Ther Adv Respir Dis
2020, Vol. 14: 1–15
DOI: 10.1177/1753466620915155

Quantitative measurement properties and score interpretation of the Cough Severity Diary in patients with chronic cough

Allison Martin Nguyen, Elizabeth Bacci, Peter Dicpinigaitis and Margaret Vernon

Abstract

Aims: The Cough Severity Diary (CSD) was developed in accordance with the FDA guidance for patient-reported outcome measures and is focused on capturing the patient's perception of cough in terms of frequency, intensity, and disruption due to their cough. The measure includes a series of seven items asking patients to rate the frequency (three items), intensity (two items), and disruptiveness (two items) of their cough. The instrument was designed to be completed daily before bedtime, has a recall period of 'today,' and responses to items are entered on an 11-point numeric rating scale ranging from 0 to 10 with anchors on each end. The objective of this analysis was to confirm the domain structure of the CSD and assess its reliability, validity, and responsiveness in adult patients with refractory or unexplained chronic cough (RCC/UCC). Criteria for defining meaningful changes in mean weekly CSD total and domain scores in the context of a clinical trial were also developed.

Methods: Pooled data from a phase II randomized controlled trial of an investigational treatment for RCC/UCC were analyzed. Participants were non-smokers, had RCC/UCC for ⩾1 year, and a baseline cough severity visual analogue scale (VAS) ⩾40 mm. CSD scores (baseline, week 4), were analyzed; the Leicester Cough Questionnaire (LCQ), cough severity VAS, Patient Global Impression of Change (PGIC), and objective cough frequency counts were used for validation. CSD domain structure (Total, Frequency, Intensity, Disruption) was assessed for scoring.

Results: A total of 253 participants were included (mean age 60.2; 76% female). Global fit of the three-factor CSD was acceptable. For the CSD total score, internal consistency (α=0.89) and test–retest reliability (intraclass correlation coefficient=0.68) were high. CSD total scores were correlated with the LCQ total (r=–0.62) and cough severity VAS (r=0.84). Participants with a PGIC score of 1 or 2 (most improved groups) had the greatest mean score improvement on the CSD Total (Day 0 to Day 28), supporting responsiveness (similar findings for subscales). A change threshold of ⩾1.3-point reduction on the total and subscale scores is appropriate to define clinically meaningful improvement.

Conclusion: The CSD is a reliable, valid, and responsive measure of cough symptom severity in patients with refractory or unexplained chronic cough and fit-for-purpose for assessing changes in cough severity in clinical trials.

The reviews of this paper are available via the supplemental material section.

Keywords: chronic cough, clinically meaningful change, Leicester Cough Questionnaire, patient reported outcomes

Received: 27 August 2019; revised manuscript accepted: 18 February 2020.

Correspondence to:
Allison Martin Nguyen
Merck Sharp & Dohme Corp., 770 Sumneytown Pike, West Point, PA 19486 USA
allison_martin@merck.com

Elizabeth Bacci
Margaret Vernon
Evidera Inc, Bethesda, MD, USA

Peter Dicpinigaitis
Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA

Introduction

Chronic cough (CC), defined as a cough lasting longer than 8 weeks,[1] is one of the most common reasons people in the United States (US) seek medical attention.[2,3] The global prevalence of CC is approximately 9%.[4] While CC is often secondary

to other conditions like chronic obstructive pulmonary disease and gastroesophageal reflux disease, a significant proportion of individuals experience CC in the absence of another identified disorder [i.e. unexplained CC (UCC)] or when an underlying condition is adequately treated [i.e. refractory CC (RCC)].[5] Patients who experience RCC/UCC have been shown to have negatively impacted health-related quality of life (HRQL), depressed mood, and inability to take part in daily activities,[6,7] in addition to social disruptions[8] and depression.[9]

A number of tools are available to assess various aspects of cough, including objective cough frequency, cough severity, and impact on HRQL. The current gold standard for assessing the efficacy of cough therapies in clinical trials is objective cough frequency.[10] The assessment of cough frequency, however, is largely restricted to clinical research settings as the method relies on the use of recording devices and manual counting of coughs from a 24-hour recording; a process that requires extensive training and specialty equipment/software and is labor-intensive and time-consuming. In the setting of clinical practice, guidelines often recommend the use of validated and reliable patient-reported outcome (PRO) instruments to assess effectiveness of interventions.[11] These instruments may be more easily administered and can evaluate the impact of cough on the patients' life.

While various PROs have been used in therapeutic trials to assess the impact of CC on HRQL, such as the Leicester Cough Questionnaire (LCQ) and the Cough-Specific Quality of Life Questionnaire (CQLQ), few measures have been developed to assess cough symptom severity. Moreover, only one PRO instrument, the Cough Severity Diary (CSD)[12,13] has been developed to measure cough symptom severity following standards for PRO development put forth by the Medical Outcomes Trust[14] and the US Food and Drug Administration (FDA),[15] in order to be suitable for use as an endpoint in clinical programs.[12] The various cough assessment tools that have been developed measure distinct yet related aspects of the impact cough has on a patient. As such, these tools complement one another and are frequently used in combination.[16] A PRO measuring cough symptom severity may play an important role in understanding the impact of cough management strategies on HRQL, particularly in clinical practice settings where cough frequency monitoring is not feasible. The content of the CSD was developed through a qualitative study in which three focus groups were conducted with patients with CC.[12] Patients indicated that cough symptom severity could be described and assessed through three domains, including: frequency, intensity, and disruption. Following the qualitative study, a small quantitative pilot study of the CSD was conducted with patients with chronic and subacute cough.[13] Confirmatory factor analysis (CFA) supported the three domains identified in the qualitative study. The CSD also demonstrated acceptable internal consistency, test–retest reliability, and construct validity; known-groups validity and responsiveness were not assessed. The sample size in this pilot study was small, thus further investigation in a larger sample is warranted, utilizing a similar patient population and assessments.

The objectives of this study were to confirm the domain structure of the CSD and assess the reliability, validity, and responsiveness in a larger study of adult patients with RCC/UCC Reliability, validity, and responsiveness were assessed as follows: CFA to evaluate domain structure, test–retest reliability (or stability) of mean weekly scores over time, stability of CSD scores in participants reporting 'no change' on the patient global impression of change (PGIC) questionnaire, convergent validity evaluated by comparing the CSD to other well established instruments [such as the LCQ, awake cough frequency, cough visual analog scale (VAS)], known-groups validity to determine whether the CSD can discriminate between groups known to differ on the LCQ and awake cough frequency, responsiveness from baseline to day 28 as measured by the PGIC scale, and several methods to evaluate meaningful change (or response) to treatment.

Materials and methods

Study sample

This study was a secondary *post hoc* analysis of a phase II, multicenter, multinational, placebo-controlled, double-blind trial conducted to assess the efficacy of three doses (7.5 mg BID, 20 mg BID, 50 mg BID) of MK-7264 or matching placebo, in reducing awake objective cough frequency in men and women 18–80 years with treatment of RCC or UCC [ClinicalTrials.gov identifier: NCT02612610]. Patients underwent screening and baseline (day 0) assessments and took the study drug every 12 h for days 1–84. Data used for this analysis were from screening

(day –14 to day –1) to day 28 and were pooled across treatment groups.

Preliminary eligibility criteria for enrollment included: no abnormality that could significantly contribute to the CC upon the conduct of a chest radiograph or computed tomography of the thorax within 5 years of assessed eligibility; diagnosis of RCC or unexplained cough for at least 1 year; and a score of ⩾40 on the cough severity VAS at screening.

Outcome measures

Cough severity diary. The CSD was designed to capture patients' assessments of cough severity through a series of seven items asking them to rate the frequency (Items 1–3), intensity (Items 4 and 5), and disruption (Items 6 and 7) of their cough. The CSD is intended to be completed in the evening prior to going to bed and has a recall period of 'today'. Responses to items are entered on an 11-point numeric rating scale (NRS) ranging from 0 to 10 with higher scores indicating greater severity. Specifically, the first item asks the participant to indicate how often he/she coughed today with NRS anchors of 'never' to 'constantly'; the second and third questions ask the participant to indicate how often their cough turned into a coughing fit ('never' to 'always') and how often they had an urge to cough today ('never' to 'constantly'); the fourth and fifth items ask the participant to rate how harsh their cough was today ('not at all' to 'extremely') and how much physical discomfort they had because of their cough today ('none' to 'extreme'); finally, items six and seven ask the participant to indicate how much their cough disrupted their activities today ('not at all' to 'could not perform activities') and how much their cough disrupted their sleep last night ('not at all' to 'could not sleep at all'). The CSD was originally developed as a paper diary and subsequently migrated to an electronic version implemented as an interactive web response system and Interactive voice response system in this study.

The total score was calculated by averaging over all seven items each day, with domain scores calculated as the mean across items within each domain. If one item was missing from a subscale, the other item(s) were used to calculate the subscale score. If more than one item was missing, the subscale score was also missing. If a subscale score was missing, the total score was also missing. Weekly mean scores were calculated if at least one day of data was completed.

Leicester Cough Questionnaire. The LCQ is a 19-item cough-specific HRQL PRO instrument which contains three domains (physical, psychological, and social), calculated as a mean score for each domain ranging from 1 to 7, and a total score calculated as the sum of the domain scores (ranging from 1 to 21). Each item on the LCQ assesses symptoms or the impact of symptoms on HRQL over the past 2 weeks using a 7-point Likert scale ranging from 1 ('all the time') to 7 ('none of the time'). Higher scores indicate better HRQL.[17] The LCQ was administered on paper at baseline, and at weeks 4, 8, and 12 during the study.

Cough severity visual analog scale and patient global impression of change PROs. The cough severity VAS is a single-item measure of the patient's perception of cough severity today'. It is scored on a 100-mm scale, where the anchors range from 'Not at all' to 'Extremely'.[18] The Cough VAS was completed on paper at baseline, and at weeks 4, 8, and 12 during the study.

The PGIC measures change is a self-reported instrument of the patient's overall improvement on a seven-point scale, ranging from 1 (very much improved) to 7 (very much worse).[19] The PGIC was completed on paper at weeks 4, 8, and 12 during the study.

Awake objective cough frequency. To determine cough frequency, a 24-hour ambulatory sound recorder (VitaloJAK, Vitalograph, Buckingham, UK), was used. Digital recordings were processed in a centralized reading center, where recordings were condensed using a computer algorithm before human analysts identified and tagged individual coughs. The output of this process was a count of coughs for each 24-hour recording period, as well as cough counts for awake and asleep portions of the day. The objective cough frequency utilized in the current analyses included only coughs recorded during awake periods.

Statistical analyses

Instrument structure

Sample and CSD descriptive statistics. Descriptive statistics [mean, standard deviation (SD), frequencies for categorical data] were calculated for demographic and clinical variables at baseline (day 0), as well as for the individual items, domain, and total score of the CSD. In addition, the percentage of CSD items exhibiting ceiling

and floor effects were calculated. An item was considered to exhibit a floor or ceiling effect if the minimum or maximum response was selected by >25% of participants, respectively.

Confirmatory factor analysis. CFA at baseline (day 0) was conducted using a structural equation modeling (SEM) approach to confirm the three-factor model proposed by Vernon and colleagues.[13] In the context of SEM, several fit statistics were used to assess the adequacy of the model to explain the data. In general, a model explains the data well if the comparative fit index (CFI) is ⩾0.9. The standardized root mean residual (SRMR) measures the mean absolute difference between observed and model-implied correlations; values of <0.1 were considered acceptable.[20] Finally, the root mean square error of approximation (RMSEA) is a measure of fit assessing the discrepancy between the predicted and observed data per degree of freedom; values <0.08 are considered acceptable.[21] Adequacy of fit was also assessed through examination of the modification indices. Factor loadings of ⩾0.40 were considered acceptable.

Analysis of measurement properties

Internalconsistency and test–retest reliability. Internal consistency reliability addresses the extent to which individual items within each scale are related to each other and with the scale as a whole. Internal consistency (Cronbach's alpha) reliability of the CSD total and domain scores was calculated at baseline (day 0). In addition, the test–retest reliability (or stability) of the mean weekly CSD total and domain scores over time was evaluated in the subset of participants categorized as 'stable' participants. Stability was defined as participants reporting 'no change' on the PGIC from baseline (day 0) to day 28. Change scores using paired *t*-tests, intraclass correlation coefficients (ICCs) and Pearson's *r* correlations were calculated between the baseline (day –6 to day 0) mean weekly CSD scores and retest mean weekly CSD scores at week 4 (day 22 to day 28).

Convergent validity. Convergent validity involves demonstrating that different measures of the same concept substantially correlate. Convergent validity was assessed for the mean weekly CSD domain and total scores by relating measures to the CSD that measure constructs similar to cough severity. The mean weekly CSD total and subscale scores at baseline were correlated (Pearson's *r*) with the LCQ, awake objective cough frequency, and VAS cough severity at baseline (day 0). Moderate to large relationships ($r > 0.30$) with the CSD were expected for all measures.

Known-groups validity. Known-groups validity is the extent to which scores from an instrument are distinguishable from groups of subjects that differ by relevant clinical or other indicators. Given there is no 'gold standard' of cough disease severity, two different anchors were used, including the LCQ and awake objective cough frequency. To assess known-groups validity of the mean weekly CSD domain and total scores at baseline, participants were stratified into tertiles using the sample distribution, according to the LCQ total score at baseline (day 0). This analysis was replicated, wherein participants were stratified into tertiles according to awake objective cough frequency. Analysis of variance was used for these analyses, with *post hoc* category comparisons *via* Scheffe's test to determine whether CSD scores discriminate between LCQ and awake objective cough frequency groups.

Responsiveness. Responsiveness is a type of validity and refers to the extent to which the instrument can detect true change in patients known to have changed in clinical status.[22] To evaluate the responsiveness of the CSD, analysis of covariance was used to compare change in mean weekly CSD domain and total scores, controlling for baseline CSD scores, from baseline to week 4, by response on the PGIC at day 28. This analysis was replicated using percentage change in awake objective cough frequency, wherein participants were considered responders/non-responders using four definitions of response: (a) ⩾30% reduction; (b) ⩾50% reduction; (c) ⩾70% reduction; and (d) using a distribution-based approach, participants with a change of ⩾0.30 SD. As recommended by the FDA,[15] effect size (ES) was calculated for the mean weekly CSD domain and total scores for the groups as defined above. Effect size is a measure of change that provides a means of standardizing the quantification for comparison between groups.[23] Effect size was calculated by subtracting the mean baseline score from the mean week 4 score and dividing by the baseline SD. Effect size was interpreted as small (0.20), moderate (0.50), or large (0.80) following the guidelines proposed by Cohen.[24]

Score interpretation. To evaluate what would constitute a meaningful change in the CSD and

define responders to treatment in a clinical trial setting, both anchor- and distribution-based approaches were used. For the anchor-based analysis, mean change in mean weekly CSD scores from baseline to the mean of week 4 was categorized by PGIC category assessed at day 28. The PGIC was grouped into five categories: 1 and 2 ('Very much improved' and 'Much improved'); 3 ('Minimally improved'); 4 ('No change'); 5 ('Minimally worse'); and 6 and 7 ('Much worse' and 'Very much worse'). The minimally important change threshold was defined as the mean CSD change score of patients who reported 'Somewhat improved' or 'Somewhat worse' on the PGIC.

In addition to the anchor-based approach for score interpretation, a distribution-based approach was used for interpreting the CSD. Specifically, the standard error of measurement (SEm) was calculated.[25,26] The SEm is estimated by multiplying the baseline standard deviation of the measure by the square root of one minus its reliability coefficient (ICC from the test–retest assessment). Shikiar and colleagues[27] found that there is a general correspondence between a meaningful change threshold and SEm; however, it is dependent upon the magnitude of the reliability coefficient. A second distribution-based approach included an assessment of half of a standard deviation of the CSD total and subscale scores at baseline (day 0). Norman and colleagues[28] suggest that half of a standard deviation of a measure represents a good approximation of the meaningful change threshold.

Finally, receiver operating characteristic (ROC) curves were evaluated to help determine the threshold values for change in mean weekly CSD total and subscale scores from baseline (day –6 to day 0) to the mean of day 22 to day 28 with the best sensitivity and specificity for predicting patients scoring a 1, 2 or 3 *versus* 4, 5, 6, or 7 on the PGIC. Positive and negative predictive values were also calculated. Youden's index was used to determine the change score on the CSD which optimized sensitivity and specificity in predicting global improvements based on the PGIC.[29]

Results

Instrument structure

Sample and CSD descriptive statistics. A total of 253 patients were included in the study. Most participants were female (76.3%), and White (92.9%), with a mean age of 60.2 (SD = 9.9). All participants (100%) had a history of CC, and over half (57.7%) had been treated for CC 1 year prior to screening. Almost all participants (97.6%) had used medication to treat their cough within 30 days of screening (Table 1).

Descriptive statistics for the CSD items and domain scores are shown in Table 2. At Baseline (day 0), the CSD item mean (SD) scores ranged from 2.6 (2.6) for Item 6 ('disrupt your activities') to 5.6 (2.2) for Item 3 ('urge to cough'). The full range of scores (0–10) was represented for all items at baseline, except for Item 1 ('often cough today'). No ceiling effects were observed, but floor effects were observed for Item 7 ('disrupt your sleep'), with the minimum response selected by 29% of participants.

Confirmatory factor analysis. The overall fit of the three-factor CSD (including frequency, intensity, and disruption factors) was found to be acceptable according to the CFI and SRMR values (0.905 and 0.048), respectively. However, the RMSEA value was slightly larger than expected (0.217, 95% confidence interval = 0.185, 0.251). All item factor loadings were found acceptable for the frequency (0.775–0.848), intensity (0.877–0.880), and disruption (0.672–0.934) subscales.

Analysis of measurement properties

Internal consistency and test–retest reliability. Cronbach's alpha (α) was high at baseline (day 0) for the CSD Total score ($\alpha = 0.923$), as well as the frequency ($\alpha = 0.839$), intensity ($\alpha = 0.869$), and disruption ($\alpha = 0.771$) domains.

For the assessment of test–retest reliability, 60 participants were considered stable from baseline (day 0) to day 28 and included in the analysis (Table 3). Acceptable test–retest reliability was demonstrated for the mean weekly CSD total and domain scores *via* the assessment of the ICCs (ICCs = 0.90, total score; 0.87–0.92 for domain scores) and Pearson's r values ($r = 0.92$, total score; 0.89–0.93 for domain scores).

Convergent validity. The CSD mean weekly total and domain scores were significantly correlated with conceptually similar measures of cough severity and frequency (Table 4). Moderate-to-large correlations were seen with related measures,

Table 1. Sociodemographic and clinical characteristics at baseline (day 0).

Characteristics	Total *n* = 253
Age, mean (SD)	60.2 (9.9)
Gender *n* (%)	
Male	60 (23.7%)
Female	193 (76.3%)
Ethnicity *n* (%)	
Hispanic or Latino	3 (1.2%)
Not Hispanic or Latino	250 (98.8%)
Race *n* (%)	
White	235 (92.9%)
Black	12 (4.7%)
Asian	3 (1.2%)
Other	3 (1.2%)
Chronic cough history	253 (100.0%)
Medications used 30 days prior to screening	247 (97.6%)
Chronic cough treatments 1 year prior to screening	146 (57.7%)

SD, standard deviation.

including the LCQ total score (r = –0.59 to –0.64) and LCQ subscale scores (r = –0.44 to –0.61), and cough severity VAS (r = 0.42 to 0.57), all significant at $p < 0.0001$.

Known-groups validity. The results of the known-groups validity analysis of mean weekly CSD scores utilizing the LCQ to define known-groups are found in Table 5. In support of known-groups validity, the CSD Total and domain scores were highest (worse) in patients in the lowest LCQ score group, indicating reduced quality of life; CSD scores decreased (improved) in correspondence with improvement in LCQ scores. For example, for the CSD total score, the mean (SE) CSD scores were 7.0 (0.32), 4.5 (0.13), and 3.1 (0.17) for the lowest (total scores of 3–8), middle (total scores of 9–13), and highest (total scores of 14–21) LCQ tertile groups, respectively. Similar results were found when investigating known-groups validity using awake objective cough frequency to define known-groups. As shown in Table 6, CSD Total and domain scores were lowest in the lowest objective cough frequency group, and increased for the two higher groups, as expected.

Responsiveness. Responsiveness of mean weekly CSD Total and subscale scores was supported when using the PGIC as an anchor. Specifically,

Table 2. CSD[1] descriptive statistics at baseline (day 0) (n = 239).

Item/domain	Mean (SD)	Range	*n*, % floor[1]	*n* % ceiling[1]
Item 1	5.4 (2.0)	1.0–10.0	0 (0.0%)	5 (2.1%)
Item 2	3.6 (3.6)	0.0–10.0	40 (16.7%)	6 (2.5%)
Item 3	5.6 (2.2)	0.0–10.0	2 (0.8%)	12 (5.0%)
Item 4	4.9 (2.4)	0.0–10.0	4 (1.7%)	7 (2.9%)
Item 5	3.9 (2.6)	0.0–10.0	23 (9.6%)	6 (2.5%)
Item 6	3.2 (2.6)	0.0–10.0	47 (19.7%)	3 (1.3%)
Item 7	2.6 (2.6)	0.0–10.0	69 (28.9%)	3 (1.3%)
Total Score	4.2 (2.1)	0.3–9.6	0 (0.0%)	0 (0.0%)
Frequency	4.9 (2.1)	0.7–10.0	0 (0.0%)	2 (0.8%)
Intensity	4.4 (2.4)	0.0–10.0	2 (0.8%)	4 (1.7%)
Disruption	2.9 (2.4)	0.0–10.0	27 (11.3%)	1 (0.4%)

[1]Floor = minimum response >25%; ceiling = maximum response >25%.
CSD, Cough Severity Diary; SD, standard deviation.

Table 3. Test–retest reliability (reproducibility) of CSD mean weekly scores: patients reporting no change on PGIC from baseline (day 0) to day 28.

CSD Total/subscales	*n*	Baseline mean (SD)	Visit 4 mean (SD)	Difference[1]	*p* value	Pearson's r^2	ICC
Total Score	60	4.3 (1.92)	4.0 (2.02)	−0.3	0.0016	0.92	0.90
Frequency	60	5.2 (1.87)	4.8 (2.06)	−0.4	0.0023	0.89	0.87
Intensity	60	4.5 (2.14)	4.2 (2.27)	−0.3	0.0078	0.90	0.89
Disruption	60	2.9 (2.36)	2.7 (2.33)	−0.3	0.0210	0.93	0.92

[1]Difference = Mean of visit 4 (day 22 to day 28) – Mean of baseline (day −6 to day 0).
[2]Pearson's product-moment correlation.
CSD, Cough Severity Diary; ICC, intraclass correlation coefficient; PGIC, Patient Global Impression of Change; SD, standard deviation.

Table 4. Pearson's correlations between mean weekly CSD scores at baseline (day −6 to day 0) and conceptually related measures at baseline (day 0).

Measure at baseline	Cough severity weekly score at baseline			
	Total	Frequency	Intensity	Disruption
LCQ total score	−0.64[a]	−0.59[a]	−0.59[a]	−0.61[a]
LCQ physical	−0.59[a]	−0.52[a]	−0.53[a]	−0.61[a]
LCQ psychological	−0.48[a]	−0.47[a]	−0.44[a]	−0.43[a]
LCQ social	−0.60[a]	−0.55[a]	−0.57[a]	−0.58[a]
Cough severity visual analogue scale	0.53[a]	0.57[a]	0.50[a]	0.42[a]

[a]Pearson's correlation coefficients $p < 0.0001$.
CSD, Cough Severity Diary; LCQ, Leicester Cough Questionnaire.

Table 5. Known-groups validity: mean weekly CSD score at baseline (day −6 to day 0) by LCQ total score groups at baseline (day 0).

CSD total/ subscales	Leicester Cough Questionnaire Group						Overall *F* value	*p* value
	Score 3–8		Score 9–13		Score 14–21			
	n	Mean (SE)	*n*	Mean (SE)	*n*	Mean (SE)		
Total Score	24	7.0 (0.32)	142	4.5 (0.13)	85	3.1 (0.17)	61.47	<0.0001
Frequency	24	7.4 (0.31)	142	5.2 (0.13)	85	3.9 (0.17)	51.78	<0.0001
Intensity	24	7.3 (0.37)	142	4.8 (0.15)	85	3.3 (0.20)	47.37	<0.0001
Disruption	24	6.1 (0.38)	142	3.3 (0.16)	85	1.8 (0.20)	53.00	<0.0001

CSD, Cough Severity Diary; LCQ, Leicester Cough Questionnaire.

participants with a PGIC score of 1 or 2 (the most improved groups) had the greatest mean score change on the CSD Total (–2.7; range –6.9–0.2), and a large effect size (Cohen's $\alpha = -1.5$; Table 7). Participants with a PGIC score of 3 ('Minimally improved') had a mean change score of –1.3

Table 6. Known-groups validity: mean weekly CSD score at baseline (day –6 to day 0) by awake objective cough frequency groups at baseline (day 0).

CSD total/subscales	Objective cough frequency group						Overall *F* value	*p* value
	0–33 Percentile		34–66 percentile		67–100 percentile			
	n	Mean (SE)	*n*	Mean (SE)	*n*	Mean (SE)		
Total Score	83	3.3 (0.19)	83	4.4 (0.19)	83	5.2 (0.19)	23.74	<0.0001
Frequency	83	3.9 (0.18)	83	5.1 (0.18)	83	5.9 (0.18)	28.68	<0.0001
Intensity	83	3.5 (0.22)	83	4.6 (0.22)	83	5.5 (0.22)	21.86	<0.0001
Disruption	83	2.2 (0.23)	83	3.1 (0.23)	83	3.8 (0.23)	11.67	<0.0001

CSD, Cough Severity Diary.

(range –5.8–1.3), and a large effect size (–0.8). Mean score change on the CSD total score was smaller with each subsequent PGIC score category group, with corresponding smaller effect sizes for each group. A similar pattern was found for each of the CSD subscales.

Awake objective cough frequency was also used as an anchor to assess responsiveness of mean weekly CSD scores. When using various thresholds for determining response for awake objective cough frequency (i.e. ⩾30%, ⩾50%, ⩾70% reduction and reduction ⩾0.30 SD), mean score changes and effect sizes on the CSD total score were always larger for those considered responders compared with non-responders using awake objective cough frequency responder thresholds (Table 8). Similar findings were evident for the CSD subscale scores. Taken together, these findings provide strong support for the responsiveness of the CSD.

Score interpretation. Using an anchor-based approach, change in mean weekly CSD scores associated with a PGIC score of 3, 'Minimally improved', were –1.3 for the CSD Total score and ranged from –1.4 to –1.1 for domain scores (Table 7). Distribution-based estimates (SEm and one-half SD) were similar, but smaller, than the anchor-based estimates for the CSD, with distribution-based estimates ranging from 0.95 to 0.99 for the CSD total score, and estimates ranging from 0.91 to 1.11 for the domain scores (Table 9). Lastly, results of the ROC analysis demonstrated that a mean (SD) change of –1.3 (–1.2) on the CSD total score was associated with the largest Youden's Index score (maximizing sensitivity and specificity) in predicting a PGIC rating of at least 'Minimally improved' while a mean (SD) change of –2.5 (–1.5) on the CSD total score was most predictive of PGIC a rating of at least 'Improved' (Table 10). After considering the findings from multiple methods, it is proposed that a change threshold of a ⩾1.3-point reduction on the CSD total and subscale scores is appropriate to define clinically meaningful improvement.

Discussion

The present findings build upon the reliability and construct validity of the CSD found in the initial pilot study investigation of patients with chronic and subacute cough.[12] There has been extensive discussion in the literature regarding the merits and limitations of the various PRO's used to evaluate cough in clinical settings.[10,11,18,30,31] Objective quantification of cough is important for the evaluation of antitussive therapies, however, evaluation of symptom severity and quality of life through PRO's reflect important considerations from the patients' perspective.[31] Accordingly, the American College of Chest Physicians recommends that validated and reliable health-related quality-of-life questionnaires be used as the measurement of choice to assess the impact of cough.[11] Multiple measures have been developed to assess the impact on cough on HRQL, such as the LCQ,[17] the CQLQ,[32] and the Chronic Cough Impact Questionnaire.[33] Other generic PRO measures of depression, anxiety, fatigue, and mood have been used to better understand the burden of disease

Table 7. Responsiveness of mean weekly CSD scores: CSD total and domain scores from baseline (day –6 to day 0) to visit 4 (day 22 to day 28) by PGIC at day 28.

CSD total/subscale, patient global impression of change (PGIC) category	*n*	Baseline mean (SD)	Visit 4 mean (SD)	Mean change score[1]		Effect size[2]
				Difference	Range	
Total Score						
PGIC Score 1 & 2	86	4.0 (1.84)	1.4 (1.16)	–2.7	–6.9–0.2	–1.5
PGIC Score 3	77	4.2 (1.64)	2.9 (1.63)	–1.3	–5.8–1.3	–0.8
PGIC Score 4	60	4.3 (1.92)	4.0 (2.02)	–0.3	–2.7–1.5	–0.2
PGIC Score 5	5	7.2 (1.84)	7.4 (1.59)	0.3	–0.2–1.4	0.1
PGIC Score 6 & 7	4	3.9 (2.36)	5.1 (2.18)	1.1	0.4–2.2	0.5
Frequency						
PGIC Score 1 & 2	86	4.7 (1.74)	1.7 (1.23)	–2.9	–7.10.7	–1.7
PGIC Score 3	77	4.8 (1.57)	3.5 (1.65)	–1.3	–5.4–1.1	–0.8
PGIC Score 4	60	5.2 (1.87)	4.8 (2.06)	–0.4	–2.7–2.7	–0.2
PGIC Score 5	5	7.5 (1.44)	7.6 (1.39)	0.1	–0.4–0.9	0.1
PGIC Score 6 & 7	4	4.6 (2.02)	6.1 (1.43)	1.5	0.7–2.2	0.7
Intensity						
PGIC Score 1 & 2	86	4.3 (2.13)	1.4 (1.40)	–2.9	–7.3–0.3	–1.4
PGIC Score 3	77	4.4 (1.88)	3.0 (1.77)	–1.4	–6.2–1.0	–0.8
PGIC Score 4	60	4.5 (2.14)	4.2 (2.27)	–0.3	–2.7–2.2	–0.2
PGIC Score 5	5	7.3 (1.73)	7.8 (1.20)	0.5	–0.3–1.6	0.3
PGIC Score 6 & 7	4	4.1 (2.94)	5.0 (3.25)	0.9	–0.6–2.6	0.3
Disruption						
PGIC Score 1 & 2	86	2.8 (2.12)	0.8 (1.05)	–2.0	–7.4–0.8	–0.9
PGIC Score 3	77	3.1 (1.91)	1.9 (1.76)	–1.1	–6.4–2.1	–0.6
PGIC Score 4	60	2.9 (2.36)	2.7 (2.33)	–0.3	–2.8–1.6	–0.1
PGIC Score 5	5	6.5 (2.78)	6.8 (2.75)	0.3	–0.8–1.8	0.1
PGIC Score 6 & 7	4	2.8 (2.43)	3.6 (2.50)	0.8	0.1–1.8	0.3

[1]Calculated as weekly mean of visit 4 (day 22 to day 28) minus baseline (day –6 to day 0).
[2]Calculated as score difference/SD of baseline score.
CSD, Cough Severity Diary; PGIC, Patient Global Impression of Change; SD, standard deviation.

associated with CC.[9,34] However, unlike these instruments that focus on HRQL and the impacts of cough, the CSD is different as it is the first patient-reported tool that directly assesses dimensions of cough severity important to patients, including frequency of cough, intensity, and disruption to sleep and activities. Thus, the development of the CSD complements measures of HRQL and objective cough frequency, as it helps to describe CC as experienced by the patient.

Table 8. Responsiveness of mean weekly CSD scores: CSD total from baseline (day −6 to day 0) to visit 4 (day 22 to day 28) by awake objective cough frequency change.

CSD total/subscale, awake objective cough frequency change	*n*	Baseline mean (SD)	Visit 4 mean (SD)	Mean change score[1]		Effect size[2]
				Difference	Range	
Total						
⩾30% Reduction	124	4.0 (1.71)	2.0 (1.59)	−2.1	−6.9–1.4	−1.2
<30% Reduction	101	4.5 (2.02)	3.7 (2.13)	−0.8	−5.8–2.2	−0.4
⩾50% Reduction	78	4.1 (1.71)	1.6 (1.15)	−2.5	−6.9–0.9	−1.5
<50% Reduction	147	4.3 (1.94)	3.4 (2.13)	−0.9	−5.8–2.2	−0.5
⩾70% Reduction	51	4.0 (1.64)	1.2 (1.04)	−2.8	−6.9–0	−1.7
<70% Reduction	174	4.3 (1.92)	3.2 (2.04)	−1.1	−5.8–2.2	−0.6
Reduction ⩾ 0.3 SD	116	4.4 (1.77)	2.3 (1.93)	−2.1	−6.9–1.0	−1.2
Reduction < 0.3 SD	109	4.1 (1.95)	3.3 (2.05)	−0.8	−5.8–2.2	−0.4
Frequency						
⩾30% Reduction	124	4.6 (1.63)	2.4 (1.64)	−2.2	−7.1–1.0	−1.4
<30% Reduction	101	5.2 (1.90)	4.4 (2.15)	−0.8	−5.4–2.2	−0.4
⩾50% Reduction	78	4.7 (1.59)	2.0 (1.25)	−2.7	−7.1–0.5	−1.7
<50% Reduction	147	5.0 (1.86)	4.0 (2.17)	−1.0	−5.4–2.2	−0.5
⩾70% Reduction	51	4.5 (1.54)	1.6 (1.09)	−2.9	−7.1–0	−1.9
<70% Reduction	174	5.0 (1.83)	3.8 (2.09)	−1.2	−5.4–2.2	−0.6
Reduction ⩾ 0.3 SD	116	5.0 (1.64)	2.8 (1.99)	−2.3	−7.1–1.4	−1.4
Reduction < 0.3 SD	109	4.8 (1.90)	3.9 (2.13)	−0.9	−5.4–2.2	−0.5
Intensity						
⩾30% Reduction	124	4.3 (1.95)	2.0 (1.78)	−2.3	−7.3–1.6	−1.2
<30% Reduction	101	4.7 (2.28)	3.9 (2.37)	−0.8	−5.9–2.6	−0.4
⩾50% Reduction	78	4.4 (1.93)	1.6 (1.37)	−2.8	−6.7–0.9	−1.4
<50% Reduction	147	4.5 (2.20)	3.5 (2.35)	−1.0	−7.3–2.6	−0.5
⩾70% Reduction	51	4.3 (1.89)	1.2 (1.23)	−3.1	−6.7–0.2	−1.7
<70% Reduction	174	4.5 (2.17)	3.4 (2.26)	−1.2	−7.3–2.6	−0.5
Reduction ⩾ 0.3 SD	116	4.7 (2.02)	2.4 (2.18)	−2.4	−7.3–1.6	−1.2
Reduction < 0.3 SD	109	4.2 (2.17)	3.4 (2.23)	−0.8	−5.9–2.6	−0.4

(Continued)

Table 8. (Continued)

CSD total/subscale, awake objective cough frequency change	*n*	Baseline mean (SD)	Visit 4 mean (SD)	Mean change score[1]		Effect size[2]
				Difference	Range	
Disruption						
⩾30% Reduction	124	2.8 (1.99)	1.2 (1.55)	−1.6	−7.4–1.8	−0.8
<30% Reduction	101	3.2 (2.43)	2.6 (2.35)	−0.7	−6.4–2.1	−0.3
⩾50% Reduction	78	2.9 (2.04)	0.9 (1.05)	−2.0	−7.4–1.4	−1.0
<50% Reduction	147	3.1 (2.29)	2.3 (2.29)	−0.7	−6.4–2.1	−0.3
⩾70% Reduction	51	2.8 (1.95)	0.7 (1.00)	−2.1	−7.4–0.0	−1.1
<70% Reduction	174	3.1 (2.27)	2.2 (2.17)	−0.9	−6.4–2.1	−0.4
Reduction ⩾ 0.3 SD	116	3.1 (2.16)	1.5 (1.86)	−1.7	−7.4–1.6	−0.8
Reduction < 0.3 SD	109	2.9 (2.25)	2.2 (2.20)	−0.6	−6.4 – 2.1	−0.3

[1]Calculated as weekly mean of visit 4 (day 22 to day 28) minus baseline (day −6 to day 0).
[2]Calculated as score difference/SD of baseline score.
CSD, Cough Severity Diary; SD, standard deviation.

Table 9. Distribution based meaningful change threshold for CSD.

CSD total/subscales	Mean (SD) at baseline	1/2 SD[1]	SEm[2]
Total Score	4.3 (1.90)	0.95	0.99
Frequency	4.9 (1.82)	0.91	1.03
Intensity	4.5 (2.15)	1.07	1.10
Disruption	3.0 (2.22)	1.11	0.93

[1]Norman and colleagues[29] suggests that one-half of a standard deviation of a measure represents a good approximation of meaningful change.
[2]Standard error of measurement (SEm) = SD*sqrt(1 – ICC), where ICC is based on awake objective cough frequency defined test–retest reliability.
CSD, Cough Severity Diary; ICC, intraclass correlation coefficient.

This investigation into the psychometric properties of the CSD shows that CSD total and domain scores demonstrated good internal consistency reliability, test–retest reliability over 28 days, and construct and known-groups validity. Strong-to-moderate correlations were seen between the CSD total and subscale scores and the LCQ, objective cough frequency, and cough severity VAS score. The CSD was also able to distinguish between groups of patients who differed across categories of cough-specific HRQL as determined by LCQ scores, and categories of awake objective cough frequency scores. Responsiveness testing demonstrated that the CSD was highly responsive to changes in cough-specific HRQL and cough frequency over time. Large effect sizes were seen for CSD change scores associated with changes in patient-reported cough severity (PGIC groups) and changes in objective cough frequency scores.

The psychometric analyses also provide support for the relevance of the three-domain structure in patients with RCC/UCC. Specifically, the overall fit of the three factor CSD (including frequency, intensity, and disruption factors) was found to be

Table 10. Receiver operating characteristic curve analysis for CSD score thresholds predictive of patient global impression of change score of 1, 2, and 3.

CSD total/subscale, change score threshold	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Youden's index
Total Score					
< -1	0.74	0.84	0.92	0.57	0.58
< -1.1	0.72	0.88	0.94	0.57	0.60
< -1.2	0.70	0.91	0.95	0.56	0.61
< -1.3	0.66	0.93	0.96	0.54	0.59
< -2	0.45	0.97	0.97	0.43	0.42
< -3	0.21	1.00	1.00	0.35	0.21
< -4	0.12	1.00	1.00	0.32	0.12
< -5	0.04	1.00	1.00	0.31	0.04
< -6	0.01	1.00	1.00	0.30	0.01
Frequency					
$\leqslant -1$	0.78	0.75	0.88	0.59	0.53
$\leqslant -1.1$	0.74	0.81	0.90	0.57	0.55
$\leqslant -1.2$	0.73	0.84	0.92	0.57	0.57
$\leqslant -1.3$	0.71	0.87	0.93	0.56	0.58
< -2	0.51	0.99	0.99	0.46	0.49
< -3	0.27	1.00	1.00	0.37	0.27
< -4	0.13	1.00	1.00	0.33	0.13
< -5	0.06	1.00	1.00	0.31	0.06
< -6	0.01	1.00	1.00	0.30	0.01
< -7	0.01	1.00	1.00	0.30	0.01
Intensity					
< -1	0.77	0.80	0.90	0.59	0.56
< -1.1	0.74	0.84	0.92	0.57	0.58
< -1.2	0.70	0.87	0.93	0.55	0.57
< -1.3	0.64	0.91	0.95	0.52	0.56
< -2	0.53	0.96	0.97	0.46	0.48
< -3	0.28	1.00	1.00	0.37	0.28

(Continued)

Table 10. (Continued)

CSD total/subscale, change score threshold	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Youden's index
< -4	0.15	1.00	1.00	0.33	0.15
< -5	0.09	1.00	1.00	0.32	0.09
< -6	0.03	1.00	1.00	0.30	0.03
< -7	0.01	1.00	1.00	0.30	0.01
Disruption					
$\leqslant -1$	0.59	0.88	0.92	0.48	0.47
$\leqslant -1.1$	0.57	0.88	0.92	0.47	0.45
$\leqslant -1.2$	0.54	0.90	0.93	0.45	0.44
$\leqslant -1.3$	0.52	0.91	0.93	0.45	0.43
< -2	0.32	0.96	0.95	0.37	0.28
< -3	0.13	1.00	1.00	0.33	0.13
< -4	0.08	1.00	1.00	0.32	0.08
$\leqslant -5$	0.04	1.00	1.00	0.31	0.04
< -6	0.02	1.00	1.00	0.30	0.02
< -7	0.01	1.00	1.00	0.30	0.01

CSD, Cough Severity Diary.

acceptable according to the CFI and SRMR values. The RMSEA value was higher than would typically be considered acceptable; however, this is to be expected in this case as models were very simplistic. In simplistic models, RMSEA may be inflated and not accurately reflect model fit.[35]

As with all research there are a few limitations that should be mentioned. The participants included in this study were primarily white women; confirmation of these results within a broader patient population would be useful. In addition, the CSD was developed because there were no measures available to adequately assess cough severity that includes frequency, intensity, and disruptions; therefore, a 'gold standard' against which to validate the CSD was not available. Regardless, analyses using the LCQ and objective cough counts showed acceptable correlations that support the validity of the CSD. The calculation of a mean weekly CSD score in this study required a minimum of 1 day of complete data which may not completely reflect a week's worth of data; current guidance for the future use of the CSD requires a minimum of 4 days of data to calculate a mean weekly score, which would be expected to further improve the psychometric properties of the measure. Finally, the results of analyses to define a clinically meaningful change on the CSD using the PGIC in this study can be considered preliminary and should be confirmed in future trials.

Importantly, this study provided the opportunity to conduct analyses to help guide the interpretation and estimation of clinically meaningful CSD results, using both anchor- and distribution-based methods. Based on results from these methods, a change of $\geqslant 1.3$ on the CSD total and subscale scores likely represents a meaningful improvement in cough severity, corresponding with a large effect size, and as such can be used to identify responders

in future clinical trials. However, further work in the context of multinational clinical trials is suggested to confirm this proposed threshold.

Conclusion

The CSD complements other measures of cough impact and objective cough frequency by providing information on patients' perceptions of meaningful improvements in cough symptom severity. The findings from the current study demonstrate that the CSD is a reliable, valid, and responsive measure of cough symptom severity in patients with RCC/UCC and fit-for-purpose for assessing changes in cough severity in clinical trials.

Acknowledgments

The authors would like to thank all participants and investigators (see the appendix) who participated in this study. In particular, the authors would like to thank Jonathan Schelfhout (Merck & Co. Inc., USA) who contributed to the CSD study.

In addition, we would like to thank Boyd Scott (Merck & Co. Inc., USA) and Desiree Deoria (Merck & Co. Inc., USA) for writing and logistical assistance in the submission of this manuscript.

Authors' Contributions

All authors confirm they were involved in the design, data collection, and/or interpretation of these study results and significantly contributed to the development of this manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA.

Conflict of interest statement

AMN is an employee of Merck & Co. Inc., USA and may own stock in the company.

EB is employed by Evidera, which provides consulting and other research services to pharmaceutical, medical device, and related organizations. In her salaried positions, she works with a variety of companies and organizations, and is precluded from receiving payment or honoraria directly from these organizations for services rendered. Evidera received funding from Merck to participate in the study and the development of this manuscript.

PD serves as a consultant to Merck.

MV reports non-financial support from Merck, personal fees from Evidera, during the conduct of the study.

ORCID iD

Allison Martin Nguyen https://orcid.org/0000-0002-6385-5570

Data sharing statement

The data-sharing policy of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co. Inc., Kenilworth, NJ, USA, including restrictions, is available at http://engagezone.msd.com/ds_documentation.php. Requests for access to the data can be submitted through the Engagezone site or *via* email to dataaccess@merck.com

Supplemental material

The reviews of this paper are available via the supplemental material section.

References

1. Irwin RS and Madison JM. The diagnosis and treatment of cough. *N Engl J Med* 2000; 343: 1715–1721.
2. Irwin RS and Madison JM. Symptom research on chronic cough: a historical perspective. *Ann Intern Med* 2001; 134: 809–814.
3. Nelson C and Woodwell D. National ambulatory medical care survey: 1993 summary. *Vital Health Stat 13* 1998; 136: 1–99.
4. Song WJ, Chang YS, Faruqi S, *et al*. The global epidemiology of chronic cough in adults: a systematic review and meta-analysis. *Eur Respir J* 2015; 45: 1479–1481.
5. Sundar KM and Daly SE. Chronic cough and OSA: a new association? *J Clin Sleep Med* 2011; 7: 669–677.
6. Chamberlain SA, Garrod R, Douiri A, *et al*. The impact of chronic cough: a cross-sectional European survey. *Lung* 2015; 193: 401–408.
7. Young EC and Smith JA. Quality of life in patients with chronic cough. *Ther Adv Respir Dis* 2010; 4: 49–55.
8. Kuzniar TJ, Morgenthaler TI, Afessa B, *et al*. Chronic cough from the patient's perspective. Mayo Clin Proc 2007; 82: 56–60.

9. Dicpinigaitis PV, Tso R and Banauch G. Prevalence of depressive symptoms among patients with chronic cough. *Chest* 2006; 130: 1839–1843.
10. Spinou A and Birring SS. An update on measurement and monitoring of cough: what are the important study endpoints? *J Thorac Dis* 2014; 6: S728–S734.
11. Boulet LP, Coeytaux RR, McCrory DC, *et al*. Tools for assessing outcomes in studies of chronic cough: CHEST guideline and expert panel report. *Chest* 2015; 147: 804–814.
12. Vernon M, Leidy NK, Nacson A, *et al*. Measuring cough severity: perspectives from the literature and from patients with chronic cough. *Cough* 2009; 5: 5.
13. Vernon M, Kline Leidy N, Nacson A, *et al*. Measuring cough severity: development and pilot testing of a new seven-item cough severity patient-reported outcome measure. *Ther Adv Respir Dis* 2010; 4: 199–208.
14. Aaronson N, Alonso J, Burnam A, *et al*. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; 11: 193–205.
15. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 2009; 74: 65132–65133.
16. Guyatt G, Gutterman D, Baumann MH, *et al*. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American college of chest physicians task force. *Chest* 2006; 129: 174–181.
17. Birring SS, Prudon B, Carr AJ, *et al*. Development of a symptom specific health status measure for patients with chronic cough: Leicester cough questionnaire (LCQ). *Thorax* 2003; 58: 339–343.
18. Morice AH, Fontana GA, Belvisi MG, *et al*. ERS guidelines on the assessment of cough. *Eur Respir J* 2007; 29: 1256–1276.
19. Guy W. *ECDEU Assessment manual for psychopharmacology*. Rockville, MD.: United States Department of Health, Education and Welfare, 1976, pp. 76–338.
20. Kline RB. *Principles and practice of structural equation modeling*. 2nd ed. New York: Guilford Press, 2005.
21. Browne MW and Cudeck R. Alternative ways of assessing model fit. *Sociol Method Res* 1992; 21: 230–258.
22. Hays RD and Revick DA. *Reliability and validity (including responsiveness)*. In: Fayers P and Hays R (eds) *Assessing quality of life in clinical trials*. 2nd ed. New York: Oxford University Press Inc., 2005.
23. Kazis LE, Anderson JJ and Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: S178–S189.
24. Cohen J. Statistical power analysis for the behavioral-sciences. *Percept Motor Skill* 1988; 67: 1007–1007.
25. Wyrwich KW, Nienaber NA, Tierney WM, *et al*. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; 37: 469–478.
26. Wyrwich KW, Tierney WM and Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861–873.
27. Shikiar R, Harding G, Leahy M, *et al*. Minimal important difference (MID) of the dermatology life quality index (DLQI): results from patients with chronic idiopathic urticaria. *Health Qual Life Outcomes* 2005; 3: 36.
28. Norman GR, Sloan JA and Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41: 582–592.
29. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32–35.
30. Schmit, Kristine M, Goode AP, *et al*. Evaluating cough assessment tools: a systematic review. *Chest* 144; 2013: 1819–1826.
31. Birring SS and Spinou A. How best to measure cough clinically. *Curr Opin Pharmacol* 2015; 22: 37–40.
32. French CT, Irwin RS, Fletcher KE, *et al*. Evaluation of a cough-specific quality-of-life questionnaire. *Chest* 2002; 121: 1123–1131.
33. Baiardini I, Braido F, Fassio O, *et al*. A new tool to assess and monitor the burden of chronic cough on quality of life: Chronic Cough Impact Questionnaire. *Allergy* 2005; 60: 482–488.
34. Hulme K, Deary V, Dogan S, *et al*. Psychological profile of individuals presenting with chronic cough. *ERJ Open Res* 2017; 3: pii: 00099-2016.
35. Bievik E and Olsson UH. Adding variables to improve fit: the effect of model size on fit assessment in LISREL. In: Cudeck R, Du Toit S and Sorbom D (eds) *Structural equation modeling: present and future*. Lincolnwood, IL: Scientific Software International, 2001, pp. 169–194.