POEM using scoring rules assigned by the developers. Also categorise into severity bandings that have been developed and published[1].
- Score bother from 0 – 10.
- Score global severity from 0 – 4.
- Explore frequency of scores for each variable to assess for data errors. Handle errors appropriately.
- Assess missing data in all variables and handle appropriately.
- If percentage of missing data for each item from RECAP appear to be particularly problematic (e.g. over 10% of participants who have filled out this page have skipped the question), we will consider why this may be and if this item should still be included in further analyses. We will consider if missing data appears to be random or non-random.

D. Descriptive statistics

Aim: *to explore how items have been answered across the sample.*

ACTIONS:

- Assess mean, SD and histogram of each RECAP item.
- We will assess the distribution of the scores for each individual item to assess if all response options are being used by the participants.
- We will consider if it would be beneficial to adapt the response options used (i.e. to potentially enhance the reliability of the scores).
- We will also assess the mean SD, and histogram for POEM overall scores, global severity scores and bother scores.

STAGE 1: Development of RECAP

E. Impact analysis

Aim: *to exclude items that are not considered relevant (i.e. are of low occurrence and low importance) across the sample and across different groups of people within the sample.*

ACTIONS:

- Impact analysis will be conducted for all participants, as well as sub-groups by age (under 5 years, 5-15 years, and 16+ years).
- Calculate the proportion of people who responded as YES in the Frequency of RECAP items in past year items.
- Calculate the mean score for the importance of RECAP items
- "Impact" = frequency x importance (min impact score = 1, max impact score = 5)
- Rank items in order of impact score for all participants and each sub-group analysis.
- Predefine items that score < 2 (maximum score of 5) on impact analysis not relevant and will consider excluding from further analysis (maximum score of 5) as a similar scale development process has used[2]. However, will be open to further discussion with expert panel and consideration of content validity.

Update from 11th Dec 2018: The expert panel decided to remove items that were < 2 on the impact analysis.

F. Regression analysis

Aim: To remove redundant items (i.e. items that do not add much further understanding to global control).

- Analysis to be conducted on full dataset.
- Sample size will be determined as at least 10 cases per predictor variable.
- Missing data will be handled by using complete case (listwise) analysis. We will do a sensitivity analysis by including those with missing data (pairwise). Multiple imputation is not considered appropriate in this context as we do not have assumptions for how the items should behave given that we are in the developmental phase.
- Test assumptions of multivariate regression analysis: multicollinearity, outliers, normality, linearity, homoscedasticity, independence of residuals
- Model building using RECAP items as independent variables and bother scale as the dependent variable.
- ~~Use a theory informed step-wise approach to model building:~~
  - ~~a. Add each item as individual predictor variables in univariate regression analysis~~

~~b. Add in all items that are significant predictors in multivariate regression analysis~~
~~c. Remove items that are no longer significant~~
~~d. Add in non-significant items one at a time to see if add anything to the overall model.~~

- Variable selection for final model based on relevance, predictive ability and feasibility:
  1. Expert panel will be asked to consider if any items should be forced to remain in the model regardless of p-value. Update from 11th December 2018: expert panel made the decision to not force any items to remain in the model. Decision to not include one item (item 1 – Thinking of all the treatments you have used over the last week, how well have your treatments been able to manage your eczema? – This was due to applicability, interpretability and conceptual reasons.)
  2. Use backward elimination technique to decide for the remaining variables which should be kept within the model. A stopping criterion of p=0.157, which is recommended for use in STATA as a proxy for Akaike's information criteria (AIC) and as the appropriate criterion for our sample size[3, 4]. AIC is a relative measure of information using the likelihood function and includes a penalty for the number of parameter estimates.
  3. During expert panel discussions on the 11th December 2018, prior to regression analysis being conducted, the panel decided that two models will be generated. One will enter all remaining items following the decisions made by the expert panel on 11th December 2018 and one will exclude the "global" items (3 Over the last week, how acceptable has your/your child's eczema been to you? and 15 Over the last week, how has your/your child's eczema been?) The expert panel will be able to review both models to decide which should be the final model to determine the items in the questionnaire.
  4. If the final model contains a large number of variables, (~~cut-off to be decided by expert panel meeting on 11th December 2018),~~those which contributed to the model the least will be removed. Number of items considered to be feasible was discussed at the expert panel meeting on 11th December 2018, but it was felt a final decision would depend on the results of the regression analysis so could not be determined at this stage.

- If resources allow, we will consider conducting an internal validation of the final model using bootstrapping methods.
- Based on the output from the regression analysis and discussion with the expert panel, we will make an informed decision about how to score the final questionnaire (i.e. whether to use weighted items or not).
- ~~We also plan to test the predictive ability of the model built by age sub-groups. These will likely be the age groups of under 5 years, 5-15 years, and 16+ years, but this will be subject to adequate sample size.~~

G. Expert panel input

Aim: To ensure face validity of the RECAP questionnaire and acceptability of the RECAP questionnaire to key stakeholders.

- We will engage with the expert panel to discuss the results from the analysis and make final decisions about the items to be included in the RECAP questionnaire.
- The feasibility of the final items considered for inclusion in the questionnaire will be of primary concern, and evidence from the literature, the think aloud interviews and the expert

panel's knowledge will guide decisions about feasibility. We will ask the expert panel a priori to the regression analyses being conducted what a feasible number of items would include. Update from 11th December 2018: Number of items considered to be feasible was discussed at the expert panel meeting, but it was felt a final decision would depend on the results of the regression analysis so could not be determined at this stage.

- ~~Conceptual issues will be considered (i.e. content validity, acceptability and comprehensiveness of the scale) and items may be added in or removed on this basis.~~

STAGE 1 OUTPUT: Chosen final set to include in RECAP.

STAGE 2: Initial assessment of measurement properties of RECAP

Aim: To assess how the RECAP items chosen for inclusion in final set function as an overall scale.

H. Descriptive statistics

- Descriptive statistics (mean, SD) and histogram. This will be conducted for sub groups of age groups of under 5 years, 5-15 years, and 16+ years.
- Assess the distribution of the total score of the set of items chosen for inclusion.

I. Floor and ceiling effects

- Consider if floor or ceiling effects occur (defined as >15% of participants achieving highest or lowest possible score). If this occurs, we will decide if we want to make further adjustments to the final questionnaire or not. ~~If this occurs this may be improved by including more items that will help distinguish individuals and/or alter descriptors in response options (the latter is the less desirable option as not been through cognitive interviewing testing).~~

INITIAL TESTING OF MEASUREMENT PROPERTIES

Aim: To assess if relationships with other instruments / known groups are consistent with assumptions about underlying construct of interest. This will be an initial testing using the same sample used for the development of the scale rather than an external validation study.

J. Convergent validity

Aim: To test hypotheses about level of convergent relationship with other instruments

ACTIONS:

Hypotheses to be tested:

- POEM and RECAP final set of items will be positively correlated by at least 0.3, interpreted as a moderate correlation[5]. It is hypothesised that POEM captures a construct that is a sub-set of the construct of interest for the RECAP scale, therefore they are expected to be correlated in the same direction but not necessarily strongly correlated.

- Will use Pearson's correlation coefficient if all the assumptions are met. If they are not, we will use Spearman's correlation coefficient.

K. Discriminative validity (known groups analysis)

Aim: To test hypotheses about ability to discriminate between known groups based on other instruments.

To test hypotheses:

- Looking at participants grouped according to the global severity categories, it is expected that those categorised with more severe eczema will have a higher mean score than those on categorised with less severe eczema.
- Looking at participants grouped according to POEM severity categories, it is expected that those categorised with more severe eczema will have a higher mean score than those on categorised with less severe eczema.

STAGE 2 Output: Testing of the adequacy of the final set of items chosen to be included in RECAP.

Note on the iterative process of the development process:
If there are concerns following initial testing, can return to previous steps to revise set of items to consider in testing.

References

1. Charman, C., et al., *Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods.* British Journal of Dermatology, 2013. **169**(6): p. 1326-1332.
2. Weller, K., et al., *Development and validation of the Urticaria Control Test: A patient-reported outcome instrument for assessing urticaria control.* Journal of Allergy and Clinical Immunology, 2014. **133**(5): p. 1365-1372.e6.
3. Sauerbrei, W., *The use of resampling methods to simplify regression models in medical statistics.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 1999. **48**(3): p. 313-329.
4. Heinze, G., C. Wallisch, and D. Dunkler, *Variable selection–A review and recommendations for the practicing statistician.* Biometrical Journal, 2018. **60**(3): p. 431-449.
5. Cohen, J., *Statistical power analysis for the behavioral sciences. 2nd*. 1988, Hillsdale, NJ: erlbaum.

SIGNATORIES TO DATA ANALYSIS PLAN:

Chief Investigator and lead PhD supervisor: _____

Signature: _____

Date: _____

PhD student: _____

Signature: _____

Date: _____