The Patient - Patient-Centered Outcomes Research (2021) 14:101–117
https://doi.org/10.1007/s40271-020-00462-3

ORIGINAL RESEARCH ARTICLE

Testing and Psychometric Validation of a Pediatric Instrument to Self-Assess Symptoms of the Common Cold

Rob Arbuckle[1] · Patricia Halstead[2] · Chris Marshall[1] · Brenda Zimmerman[2] · Kate Bolton[1] · Antoine Regnault[3] · Cathy Gelotte[2]

Accepted: 19 September 2020 / Published online: 11 November 2020

Abstract

Objectives Given the lack of validated patient-reported outcomes (PRO) instruments assessing cold symptoms, a new pediatric PRO instrument was developed to assess multiple cold symptoms: the Child Cold Symptom Questionnaire (CCSQ). The objective of this research was to evaluate the measurement properties of the CCSQ.

Methods This observational study involved daily completion of the self-report CCSQ by children aged 6–11 years in their home for 7 days. These data were used to develop a scoring algorithm and item-scale structure and evaluate the psychometric properties of the resulting scores. Analyses included evaluation of item and dimensionality performance (item response distributions and confirmatory factor analysis) and assessment of test–retest reliability in stable patients, construct validity (convergent and known groups validity), and preliminary responsiveness. Qualitative exit interviews in a subgroup of the children with colds and their parents were conducted.

Results More than 90% of children had no missing data during the testing period, reflecting an excellent completion rate. For most items, responses were distributed across the options, with approximately normal distributions. Test–retest reliability was adequate, with intra-class correlation coefficients ranging from 0.63 to 0.83. A logical pattern of correlations with the validated Strep-PRO instrument provided evidence supporting convergent validity. Single- and multi-item symptom scores distinguished between children who differed in their cold severity based on global ratings, providing evidence of known groups validity. Preliminary evidence indicates the CCSQ is responsive to changes over time.

Conclusions The findings demonstrate that the CCSQ items and multi-item scores provide valid and reliable patient-reported measures of cold symptoms in children aged 6–11 years. They provide strong evidence supporting the validity of these items and multi-item scores for inclusion as endpoints in clinical trials to evaluate the efficacy of cold medicines.

Key Points for Decision Makers

The psychometric validation of a child self-report measure of common cold symptoms in children aged 6–11 years is described.

The single-item and multi-item scores are valid and reliable in children aged 6–11 years.

The measure is appropriate for assessing cold symptoms in clinical trials.

Chris Marshall and Kate Bolton were working for Adelphi Values (known at the time as Mapi Values) at the time the research was conducted. Patricia Halstead, Brenda Zimmerman and Cathy Gelotte were working for McNeil Consumer Healthcare at the time the research was conducted. Antoine Regnault is affiliated to Modus Outcomes and was working for Mapi Values at the time the work was conducted.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40271-020-00462-3) contains supplementary material, which is available to authorized users.

✉ Rob Arbuckle
rob.arbuckle@adelphivalues.com

Extended author information available on the last page of the article

1 Introduction

The common cold, an upper respiratory tract infection (URI), is the most common acute illness in the United States (US) among both pediatric and adult populations, leading to more doctor visits and loss of work than any other illness [1, 2]. Over-the-counter (OTC) cough and cold medications are marketed widely for relief of common cold symptoms and are "Generally Recognized as Safe and Effective" (GRASE) under the OTC monograph system (Title 21 of the *Code of Federal Regulations*, Part 341). While data demonstrate the efficacy of these products in adults, they are inconclusive in children, likely due in part to difficulties evaluating common cold symptom severity and changes over time in a young age group [3].

Patient-reported outcomes (PRO) measures of symptoms and associated quality of life for upper respiratory infections in adults have been developed, such as the InFLUenza Patient-Reported Outcome (FLU-PRO©) [4–7], the Influenza Intensity and Impact Questionnaire (R-iiQ™/FluiiQ™) [8], and the Wisconsin Upper Respiratory Symptom Survey (WURSS-21). Similarly, clinician- and parent/caregiver-report measures exist for use in infants, such as the ReSVinet [9, 10] and the Pediatric RSV Severity and Outcome Rating Scales (PRESORS) [11]. The only instrument developed specifically for use in 6- to 11-year-olds is the Canadian Acute Respiratory Illness and Flu Scale (CARDIFS), which is a parent/caregiver report. Given the lack of existing *self-report* instruments assessing cold symptoms in children aged 6–11 years that meet the Food and Drug Administration's (FDA's) PRO Guidance for Industry [12], a new pediatric PRO instrument was developed to assess multiple cold symptoms (nasal congestion, runny nose, pain [including headache, sinus pain/pressure, body aches/muscle aches, and sore throat], chest congestion, and cough): the Child Cold Symptom Questionnaire (CCSQ). The aim was to create a PRO instrument that could support symptom-specific endpoints in clinical trials to assess the efficacy of cold medicines in children aged 6–11 years. The PRO development was supported by extensive qualitative research with both children and parents, including concept elicitation and cognitive debriefing activities to test alternative wordings of various symptom items and recall periods, and provide evidence of content validity [13]. Following that qualitative research, this article reports results of an observational study to develop scoring, with possible item deletion, and to evaluate the psychometric measurement properties for the CCSQ.

2 Methods

2.1 Study Design

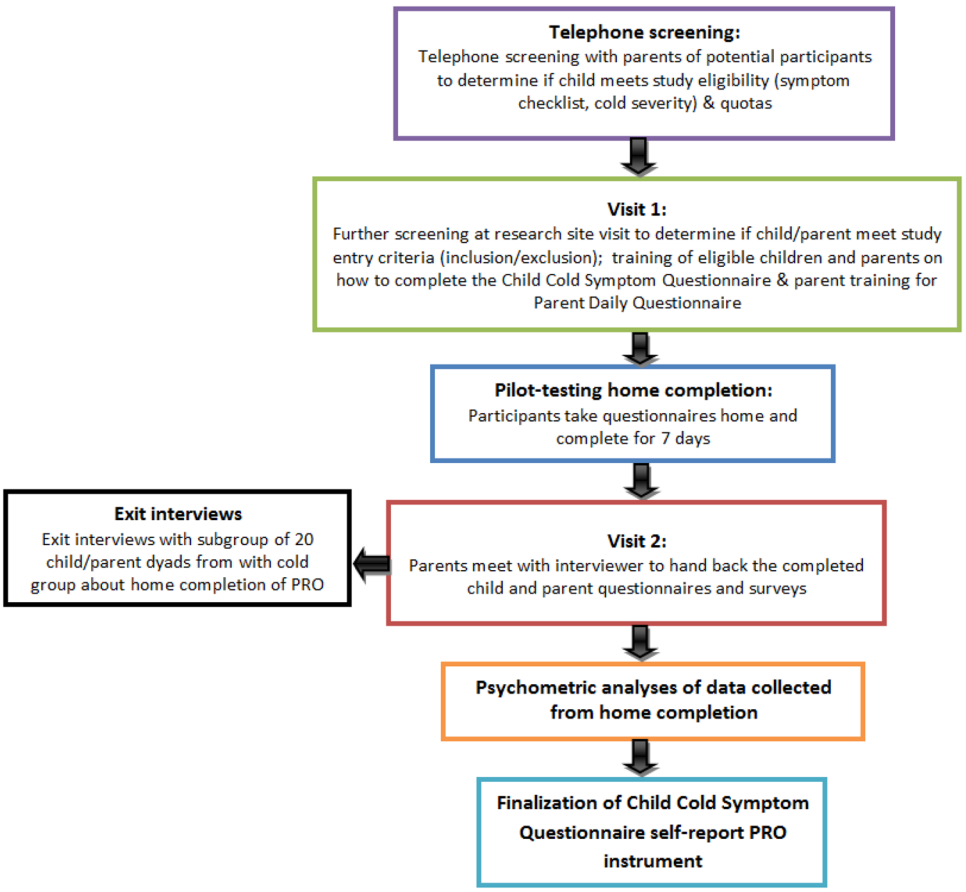
This was a multiple-center, observational, non-interventional study in which PRO data were collected from children in four cities in the US: Philadelphia, St Louis, Chicago, and St. Paul. An overview of the study is provided in Fig. 1.

2.2 Sample and Recruitment

A sample of 200 children aged 6–11 years was targeted: 150 who were experiencing cold symptoms ("cold eligible sample") and 50 healthy controls who were not ("control eligible sample"). Psychometrics do not simply rely on inferential statistics but imply the simultaneous consideration of multiple aspects generally based (at least in the context of classical test theory used in this paper) on the magnitude of association between variables using correlations. Hence, no formal calculation of sample size can be determined. Instead, sample sizes are defined to allow robust results to be obtained. When factor analyses are involved, a typical rule of thumb is that the ratio of individuals (participants) to variables (PRO items) is between 5 and 10 [14]. As the CCSQ included 32 items (for morning and evening completion), a minimum sample of 150 children with colds was targeted. The control group (children without colds) was included to allow the scores of children with colds to be compared with those without colds. It was decided that a sample of 50 children with no cold should be adequate to allow the scores to be reliably estimated in this subgroup and the comparison to be valid.

The sample was recruited and enrolled by a specialist patient recruitment agency (Global Market Research Group [GMRG]). A GMRG researcher confirmed that the child met initial inclusion/exclusion criteria through telephone screening with parent/caregivers, after which the parent/caregiver and child were invited to attend further screening at a research facility at visit 1. An approximately even distribution of male and female participants and children aged 6–8 years and 9–11 years was targeted. Diagnosis of cold symptoms was not confirmed by a clinician since most parents do not take a child with the common cold to the doctor. Instead, a parent-completed checklist was initially completed by telephone to identify the symptoms the child was currently experiencing to determine eligibility for enrollment. Those symptoms were further

Fig. 1 An overview of the study



confirmed by the child completing a checklist (without help from the parent) at visit 1.

Participants were required to be in the typical or higher school grade for their age group and reading at an appropriate level for their age. The child's parent/caregiver had to be one of the child's primary caregivers who could observe the child daily and answer the parent study questions. Exclusion criteria included known allergies (allergic rhinitis/hay fever) and/or asthma; treatment with antibiotics; or difficulty completing a cold symptom checklist at visit 1 (without help from the parent/caregiver). Child/parent dyads who met the inclusion/exclusion criteria were asked to answer the PRO questions twice a day for 7 days.

Cold eligible sample Visit 1 at the research facility was scheduled no more than 4 days following the onset of cold symptoms. A symptom checklist (not the CCSQ) was used to screen children into the study. This checklist included nine cold symptoms, which the child rated on a response scale of "not at all," "a tiny bit," "a little," "some," and "a lot." To be included in the cold eligible sample, the child had to choose a response of at least "some" for at least one of the nasal symptoms (stuffy nose or runny nose) and at least "a little" for one pain symptom and one other symptom.

Control eligible sample Children were excluded from the control sample if they had had a common cold in the 14 days preceding enrollment. Control participants had to choose a response of "not at all" or "a tiny bit" for all items on the symptom checklist and "a tiny bit" for a maximum of two items.

While the child completed the symptom checklist at visit 1, the recruiter assessed whether the child was able to read, understand, and complete the checklist independently or with minimal support. If not able, the child was deemed a screen failure. If a child/parent dyad was successful in meeting study criteria, the child was trained on how to complete the CCSQ and the parent was trained on how to complete the Parent Daily Questionnaire. They were then asked to answer daily questions contained in the child and parent booklets at home over the next 7 days. A date was arranged for the parent to return the completed booklets at the facility (visit 2). Children and parent/caregivers received monetary compensation for their time to participate in the study or, if deemed ineligible, a smaller amount for the screening visit. The children were compensated with $75 and the parent/caregivers were compensated with $100. If the parent and child participated

in an exit interview, then the child received an additional $10 and the parent received an additional $20. If the child attended visit 1 but was not eligible, the child received $10 and the parent $25.

2.3 Ethics

The study was conducted in accordance with the Declaration of Helsinki and approved by Copernicus, an independent review board (IRB) in the US (IRB approval #MAP2-11-470). Written informed consent was obtained from a parent/guardian of all children who participated in the study and written and verbal assent was obtained from all of the children.

2.4 Study Assessments

CCSQ This new, child-completed cold symptom questionnaire consisted of a set of 32 daily items; 15 items completed first thing in the morning and 17 completed in the evening before bed for 7 days [13]. An example is provided in Fig. 2, and the whole instrument is provided in the Supplemental File A (see the electronic supplementary material). All items have a 5-point, verbal-descriptor response

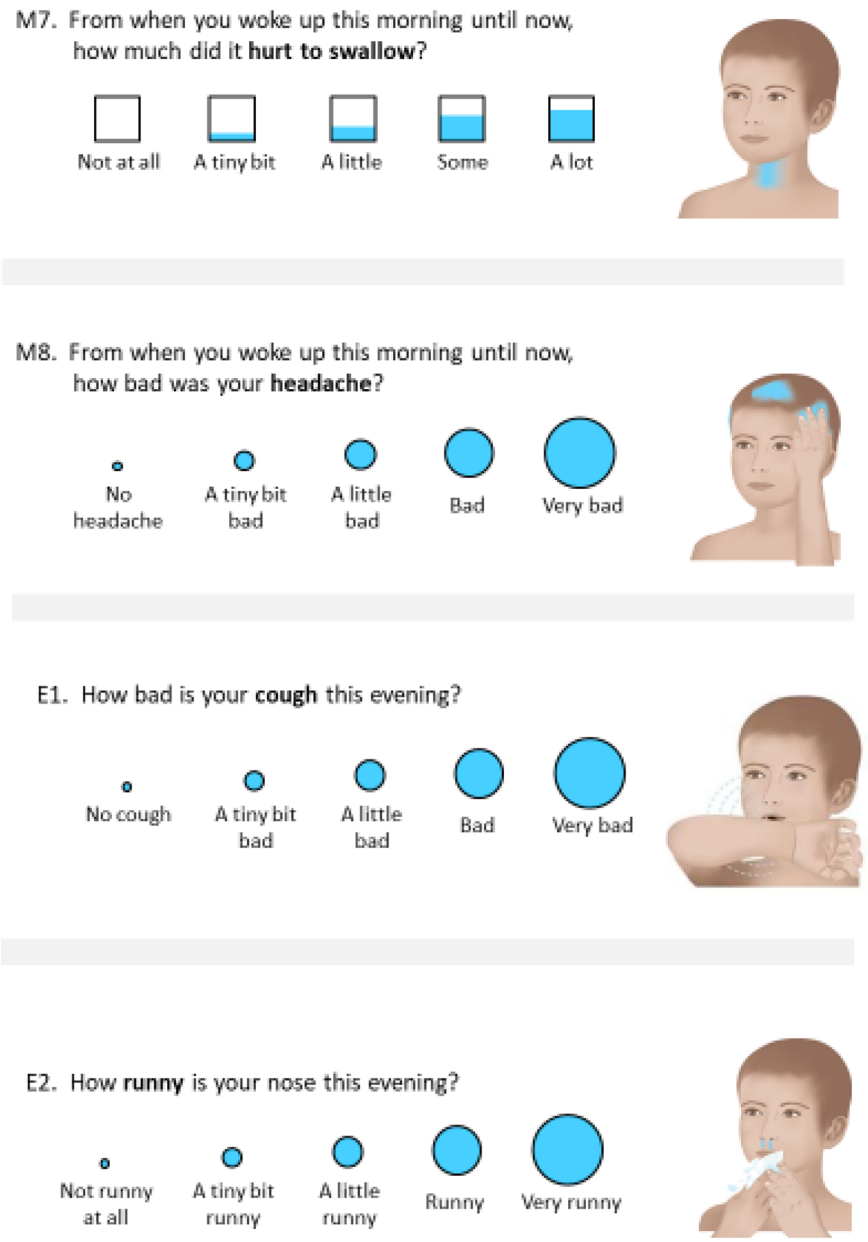


Fig. 2 Example items showing response scales and symptom illustrations © Johnson & Johnson

scale, scored from 0 to 4, that is represented pictorially with either circles of increasing size or boxes of increasing filled volume. Each question is associated with an illustration of a gender-neutral child without an emotional expression on which the location of the symptom being evaluated is shaded blue.

The questions assess the severity of eight cold symptoms of interest (nasal congestion [stuffy nose], runny nose, sinus pain/pressure, headaches, body/muscle aches, sore throat, cough, and chest congestion). All items had recall periods of less than 24 h, but the specific wording varied between items to allow the validity and reliability of various recall periods to be evaluated, allowing the optimal recall periods for each symptom to be chosen in the future. Recall periods tested in the morning items included "right now," "this morning," "from when you woke up this morning until now," and "last night in bed." The recall periods tested in the evening items included "right now," "this evening," and "for all of today." The questions were completed with pen on paper booklets, and the illustrations helped clarify what symptom the child should be thinking of for each question.

Other patient- and observer (parent)-reported outcome measures were administered to support the validation of the new instrument, which are detailed below.

- *Strep-PRO* The Strep-PRO is a seven-item questionnaire developed to assess the symptoms of Group A streptococcus (GAS) pharyngitis in children aged 5–15 years [15]. While not specific to the common cold, it is a validated child-reported measure of several symptoms relevant to the cold (e.g., sore throat), and so was included to evaluate convergent validity. This questionnaire was completed in the evening of days 1, 2, and 3.
- *Child Global Impression of Severity* (*CGI-S*) This single-item global assessment asked, "How bad is your cold today?" with a 0–4 verbal descriptor response scale ("no cold," "a tiny bit bad," "a little bad," "bad," and "very bad"). This item was completed each evening on days 1–7.
- *Parent Global Impression of Severity* (*PGI-S*) This single-item global assessment asked parents "How severe has your child's cold been over the past 24 h?" with a 0–4 verbal descriptor response scale ("no symptoms," "very mild," "mild," "moderate," and "severe"). This item was completed each evening on days 1–7.
- *Parent Global Impression of Change* (*PGI-C*) This single-item global assessment asked parents "How has your child's cold changed since Visit 1 when you started the study?" with a 0–4 response scale with verbal descriptors ("a lot better," "a little better," "the same," "a little worse," and "a lot worse). This item was completed each evening on days 1–7.
- *Parent Survey* A survey was completed by parents on the final evening of the study period. This included nine Likert-type and four open-response questions asking for feedback on the experience of completing the PRO questions for 7 days.

Exit interviews 30-min cognitive debriefing exit interviews were conducted at the end of the study in a subsample of 20 cold eligible children and their parents as a mixed-methods approach to instrument development. The exit interviews took place in person at visit 2 after the parents returned the completed questionnaires. The 20 child/parent dyads comprised the first 14 dyads with children aged 6–8 years old and the first six dyads with children aged 9–11 years old where the children had current cold symptoms and were willing to participate. These interviews were designed to further explore the content validity of the PRO beyond the previous qualitative study [13] and identify any challenges associated with the feasibility of completing the PRO twice daily. A semi-structured interview guide was developed that included a mix of open-ended questions and more direct cognitive debriefing questions. Feedback was solicited on how easy or difficult the child found answering the PRO questions and how easy or difficult completion of the PRO fit into daily routines. Due to time constraints, not every PRO item was cognitively debriefed with every child–parent/caregiver dyad, but all items had been debriefed in the previous qualitative study. The findings contributed to item reduction and scoring development decisions.

All exit interviews were audio-recorded and transcribed verbatim; the verbatim transcripts were qualitatively analyzed using thematic analysis methods and Atlas.ti software [16, 17]. Thematic analysis is a foundational, theory-free, qualitative analysis method, which offers flexibility to provide a rich, detailed, and complex synthesis of data that meets a very specific and applied aim [18]. An induction–abduction approach was taken to identifying themes in the data where themes were identified both by topics emerging directly from the data (inductive inference) and by applying prior knowledge (abductive inference). This enabled the analysis to remain rooted in the data, allowing participants to identify areas of importance for them, but also taking into consideration prior knowledge.

2.5 Psychometric Analyses

The first stage of the analyses was to develop the item-scale structure of the CCSQ that would be taken forward, including consideration of deletion of poorly performing or redundant items. This was determined based on the properties of quality of completion, item response distributions, confirmatory factor analysis (CFA), earlier qualitative findings, and the clinical relevance and importance of items.

The remaining analyses (test–retest reliability, convergent validity, known groups validity, and ability to detect change over time) were then performed to evaluate the psychometric properties of the resulting item-scale structure. All analyses were specified a priori in a statistical analysis plan and are detailed in Table 1.

The central limit theorem ensures that parametric tests can be used with large samples ($n > 30$), even if the hypotheses of normality are violated [19]. The *t* test, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) on ordinal data have been shown to be robust to violations of normality with small samples as well [20, 21]. Therefore, the following parametric tests were used for the comparison of variables between groups of subjects:

- *T* test when comparing two groups of participants
- ANOVA when comparing three groups of participants or more

Other statistics were also calculated:

- A Chi-square test was used to compare a qualitative variable between groups of subjects. If the underlying assumptions for the Chi-square test were not met, a Fisher exact test was used.
- The Pearson correlation coefficient was used when analyzing the relationship between quantitative variables.
- A paired *t* test was used to compare the change in a quantitative variable to 0.
- The intra-class correlation coefficient (ICC) was used for the evaluation of test–retest reliability [22].
- Effect size (ES), standardized response mean, and Guyatt's statistics were used for the evaluation of ability to detect change over time.

The following methods were used to analyze the structure of the CCSQ:

- CFA for the exploration of potential multi-item score structures [23, 24].
- Multi-trait analysis for the evaluation of the relationships between single items and hypothesized multi-item scores [25].

The emphasis of psychometric analysis is on evaluating the magnitude of relationships among variables and the overall pattern of results rather than on significance testing. Because of this, no adjustments were used for multiplicity of tests for psychometric analysis. Where specific significance tests were used, the threshold for statistical significance was fixed at 5% for each test. All data processing and analyses were performed with SAS software for Windows (Version 9.2, SAS Institute, Inc., Cary, NC, USA).

3 Results

3.1 Demographics and Clinical Characteristics of Sample

Fifteen of 200 children screened did not meet the inclusion criteria, resulting in a full eligible sample of 138 children with colds and 47 controls. The mean age of the full eligible sample was 9.1 years (SD = 1.7, range = 6.0–11.9 years), with similar mean ages in the cold (9.1 years) and control (9.0 years) subgroups. There were slightly more female children than male (52.9%); 77.3% of the children were Caucasian, 10.2% were black/African American, and 8.1% were Hispanic/Spanish American/Latino. Overall, 28.3% of the children experiencing cold symptoms had been treated at home with at least one OTC medicine in the 7 days prior to visit 1 and 21.0% expected to give their child medicine in the next 7 days. The first participant was enrolled on 22 December 2011, and the final participant was interviewed on 12 February 2012.

Results from the symptom checklist completed at visit 1 indicate that children responded "some" and "a lot" most often to the symptoms of "runny nose," "stuffy nose," and "coughing" (70.3%, 64.5%, and 55.1%, respectively). The greatest number of "not at all" responses was found for "aching arms and legs," "hurting face around eyes and nose," "hard to breathe," and "hurting head" (48.6%, 44.2%, 39.9%, and 33.3%, respectively).

3.2 Quality of Completion

Quality of completion was excellent, with a mean of zero missing data per child and a maximum number of missing items per child of one. Seven items were missed at least once, but no more than twice. There was no pattern of any one item being missed more often than others (candidate for deletion) or of items being missed more often in the younger 6–8 age group.

3.3 Item Response Distributions

Item-level response distributions on the morning and evening of day 2 for the CCSQ demonstrated that most items provide a good spread of responses across the response scale, with approximately normal distributions. These are provided in Supplemental File B (see the electronic supplementary material). There was some evidence of floor effects (a high percentage of children choosing the lowest possible score, in this case 0) and/or slightly positively

skewed distributions towards the lower end of the scale for the "chest congestion," "headache," and "arm and leg aches" items. This was limited and not considered of concern because the low scores are consistent with these symptoms not being commonly reported for children with colds. There was some evidence of relatively strong floor effects in children aged 6–8 years compared with those aged 9–11 years for some items. These were the items assessing "being kept awake by cough," "being kept awake by a stuffy nose," "how clear the nose was after blowing," "pain around eyes and nose," "tightness on face," "runny nose," "wipe or blow your nose," "hard to breathe deep into chest," "tight chest," "headache," "sore throat," and "arms and legs ache." Notably, apart from the items assessing "how clear the nose was after blowing" and "runny nose," these items were deleted (see Sect. 4.4). The most highly endorsed symptoms (for both age groups and in the total sample) were "cough," "runny nose," and "stuffy nose." The least highly endorsed symptoms were the items assessing "tight chest," "pain around eyes and nose," "sore throat," "headache," and "arm and leg aches."

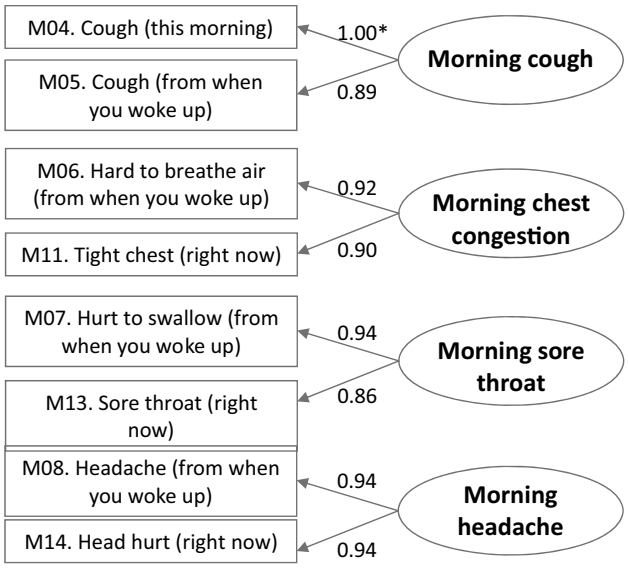


Fig. 3 Definition of multi-item scores without nasal dimensions showing standardized factor loadings (cold eligible sample, $N = 138$). *Standardized factor loadings

3.4 Development of the CCSQ Scores

The appropriateness of grouping the items into seven hypothesized multi-item scores (comprising 16 items) was tested using CFA; fit indices and standardized factor loadings were examined. These seven domains assessed the various types of cold symptoms separately for the overnight, morning, and day timeframes (morning nasal congestion, morning cough, morning chest congestion, morning sore throat, morning headache, evening nasal congestion, and day nasal congestion). Standardized factor loading of items on their hypothesized domains were high for most items (all > 0.80), with the only exception being the nasal congestion items. Nasal congestion included the "stuffy nose" items (M09, E11, and E03) and the "clear nose" items (M10 and E04), which all loaded at < 0.58 for morning, day, and evening scores. Nasal congestion also included items asking about "pain around eyes and nose" (sinus pain), which had a high loading on both the morning and evening nasal domain (all > 0.72). Overall these data suggest that the "pain around eyes and nose" items do not fit with the other nasal items that assess "stuffy nose" or "clear nose" symptoms.

Repeating the analysis without the "pain around eyes and nose" items resulted in all a priori CFA thresholds being satisfied (goodness of fit index = 0.949; root mean square error of approximation = 0.030; comparative fit index = 0.996), suggesting that this revised structure was acceptable. This involved grouping the items into the scores of "morning cough," "morning chest congestion," "morning sore throat," "morning headache," "morning nasal congestion," and "evening nasal congestion." All standardized loadings were high across the multi-item dimensions tested (all > 0.86, Fig. 3).

High item loadings on each factor and very high inter-item correlations suggested a high level of redundancy among items within scores. This is unsurprising, as the items within each score are measuring aspects of the same cold symptom. Based on these findings and importantly considering the findings of the previous qualitative research, the CCSQ was reduced from 32 to 15 single items. Items deleted were those that were less well understood as well as showing redundancy and/or poor psychometric performance. The CFA results indicate that these 15 items could be scored as single-item scores or form five multi-item scores, as outlined in Table 2 (morning and evening nasal scores, morning and evening aches and pains scores, and a "day" nasal score). The high correlations of items within scores suggest that single-item scores are likely sufficient, but both options were tested in the remaining analyses. This version of the CCSQ is provided in Supplemental File A.

3.5 Changes in Score Distributions over Time

For all single- and multi-item scores, the mean scores in the cold eligible sample decreased steadily from day 1 to day 7, consistent with the children's colds improving over the 7-day observation period and providing evidence that the instrument is responsive to change.

3.6 Test–Retest Reliability

Stable children were defined as those whose cold was rated as "unchanged or almost the same" compared to the beginning of the study by their parent on the PGI-C; 106 children met this criterion. The stability of CCSQ scores for this subgroup was assessed between days 1 and 2 for evening scores, and between days 2 and 3 for morning scores.

For the evening scores (Table 3), the single-item scores for the "cough," "stuffy nose," "sore throat," "headache," and "day cough" domains and all three composite multi-item scores had ICCs > 0.70, indicating good test–retest reliability. The remaining four items had ICC scores between 0.60 and 0.70 ("runny nose," "clear nose," "day wipe or blow nose," and "day stuffy nose"). Given that the cold symptoms are variable and fast changing, this seems acceptable. Reliability coefficients for single-item measures are also typically lower than for multi-item scores. Results for the morning scores were similar.

When evaluated within age subgroups, test–retest reliability in the 6- to 8-year-old group was almost as strong as in the 9- to 11-year-old group. For the morning scores, the ICC [and CCC (concordance correlation coefficient) and Pearson correlations] for the 6–8 year olds were lower than those for the 9–11 year olds for only stuffy nose (0.62 vs. 0.70), clear nose (0.59 vs. 0.77), sore throat (0.66 vs. 0.74) (these data are provided in Supplemental File B).

3.7 Construct Validity

3.7.1 Convergent Validity

To evaluate convergent validity, correlations of the CCSQ single- and multi-item scores with Strep-PRO scores were evaluated (see Table 3 for evening of day 1). As hypothesized, the highest correlations (> 0.60) were between pairs of items measuring similar symptoms (e.g., the correlation between item assessing "sore throat" and "throat hurt" was $r = 0.82$). Low correlations were reported between scores that would not be expected to correlate highly (e.g., "stuffy nose" and "throat hurt," $r = 0.21$), providing further evidence of concurrent or convergent/divergent validity. Results were equally strong on the evenings of day 1, day 2, and day 7, and the mornings of day 2, day 3, and day 7. The results were also similar when the correlations were examined in the age subgroups (see Supplemental File B).

3.8 Known Groups Validity

3.8.1 Comparison Between Child Cold Symptom Scores and Child Global Impression of Severity (CGI-S)

The single-item and multi-item scores from the CCSQ were compared among groups who differed in overall cold severity as reported by the children on the CGI-S item. There

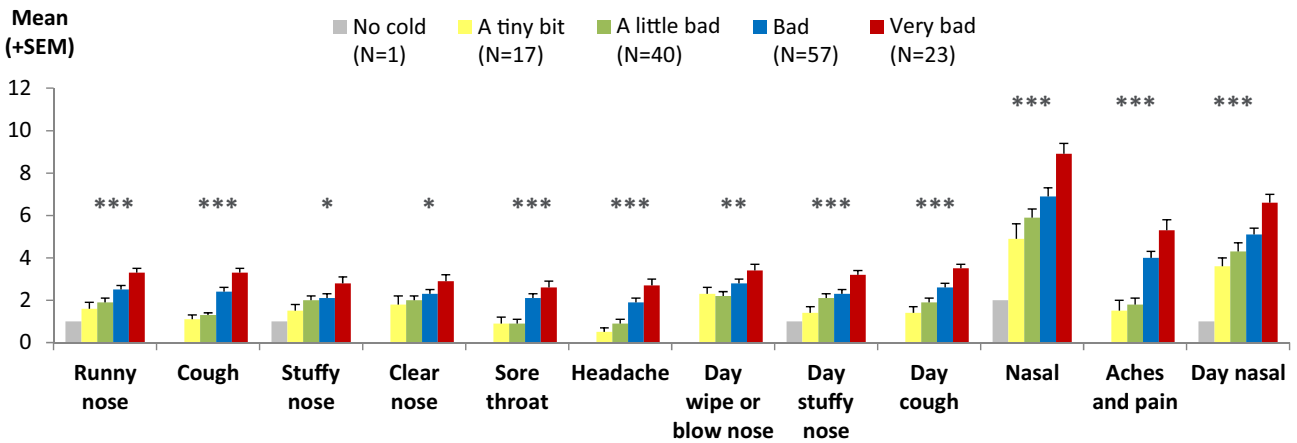


Fig. 4 Known groups validity: ANOVA comparison of CCSQ evening scores according to CGI-S defined groups at day 1 ($N = 138$). Note, "nasal," "aches and pain," and "day nasal" are multi-item scores (made up of 3, 2, and 2 items, respectively) and therefore have possible score ranges of 0–12, 0–8, and 0–8, respectively, rather than 0–4 as for the other items. *ANOVA* analysis of variance, *CCSQ* Child Cold Symptom Questionnaire, *CGI-S* Child Global Impression of Severity, *SEM* standard error of the mean. *ANOVA showed statistically significant differences at the $P < 0.05$ level. **ANOVA showed statistically significant differences at the $P < 0.01$ level. ***ANOVA showed statistically significant differences at the $P < 0.001$ level

was a pattern of significantly higher mean CCSQ evening scores (indicating worse symptoms) for the children who also scored higher (worse) on the CGI-S on the evening of day 1 ($P < 0.05$ for all scores), with the expected monotonic increases across the groups (Fig. 4). Children reporting the worst colds that evening also reported worse scores on individual symptoms, providing evidence of known groups validity. The only score that did not have a clear stepwise increase in scores across severity groups was the "day wipe or blow your nose" score. Similar patterns of results were found for the morning scores, and when evaluated within age subgroups (6–8 vs. 9–11 years).

3.8.2 Comparison Between Child Cold Symptom Scores and Parent Global Rating of Severity (PGI-S)

A pattern of higher scores (indicating worse symptoms) was reported by children whose parents reported worse scores on the PGI-S, with statistically significant differences among the groups for all CCSQ scores ($P < 0.05$). Thus, a child's ratings of symptoms were generally aligned with the parent's ratings of overall cold severity. The only score that did not show a clear stepwise increase across the PGI-S defined severity groups was the "day wipe or blow nose" score. Results were very similar for the morning scores and when performing the analysis by age group (6–8 vs. 9–11 years).

3.9 Ability to Detect Change over Time

Changes in the CCSQ scores were compared among children defined as "improved," "unchanged," and "worsened" on the PGI-C score and changes in the CGI-S and PGI-S scores between day 1 and day 2 and between day 1 and day 7. These results provide evidence that the CCSQ can detect changes over time, regardless of the rating used to define change. Generally, there were improvements in CCSQ scores for children rated as "improved," negligible changes for children rated as "unchanged," and worsening or negligible changes for those rated as "worsened." Weaker results (less clear differences among these groups) were found when the change groups were defined using the parent-completed PGI-C as opposed to the child-completed CGI-S; this may reflect weakness with the global change measure, which relies on the parents mentally averaging and recalling back to the beginning of the study, rather than weaknesses in the CCSQ. It may also suggest that parents are not able to observe their child's cold symptoms closely enough to provide a valid rating—which is part of the rationale for developing a child-report measure rather than a parent-report measure.

As an example, Fig. 5 shows the change in CCSQ evening scores between the evenings of day 1 and day 2 compared among groups defined according to changes in the CGI-S over the same time period. The CCSQ evening scores demonstrated improvement and worsening as expected for the CGI-S groups. For the children in the "improved" group, small to moderate decreases in CCSQ evening scores were observed (ES range − 0.11 to − 0.38), compared to mostly small or moderate increases in scores (ES range − 0.24 to 0.39) in the "worsened" group. The changes in CCSQ scores for the "no change" group were small or negligible (ES range − 0.19 to 0.11). Differences in changes in score between the "improved," "no change," and "worsened" groups were statistically significant for all CCSQ scores

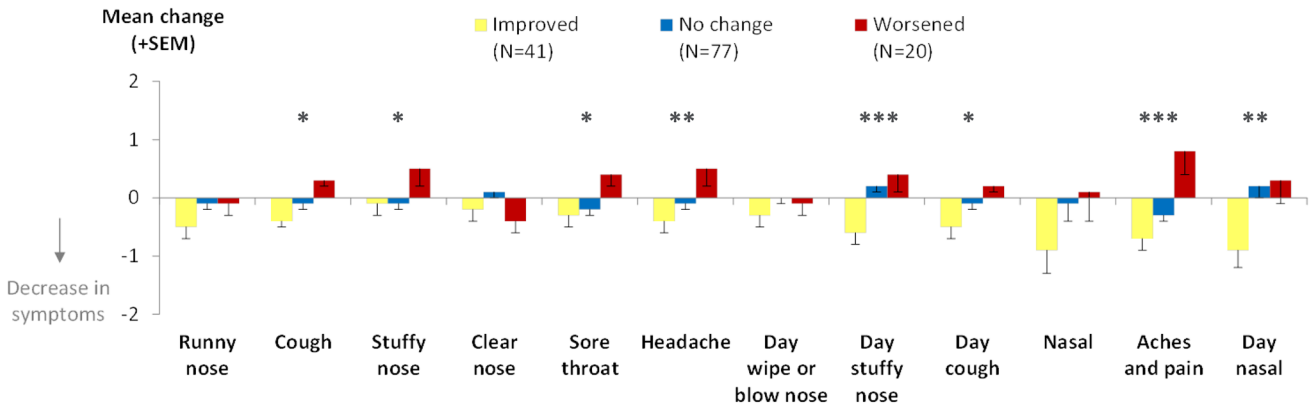


Fig. 5 Change over time: ANOVA comparison of changes in CCSQ evening scores between day 1 and day 2 for change groups defined by CGI-S changes between day 1 and day 2 ($N = 138$). *ANOVA* analysis of variance, *CCSQ* Child Cold Symptom Questionnaire, *CGI-S* Child Global Impression of Severity, *SEM* standard error of the mean. Note, "nasal," "aches and pain," and "day nasal" are multi-item scores (made up of 3, 2, and 2 items, respectively) and therefore have possible change score ranges of 0–12, 0–8, and 0–8, respectively, rather than 0–4 as for the other items. *ANOVA showed statistically significant differences among change groups at the $P < 0.05$ level. **ANOVA showed statistically significant differences among change groups at the $P < 0.01$ level. ***ANOVA showed statistically significant differences among groups at the $P < 0.001$ level

Table 1 Description and criteria for psychometric validation analyses performed

Property	Description/definition	Criteria for consideration
Quality of completion	Evaluation of frequency and percentage of missing items per child, frequency and percentage of missing data per item, number of children with at least one missing item, number of missing questionnaires for each planned assessment	Items with high levels of missing data were considered for deletion
Item response distributions	Examined in the total sample and within age subgroups (6–8 and 9–11 years) to evaluate if any items exhibited a skewed distribution, floor/ceiling effects, or bimodal distribution or if any particular response options were overly favored. Floor effects refer to a high percentage of children with the lowest (best) possible score for an item or score, and ceiling effects refer to a high percentage of children with the highest (worst) possible score for an item or score	Items with evidence of problematic distributions were flagged and considered for deletion
Confirmatory factor analysis (CFA)	Performed to evaluate the a priori hypothesis for aggregating the items into multi-item domains by calculating a score for each type of cold symptom separately during the overnight, morning, day, and evening timeframes	The quality of the CFA models was assessed according to the following goodness-of-fit indices: [14, 25, 31] Root mean square error of approximation (RMSEA): good fit if RMSEA < 0.05, acceptable fit if RMSEA < 0.08 Root mean square residual (RMR) and standardized RMR: good fit if RMR < 0.05 Goodness of fit index (GFI) and adjusted GFI (AGFI): good fit if GFI or AGFI > 0.90 Normed fixed index (NFI) and comparative fit index (CFI): good fit if NFI or CFI > 0.90 In addition, the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) of the models underpinned by hypotheses A and B were computed and compared. AIC and BIC are indices of fit that can be compared between non-nested models. The strength of the relation of the items to the unobserved variables that they are assumed to measure was evaluated by the standardized loadings. Potential improvement in the model fit was observed using modification indices, which indicated whether or not there were better item grouping
Score distributions	Score distributions of the resulting multi-item scores and single-item responses were described in the cold eligible sample and by age category (6–8 years; 9–11 years)	Percentages of participants scoring at the lowest and highest possible values of scores and response scales were evaluated to detect potential floor or ceiling effects
Test–retest reliability	Evaluated to establish the ability of the instrument to give reproducible results when administered twice, over a given time period, in a sample with stable health [31] Evaluated between day 1 and 2 for evening assessments and day 2 and 3 for the morning assessments to maximize the likelihood of capturing an adequate sample of patients with stable colds. Assessed by calculating intra-class correlation coefficients (ICCs) in children whose cold severity was unchanged according to the Parent Global Impression of Change (PGI-C)	This analysis was considered exploratory given that cold symptoms may not be stable, even when comparing 2 consecutive days. Therefore, while the normal threshold of ICC > 0.70 was targeted, it was accepted that it may not be achieved

Table 1 (continued)

Property	Description/definition	Criteria for consideration
Convergent validity	Involved examining correlations between the scores of the instrument under study and those of a validated instrument assessing related constructs. Correlations of the Child Cold Symptom Questionnaire (CCSQ) single- and multi-item scores with the Strep-PRO scores [15] were computed in the cold eligible sample and the full eligible sample	Evidence of convergent or concurrent validity was considered to have been demonstrated if there was a logical pattern of correlation among scores, with scores measuring similar or related symptoms correlating more highly than scores measuring unrelated symptoms. In particular, the Strep-PRO item scores for "sore throat" and "headache" were expected to correlate most highly (and with correlation coefficients > 0.6) with CCSQ scores assessing corresponding symptoms. The Strep-PRO item score for "pain swallowing" was expected to correlate most highly with CCSQ sore throat
Known groups validity	Known groups validity involved comparing scores among groups that would be expected to differ on the construct of interest [14]. The single-item and multi-item scores from the CCSQ were compared among groups defined according to the rating of overall cold severity completed by the children (Child Global Impression of Severity [CGI-S]). Mean scores from the CCSQ were compared among groups defined according to: CGI-S responses for that day Parent Global Impression of Severity (PGI-S) responses for that day Children with and without current cold	Statistically significant differences for analysis of variance (ANOVA) or *t* test comparisons among the groups being compared and step-wise, monotonic differences in mean scores were considered adequate evidence of known groups validity
Ability to detect change over time	The ability of scores of the CCSQ to detect changes over time in individuals who had changed with respect to the symptom measured [32–34]. Given that the typical transitional period for cold symptoms is 2–3 days, it was considered reasonable to assume that a large proportion of children would experience an improvement over the 7-day study Changes in the CCSQ scores from day 1 to day 2 and from day 1 to day 7 were described and compared among children considered "improved," having "no change," and "worsened," as determined using the ratings on the PGI-C, CGI-S, and PGI-S	Change scores were compared among the change groups using ANOVA. A paired *t* test was also computed to compare the change in the CCSQ scores to 0 within each group Effect size (ES), standardized response mean, and Guyatt's statistic were calculated to evaluate the magnitude of changes in CCSQ scores over time. Based on Cohen's guidance for interpretation of ES [35], changes in score were interpreted as: ES around 0.20: small change ES around 0.50: moderate change ES around 0.80: large change Statistically significant changes over time and at least moderate ESs were considered evidence of responsiveness

Table 2 Final scoring structure of the single-item and multi-item scores

Composite multi-item scores	Single-item scores	Morning item	Evening item
Items retained			
Nasal	Runny nose	M03. Runny nose (from when you woke up)	E02. Runny nose (this evening)
	Stuffy nose	M09. Stuffy nose (right now)	E03. Stuffy nose (right now)
	Clear nose	M10. Clear nose (right now)	E04. Clear nose (right now)
	Cough	M04. Cough bad (this morning)	E01. Cough bad (this evening)
Aches and pain	Sore throat	M13. Sore throat (right now)	E07. Sore throat (right now)
	Headache	M14. Head hurt (right now)	E08. Head hurt (right now)
Day nasal	Day wipe or blow		E10. Wipe or blow nose (for all of today)
	Day stuffy nose		E11. Stuffy nose (for all of today)
	Day cough		E12. Cough amount (for all of today)

except for "runny nose," "clear nose," "day wipe or blow nose," and "nasal" scores.

3.10 Findings from Parent Survey and Exit Interviews

Most parents provided very positive feedback on the CCSQ ($n = 148$, 80.0%), reporting that it was "very easy" or "easy" to fit answering the questions into their schedule over the past 7 days ($n = 158$, 85.4%), to think of the answers to questions every day ($n = 169$, 91.4%), for the child to answer the questions by himself/herself ($n = 156$, 84.3%), for the child to recall symptoms from the previous night ($n = 142$, 76.8%), and for the child to recall symptoms from the daytime ($n = 138$, 74.6%).

During the survey, almost all parents ($n = 160$, 86.5%) reported that their children were "willing" or "very willing" to answer the questions twice a day, and needed little or no help during the first 3 days ($n = 153$, 82.7%) and the last 3 days ($n = 172$, 93.0%). Some parents also commented that the illustrations helped the children to understand the questions ($n = 18$, 9.7%).

The exit interviews with 20 children and their parents also provided support for the face and content validity of the items. All but two children were able to complete the questions with little or no help. Ten children (50.0%) reported that they asked for help from their parent on the first day, but were then able to answer on their own for the remaining days: "I helped—I kind of went through it with him the first day and after that, he was able to do it by himself" (parent of 9-year-old).

The qualitative data also supports the high completion rate evident from the descriptive statistics, with two reasons evident: children appeared to enjoy completing their booklets ["I got to express my feelings." (9-year-old girl)] and parents took responsibility for the child's booklet, reminding their children to complete it. Most children also said they found the illustrations helpful in answering the questions: "If there was no picture, I'd be so confused." (7-year-old girl).

During the interviews, high levels of understanding were demonstrated by children of all ages. Only two children (aged 6 and 8 years) struggled to read any items. They were able to read with the aid of the interviewer/parent and seemed to understand the question without help once it had been read to them.

4 Discussion

The findings of this psychometric evaluation study provide evidence that the CCSQ items and multi-item scores provide valid and reliable patient-reported measures of cold symptoms in children aged 6–11 years that are not too burdensome for children to complete twice daily for 7 days, even for 6- to 8-year-old respondents. There is preliminary evidence that the questionnaire is responsive to change (albeit based only on data from a non-interventional study). The fast changing, variable nature of cold symptoms means that any treatment benefits are likely to be difficult to detect, so an instrument with strong psychometric properties is essential.

There were exceptionally low levels of missing data for both age groups throughout the study. Evaluation of item distributions in this study suggests that the response scales capture the variability of symptom scores over the course of a cold. Construct validity testing supported a priori hypotheses concerning relationships with the Strep-PRO and global items assessing cold severity. The most highly endorsed items (based on the response distributions) were also those that were associated with the cold symptoms selected most often by children upon enrolment in this study and are more prevalent in children [26, 27]. Additionally, they were items

Table 3 Test–retest reliability between the day 1 and day 2 evening scores and convergent validity Pearson correlations between the Child Cold Symptom Questionnaire (CCSQ) and the Strep-PRO scores on day 1

Type of score	Symptom	Test–retest reliability of stable evening scores (days 1 and 2) based on the PGI-C ("unchanged or almost the same") ($N=106$)					Convergent validity Pearson correlations between the day 1 evening scores and the Strep-PRO day 1 scores ($N=138$)							
		Mean change (SD)	*P* value	ICC	CCC	Pearson correlation	Throat hurt	Head hurt	Hurt to swallow	Fever or feel warm	Food eating	Playing	Feeling tired	Total score
Single-item scores	Runny nose	−0.3 (1.0)	0.007	0.64	0.63	0.66	0.34*	0.24*	0.25*	0.26*	0.24*	0.23*	0.18*	0.34*
	Cough	−0.1 (1.0)	0.285	**0.73**	0.73	0.74	**0.59***	0.44*	0.48*	0.33*	0.41*	0.37*	0.42*	**0.61***
	Stuffy nose	−0.1 (1.0)	0.259	**0.72**	0.71	0.72	0.21*	0.30*	0.17*	0.19*	0.12	0.16	0.18*	0.27*
	Clear nose	0.0 (1.2)	0.934	0.63	0.63	0.63	0.16	0.28*	0.18*	0.25*	0.28*	0.28*	0.06	0.29*
	Sore throat	−0.2 (0.9)	0.043	**0.76**	0.76	0.77	**0.82***	0.35*	**0.82***	0.37*	0.46*	0.35*	0.33*	**0.74***
	Headache	−0.1 (0.9)	0.294	**0.80**	0.80	0.80	0.39*	**0.79***	0.42*	0.20*	0.36*	0.32*	0.41*	**0.61***
	Day wipe or blow	−0.1 (1.1)	0.239	0.64	0.64	0.65	0.29*	0.18*	0.25*	0.22*	0.23*	0.22*	0.15	0.31*
	Day stuffy nose	−0.1 (1.1)	0.586	0.68	0.68	0.68	0.24*	0.30*	0.25*	0.27*	0.15	0.20*	0.30*	0.34*
	Day cough	−0.2 (1.0)	0.044	**0.73**	0.73	0.73	0.54*	0.35*	0.45*	0.29*	0.34*	0.30*	0.41*	0.54*
Composite multi-item scores	Nasal	−0.4 (2.3)	0.088	**0.72**	0.71	0.72	0.32*	0.38*	0.27*	0.32*	0.29*	0.31*	0.19*	0.41*
	Aches and pain	−0.3 (1.4)	0.033	**0.83**	0.83	0.84	**0.72***	**0.68***	**0.74***	0.35*	0.49*	0.40*	0.45*	**0.80***
	Day nasal	−0.2 (1.6)	0.240	**0.71**	0.70	0.71	0.33*	0.30*	0.32*	0.31*	0.24*	0.26*	0.28*	0.41*

ICC > 0.70 (thus surpassing the a priori threshold) are highlighted in bold

ICC intra-class correlation coefficient, *PGI-C* Parent Global Impression of Change

*Convergent validity Pearson correlations that were statistically significant at the $P<0.05$ level; the highest convergent validity correlations (≥0.60) are highlighted in bold

associated with symptoms most commonly reported in the previous qualitative research, and best understood by the children when the presence of symptoms could be explored through more open-ended qualitative enquiry and did not rely on comprehension of specific terms. Therefore, it is uncertain whether ease of understanding influenced the response distributions for the items or whether these items reflect the most commonly experienced cold symptoms. Nevertheless, these findings provide some evidence that the items assessing simpler concepts and that were more simply worded may have stronger validity and reliability for the younger children.

This study provides strong evidence supporting the validity of the items and multi-items scores for inclusion as endpoints in clinical trials to evaluate the efficacy of cold medicines. Whether multiple- or single-item scores are more appropriate for inclusion as endpoints in a clinical trial evaluating a cold treatment may depend upon the aim of the trial and the symptoms targeted by the product. The CCSQ provides a battery of potential endpoints for measurement of various cold symptoms, which provides flexibility to measure the symptoms of interest. Similarly, the most appropriate recall period will depend upon the specific context of use in terms of study design, such as the dosing interval or time of day. All recall periods tested during this study were found to have strong validity and reliability. In fact, four items addressing nasal congestion were used to construct the primary (M09) and secondary (M02, M03, and E04) endpoints in a placebo-controlled clinical trial for the decongestant pseudoephedrine with children aged 6–11 years old [28]. Wording of the PRO items were identical, except "stuffy nose" was used in M03 instead of "runny nose," and the word "once" was not included in E04. The endpoints detected differences between treatment and placebo groups, thus providing additional evidence that children in this age group can self-report on the severity and frequency of subjective symptoms in a clinical trial.

While the possibility that some symptoms reported were due to allergies or other respiratory conditions cannot be ruled out, the trajectory observed of the cold symptoms reducing to very low levels in almost all participants during the 7-day study period was consistent with a common cold. Feedback from exit interviews and the extremely low level of missing data across all ages suggest that the instrument was well accepted by the children and easily fitted into their daily schedule. However, since this was a paper instrument, the timeliness of completion cannot be verified.

As noted in the introduction, the CARIFS [7] is a parent/caregiver report measure developed for use in babies, infants, and children up to age 12. In terms of the symptoms assessed, there is a high degree of overlap/consistency, with both measures assessing headache, sore throat, cough, and nasal congestion. It is also notable that the CARIFS does not include assessment of chest congestion or sinus pain, both removed during this validation study from the CCSQ. In terms of differences, the CARIFS does include items assessing fatigue/tiredness, muscle aches or pains, fever, and vomiting. However, these are all symptoms that are more relevant to influenza than the common cold. Moreover, as a parent/caregiver report, in addition to assessing symptoms, the CARIFS also includes items assessing observable behaviors associated with colds and influenza, such as the child appearing "irritable, cranky, fussy" or "not playing well," which would not be appropriate for inclusion in the self-report CCSQ.

Similar to the correlations observed between the CCSQ and the Strep-PRO, the CARIFS has been shown to have moderate correlations with physician, nurse, and parent global assessments. Moreover, similar to the improvements in CCSQ scores over the 7-day study period, the CARIFS was shown to improve over 14 days. Neither CFA nor test–retest reliability were evaluated for the CARIFS.

As an adult measure of upper respiratory symptoms, comparison with the WURSS-11 is arguably less relevant [6]. However, it is notable that, CFA of the WURSS-11 also supported grouping of the items in multi-item symptom scores. It is also interesting to note that, similar to the CCSQ, the evidence of test–retest reliability for the Flu-PRO has also been shown to be somewhat mixed, with the common threshold of 0.70 not always met [4]. This provides support for the hypothesis that this reflects the fluctuating nature of cold/flu symptoms.

Because of age and developmental changes, it is important to demonstrate that a pediatric PRO instrument has content validity and strong psychometric properties within narrow age bands [12, 29, 30]. The qualitative research that preceded this study suggested that the PRO items tested in this study can be used with confidence in children aged 9 years and older, and in children aged 6–8 years with initial adult supervision to explain the more difficult concepts [13]. The evidence presented here of strong reliability and validity, even in the 6–8-years age group supports those findings and suggests the instrument is appropriate to use in children as young as 6 years, so long as they can read the items. Although the item response distributions and score distributions were examined separately in the two age subgroups (6–8 years and 9–11 years), one limitation of the present study is that the sample size was not sufficient to support further age analyses within these subgroups. Between the ages of 6 and 8 years, substantial differences exist in development of reading ability, concept understanding, and general comprehension for most children. In this regard, we believe the visual cues for each symptom provided by the illustrations of a child demonstrating a cold symptom, with added blue shading to direct attention to the affected area, helped the youngest children understand the item, even if their reading

comprehension was lower. This was also the aim of the increasing sizes of the circles and filled boxes linked to the response options. Smiley and sad faces often used in other children's PRO tools can be confusing because they convey emotion, which may confuse a child if they do not apply.

Another limitation is that the PRO questions were developed and this validation study conducted only in the US. For the instrument to be used in other countries and cultures, appropriate translation and linguistic validation (involving two forwards and one backwards translation at a minimum) would be required in addition to psychometric evaluation. While black/African American (10.2%) and Hispanic/Spanish American/Latino (8.1%) children participated in the study, future studies may benefit with larger proportions of non-Caucasian participants. For the assessment of convergent validity, the lack of existing measures of cold symptoms for use in this age range meant that options were limited when trying to identify measures against which CCSQ scores could be compared. Ideally specific, validated measures of nasal symptoms and other concepts would have been included. Finally, the instrument was developed using a pen/paper questionnaire. If transformed into an electronic PRO (ePRO)—for example, completed on a hand-held touch screen device—then the ePRO version would need confirmation that it had equivalent or at least equally strong content validity and psychometric validity.

A copy of the version of the CCSQ with 15 items that emerged from this study is provided in Supplemental File C along with the conditions for use. The Child Cold Symptom Questionnaire is available for educational, research, or clinical use at no cost, provided such use includes an attribution statement that reads: "The Child Cold Symptom Questionnaire was developed by Johnson & Johnson Consumer Inc., McNeil Consumer Healthcare Division." Any other use must receive permission from Johnson & Johnson Consumer Inc., McNeil Consumer Healthcare Division.

5 Conclusions

The findings of this study demonstrate that the CCSQ provides a valid and reliable assessment of cold symptoms for use across the 6–11-year age range. They provide evidence supporting the validity of the items and multi-items scores for inclusion as endpoints in clinical trials to evaluate the efficacy of cold medicines. Further study is recommended to evaluate the ability of individual items and the instrument to detect changes due to symptomatic treatment and to identify minimal levels of change that could be considered meaningful in such a context.

Acknowledgements The authors would like to acknowledge the contributions of Helene Gilet, who contributed to the analysis of the data. The authors would also like to thank all the children and parents who participated in the research. The study was sponsored by McNeil Consumer Healthcare. Editorial support was provided by R. Hall from Adelphi Values, and funded by members of the Consumer Health Products Association Pediatric Cough and Cold Task group.

Declarations

Funding This research was funded by McNeil Consumer Healthcare, a Division of Johnson & Johnson Consumer, Inc. Editorial support was provided by Rebecca Hall from Adelphi, and funded by member companies of the Pediatric Cough and Cold Task group of the Consumer Health Products Association (Washington, DC).

Competing interests R. Arbuckle is an employee of Adelphi Values (known as Mapi Values at the time the research was conducted), a health outcomes agency commissioned to conduct this research. At the time of the study, C. Marshall, K. Bolton, and A. Renault were also employees of Mapi Values, but all have since left the company. P. Halstead, B. Zimmerman, and C. Gelotte are former employees of McNeil Consumer Healthcare, a Division of Johnson & Johnson (J&J) Consumer, Inc., and hold stock or stock options in J&J. C. Gelotte has received consultant fees from J&J and an honorarium from the Consumer Health Products Association. The authors have no other conflicts of interest regarding the content of this article.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved and overseen by Copernicus, an independent review board (IRB) in the US (IRB approval #MAP2-11-470).

Consent to participate Written informed consent was obtained from a parent/guardian of all children who participated in the study and written and verbal assent was obtained from all of the children.

Consent for publication As part of the written informed consent form, all parent/guardian consented to study results being published in anonymised, aggregated form.

Data availability statement The data described in this article are not publicly available in further detail.

Code availability The code for the analyses described in this article is not publicly available.

Authors' contributions All authors contributed to the design of the study and development of the items and response options. RA, KB, and CM were responsible for conducting the research, including involvement in the study design, development of study documents, and data collection and interpretation. AR was responsible for the analysis of the data—he drafted the analysis plan (which was reviewed by all other authors), was involved in the conduct of all analysis, and contributed substantially to interpretation of the findings. PH, BZ, and CG contributed to the design of the study, reviewed all study documents, and were involved in the interpretation of the data. All authors were responsible for reviewing and revising the manuscript and have given approval for this version to be published.

References

1. Fashner J, Ericson K, Werner S. Treatment of the common cold in children and adults. Am Fam Phys. 2012;86(2):153.
2. Hsiao J, Cherry D, Beatty P, Rechtsteiner E. National Ambulatory Medical Care Survey: 2007 summary. National health statistics report; no 27. Hyattsville, MD: National Center for Health Statistics. 2010.
3. Medicines and Healthcare Products Regulatory Agency. Overview—Risk: benefit of over-the-counter (OTC) cough and cold medicines. London: MHRA; 2009.
4. Powers JH III, Bacci ED, Guerrero ML, et al. Reliability, validity, and responsiveness of influenza patient-reported outcome (FLU-PRO©) scores in Influenza-Positive patients. Value Health. 2018;21(2):210–8.
5. Powers JH III, Bacci ED, Leidy NK, et al. Performance of the in FLUenza patient-reported outcome (FLU-PRO) diary in patients with influenza-like illness (ILI). PLoS ONE. 2018;13(3):e0194180.
6. Obasi CN, Brown RL, Barrett BP. Item reduction of the Wisconsin Upper Respiratory Symptom Survey (WURSS-21) leads to the WURSS-11. Qual Life Res. 2014;23(4):1293–8.
7. Jacobs B, Young NL, Dick PT, et al. CARIFS: the Canadian acute respiratory illness and flu scale. Pediatr Res. 1999;45(7):103–103.
8. Osborne RH, Norquist JM, Elsworth GR, et al. Development and validation of the influenza intensity and impact questionnaire (FluiiQ™). Value in health. 2011;14(5):687–99.
9. Justicia-Grande AJ, Martinón-Torres F. The ReSVinet score for bronchiolitis: a scale for all seasons. Am J Perinatol. 2019;36(suppl S2):S48–53.
10. Justicia-Grande AJ P-SJ, Cebey-López M, et al. Development and validation of a new clinical scale for infants with acute respiratory infection: the ReSVinet scale. PLoS One. 2016;11(6):e0157665.
11. Scott J, Tatlock S, Kilgariff S, et al. Clinician (ClinRO) and Caregiver (ObsRO) reported severity assessments for respiratory syncytial virus (RSV) in infants and young children: Development and qualitative content validation of Pediatric RSV Severity and Outcome Rating Scales (PRESORS). In: Paper presented at: QUALITY OF LIFE RESEARCH2018. 2018.
12. Administration UFaD. Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. In: US Department of Health and Human Services Food and Drug Administration, December 2009.2013.
13. Halstead P, Arbuckle R, Marshall C, Zimmerman B, Bolton K, Gelotte C. Development and content validity testing of patient-reported outcome items for children to self-assess symptoms of the common cold. Patient. 2020;13:235–50.
14. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill; 1994.
15. Shaikh N, Martin JM, Casey JR, et al. Development of a patient-reported outcome measure for children with streptococcal pharyngitis. Pediatrics. 2009;124(4):e557–e56363.
16. Muhr T. ATLAS. ti. Berlin: Scientific Software Development. 1997.
17. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods. 2006;18:59–82.
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
19. Rao PV. Statistical research methods in the life sciences. California: Duxbury Press; 1998.
20. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. Stat Med. 1987;6(1):79–90.
21. Sullivan LM, D'Agostino R Sr. Robustness and power of analysis of covariance applied to ordinal scaled data as arising in randomized controlled trials. Stat Med. 2003;22(8):1317–34.
22. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. Control Clin Trials. 1991;12(4):S142–S158158.
23. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. Mahwah: Laurence Erlbaum Associates Inc, Publishers; 2001.
24. Kline RB. Principles and Practice of Structural Equation Modeling. New York: Guilford; 2005.
25. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56(2):81.
26. Pappas DE, Hendley JO, Hayden FG, Winther B. Symptom profile of common colds in school-aged children. Pediatr Infect Dis J. 2008;27(1):8–11.
27. Troullos E, Baird L, Jayawardena S. Common cold symptoms in children: results of an internet-based surveillance program. J Med Internet Res. 2014;16(6):e144.
28. Gelotte CK, Albrecht HH, Hynson J, Gallagher V. A multicenter, randomized, placebo-controlled study of pseudoephedrine for the temporary relief of nasal congestion in children with the common cold. The J Clin Pharmacol. 2019;59(12):1573–83.
29. Arbuckle R, Abetz-Webb L. "Not just little adults": qualitative methods to support the development of pediatric patient-reported outcomes. Patient. 2013;6(3):143–59.
30. Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. Value Health. 2013;16(4):461–79.
31. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. Abingdon: Routledge; 2013.
32. Sloan J, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. Drug Inf J. 2003;37(1):23–31.
33. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998;16(1):139–44.
34. Cohen J. Statistical power analysis for the behavioral sciences (rev. New Jersey: Lawrence Erlbaum Associates Inc; 1977.
35. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989;27(3):S178–S189189.

Affiliations

Rob Arbuckle[1] · Patricia Halstead[2] · Chris Marshall[1] · Brenda Zimmerman[2] · Kate Bolton[1] · Antoine Regnault[3] · Cathy Gelotte[2]

Patricia Halstead
halstead@comcast.net

Chris Marshall
cmarshall@teamdrg.com

Brenda Zimmerman
bbzimmer@comcast.net

Kate Bolton
kate.skivington@jointhedotsmr.com

Antoine Regnault
antoine.regnault@modusoutcomes.com

Cathy Gelotte
cathy.k.gelotte@gmail.com

[1] Adelphi Values, Adelphi Mill, Grimshaw Lane, Bollington SK10 5JB, CH, UK

[2] McNeil Consumer Healthcare, a Division of Johnson and Johnson Consumer Inc., 7050 Camp Hill Rd, Fort, Washington, PA 19034, USA

[3] Modus Outcomes, 61 Cours de La Liberté, 69003 Lyon, France