

Validation of a visual analog scale for assessing cough severity in patients with chronic cough

Allison Martin Nguyen , Elizabeth D. Bacci, Margaret Vernon, Surinder S. Birring, Carmen La Rosa, David Muccino and Jonathan Schelfhout

Abstract

Introduction: Patients with chronic cough experience considerable burden. The cough severity visual analog scale (VAS) records patients' assessment of cough severity on a 100-mm linear scale ranging from "no cough" (0 mm) to "worst cough" (100 mm). Although cough severity scales are widely used in clinical practice and research, their use in patients with refractory or unexplained chronic cough has not been formally validated.

Methods: This analysis includes data from a phase 2b randomized controlled trial of the P2X3-receptor antagonist gefapixant for treatment of refractory or unexplained chronic cough (NCT02612610). Cough severity VAS scores were assessed at baseline and Weeks 4, 8, and 12. The cough severity VAS was validated using several outcomes, including the Cough Severity Diary (CSD), Leicester Cough Questionnaire (LCQ), patient global impression of change (PGIC) scale, and objective cough frequency. Validation metrics included test-retest reliability, convergent and known-groups validity, responsiveness, and score interpretation (i.e., clinically meaningful change threshold).

Results: The analysis included 253 patients (median age, 61.0 years; females, 76%). Test-retest reliability of the cough severity VAS was moderate (intraclass correlation coefficient, 0.51). The cough severity VAS had acceptable convergent validity with other related measures (Pearson r of 0.53 and -0.41 for CSD and LCQ total scores, respectively; $p < 0.0001$ for each). Known-groups validity was supported by significant differences in mean cough severity VAS scores across severity groups defined using CSD, LCQ, and cough frequency tertiles. A large effect size was observed in patients with the greatest improvements in PGIC (Cohen $d = -1.8$). A ≥ 30 -mm reduction in the cough severity VAS was estimated as a clinically meaningful change threshold for clinical trials in chronic cough.

Conclusions: The cough severity VAS is a valid and responsive measure. A cough severity VAS reduction of ≥ 30 mm can discriminate clinically meaningful changes in chronic cough severity in clinical studies.

Keywords: clinically meaningful change, cough monitoring, Cough Severity Diary, idiopathic cough, Leicester Cough Questionnaire, minimal important difference, objective cough frequency, patient-reported outcomes, responder threshold

Received: 17 March 2021; revised manuscript accepted: 8 September 2021.

Introduction

Chronic cough, defined in adults as a cough lasting more than 8 weeks, is a burdensome medical condition with an estimated prevalence ranging from approximately 4% to 11%.¹⁻⁵ Patients with

chronic cough experience considerable burden that can be long lasting, as chronic cough often persists for many years, during which patients may undergo numerous doctor visits to assess and manage their cough.⁶⁻¹¹ Although chronic cough

Ther Adv Respir Dis

2021, Vol. 15: 1-13

DOI: 10.1177/
17534666211049743

© The Author(s), 2021.

Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:
Jonathan Schelfhout
Merck & Co., Inc., 2000
Galloping Hill Road,
Kenilworth, NJ 07033,
USA.

jonathan.schelfhout@merck.com

Allison Martin Nguyen
Carmen La Rosa
David Muccino
Merck & Co., Inc.,
Kenilworth, NJ, USA

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

Surinder S. Birring
Centre for Human &
Applied Physiological
Sciences, School of Basic
& Medical Biosciences,
Faculty of Life Sciences &
Medicine, King's College
London, London, UK

typically is associated with an underlying condition, some patients with chronic cough have a cough that persists despite optimal treatment of underlying conditions according to practice guidelines (refractory chronic cough (RCC)) or have no identified diagnosable cause for cough despite extensive assessment (unexplained chronic cough (UCC)).^{1,2,12} There are currently limited treatment options for patients with RCC or UCC, as there are no treatments with approved indications; however, the development of novel antitussives for the treatment of RCC or UCC is an active area of clinical research.^{13–15}

Previous research suggests that patients with chronic cough experience aspects of burden that can be categorized into distinct but related facets, including cough frequency, cough intensity, disruption of daily activities, and the impact of cough on quality of life.^{6,8,16} Several complementary clinical outcome assessments have been used to measure these aspects of burden and to evaluate treatment benefit in patients with chronic cough who have been treated with pharmacologic or nonpharmacologic approaches.^{17,18} Objective cough frequency is often used to assess the efficacy of antitussives in clinical trials, but use of cough-monitoring devices in a clinical practice is impractical.¹⁹ Previously validated cough-related patient-reported outcomes (PROs) include the Leicester Cough Questionnaire (LCQ),²⁰ Cough-Specific Quality-of-Life Questionnaire,²¹ and Cough Severity Diary (CSD),^{22,23} all of which can also be used to monitor patients' responses to treatment. These tools evaluate the effect of cough on quality of life and cough symptom burden.

The cough severity visual analog scale (VAS) is a brief, easily administered PRO that is widely used in clinical practice and clinical research to assess cough severity in both acute and chronic cough.^{13–15,24,25} The cough severity VAS is often 100 mm in length and captures patients' self-assessment of cough severity ranging from 0 mm (no cough) to 100 mm (worst cough). Despite the cough severity VAS being one of the most widely used tools to assess cough in clinical trials and clinical practice, it has not been formally validated, and its clinically meaningful change threshold is unknown. Formal validation of this instrument is important to ensure that changes in patients' cough severity in response to therapeutic interventions can be reliably assessed in both clinical practice and clinical research settings.

The aim of this analysis was to assess and report the psychometric properties of the cough severity VAS, including test–retest reliability, convergent validity, known-groups validity, and responsiveness in patients with RCC or UCC. Furthermore, analyses to estimate meaningful changes in cough severity VAS scores were conducted to guide score interpretation.

Methods

Data source

This study is a post hoc analysis of the psychometric properties of the cough severity VAS using data collected in a recent phase 2b study of the P2X3-receptor antagonist gefapixant (Clinical Trials.gov identifier: NCT02612610), for which primary results have been previously reported.¹³ This clinical trial assessed the efficacy and safety of 3 doses of gefapixant or matching placebo in adult patients with RCC/UCC. Key eligibility criteria at screening included RCC or UCC (per American College of Chest Physicians and British Thoracic Society guidelines) for ≥ 1 year, a cough severity VAS score of ≥ 40 mm, and no substantial abnormalities contributing to cough within the past 5 years as determined by a chest x-ray; additional study eligibility criteria are shown in Table 1. Blinded data from all patients receiving any dose of gefapixant or placebo were pooled into a single population for validation of the cough severity VAS.

Outcome measures

Outcomes collected in the phase 2b study that were used to evaluate the cough severity VAS in this analysis included the CSD, the LCQ, the patient global impression of change (PGIC) scale, and objective awake cough frequency. The cough severity VAS used in this study was measured on a 100-mm scale ranging from “no cough” (0 mm) to “worst cough” (100 mm), with a recall period of today. The CSD includes 7 items capturing the patients' perceptions of cough in terms of frequency, intensity, and disruptiveness on an 11-point scale ranging from 0 to 10, with a recall period of today. The CSD total score is calculated by averaging scores from all 7 items, with higher scores reflecting greater severity. Total CSD scores for this study were calculated as the average total daily CSD score over the previous week. The LCQ is a 19-item cough-specific

Table 1. Key inclusion and exclusion criteria for participants enrolled in NCT02612610.

Key inclusion criteria	Key exclusion criteria
<ul style="list-style-type: none"> Adults aged < 80 years Chest imaging (chest radiograph or CT thorax) within the last 5 years confirming no abnormalities contributing to cough Diagnosis of RCC or UCC for ≥ 1 year according to ACCP/BTS guidelines Score of ≥ 40 mm on the cough severity VAS at screening Women with child-bearing potential and male participants and their partners of child-bearing potential must have used 2 acceptable forms of birth control from screening through the follow-up visit 	<ul style="list-style-type: none"> Current smoker or history of smoking within 6 months before study entry Initiation of treatment with an ACE inhibitor within 4 weeks of baseline visit History of opioid use within 1 week of baseline visit Requirement for concomitant therapy using prohibited medications during study period (including pregabalin, gabapentin, thalidomide, dextromethorphan, guaifenesin, benzonatate, opioids, and ACE inhibitors) FEV₁/FVC of < 60% History of respiratory tract infection or recent pulmonary status change within 4 weeks of baseline visit History of cystic fibrosis/bronchiectasis BMI of < 18 kg/m² or ≥ 40 kg/m² History of renal diseases, kidney/bladder stones, conditions/disorders predisposed to nephrolithiasis or conditions affecting drug absorption
<p>ACCP, American College of Chest Physicians; ACE, angiotensin-converting enzyme; BMI, body mass index; BTS, British Thoracic Society; CT, computed tomography; FEV₁/FVC, percent of forced vital capacity exhaled in first second; RCC, refractory chronic cough; UCC, unexplained chronic cough; VAS, visual analog scale.</p>	

health-related quality-of-life questionnaire comprising 3 domains that assess the impact of cough on physical, psychological, and social functioning, with a recall period of the past 2 weeks. Each item is rated using a 7-point Likert-type scale; a higher total score, calculated by summing the domain scores, indicates a better quality of life. The PGIC scale measures change in a patient's overall status on a 7-point scale, ranging from "very much improved" (PGIC score of 1) to "very much worse" (PGIC score of 7). Objective cough frequency was monitored via ambulatory cough recorders (VitaloJAK; Vitalograph, Buckingham, UK), which were worn by patients for 24-h periods; awake cough frequency was derived from the corresponding 24-h sound recordings during times in which the patient was awake. Outcomes in the phase 2b study were assessed at baseline and Weeks 4, 8, and 12; PGIC was assessed at Weeks 4, 8, and 12. However, baseline and Weeks 4 and 12 were the primary time points used for the current analysis.

Validation metrics and statistical analyses

All analyses were conducted following an a priori validation analysis plan. All statistical tests used a significance level of $p < 0.05$.

Test-retest reliability. Cough severity VAS scores were assessed for consistency over time in a subset of patients characterized as "stable" from baseline to Week 4 using 2 anchors: (1) patients with

a $\leq 10\%$ change in awake cough frequency and (2) patients reporting "no change" on the PGIC scale. Because of the long recall period for patients to report their change using the PGIC scale, this analysis was considered exploratory. Intraclass correlation coefficient (ICC) and change scores using paired t tests, as well as 95% confidence intervals, were calculated between test scores at baseline and retest scores at Week 4 in the stable subpopulations. ICC values between 0.5 and 0.75 have been suggested to indicate moderate reliability, whereas values < 0.5 may reflect poor reliability.²⁶

Convergent validity. The cough severity VAS was assessed to determine the extent to which the scores were consistent with other measures assessing a similar construct (cough severity) and closely related constructs (cough-specific quality of life). The cough severity VAS was therefore correlated with related cough metrics (i.e., the CSD and LCQ) at baseline using Pearson r . Correlation coefficients with absolute magnitudes from 0.30 to < 0.50 and ≥ 0.50 to 0.70 were interpreted to reflect low and moderate correlations, respectively.²⁷

Known-groups validity. Known-groups validity was evaluated to capture the ability of the cough severity VAS to discriminate between different levels of cough severity. Patients were stratified into tertiles using the sample distribution for the CSD total score, LCQ total score, and awake cough frequency at baseline and Week 4. Cough severity VAS scores were compared between these

tertiles using analysis of variance, with post hoc comparisons assessed via the Scheffé test to determine whether cough severity VAS scores significantly discriminate between tertiles.

Responsiveness. To evaluate the responsiveness of the cough severity VAS in detecting meaningful changes in other validated instruments, analysis of covariance was used to compare change in mean scores, controlling for baseline scores, by response on the PGIC scale from baseline to Week 4 and by change in awake cough frequency from baseline to Week 4. For the analysis of cough frequency, patients were considered responders by using 4 definitions of response: $\geq 30\%$, $\geq 50\%$, and $\geq 70\%$ reductions in awake cough frequency and a reduction of ≥ 0.30 standard deviations (SD) in awake cough frequency. The $\geq 30\%$ -reduction cutoff for response was previously defined as a minimal clinically important difference in objective cough counts and has been used in clinical trials of RCC and UCC (ClinicalTrials.gov identifiers: NCT03449134, NCT03449147, NCT04562155).^{28,29} The effect size of mean cough severity VAS scores for the groups defined above was calculated using a standardized mean difference by subtracting the mean score at baseline from the mean score at Week 4 and dividing by the baseline SD. Previous guidance suggests that effect sizes of 0.20, 0.50, and 0.80 be interpreted as small, moderate, and large, respectively.³⁰

Score interpretation. Anchor- and distribution-based approaches were used to derive clinically meaningful thresholds for change in the cough severity VAS. Anchor-based approaches were assessed using changes from baseline to both Week 4 and Week 12. The PGIC scale was grouped into 5 categories as follows: 1 or 2 (very much improved, much improved), 3 (minimally improved), 4 (no change), 5 (minimally worse), and 6 or 7 (much worse, very much worse). The meaningful change threshold for the cough severity VAS was estimated as the mean change in cough severity VAS scores corresponding to patients reporting themselves as at least “minimally improved” (PGIC score of 3); the threshold was also estimated in patients who reported themselves as “very much improved” or “much improved” (PGIC score of 1 or 2).

Receiver operating characteristic (ROC) curves were also evaluated to determine the threshold

values for change in mean cough severity VAS score from baseline to Week 4 that would be most predictive of patients rating themselves as “minimally improved” (PGIC score of 3) or “very much improved/much improved” (PGIC score of 1 or 2). Two ROC curves were therefore analyzed: one for predicting patients reporting 1 or 2 versus 3 to 7 on the PGIC scale and one for predicting patients reporting 1 to 3 versus 4 to 7 on the PGIC scale. The threshold for each curve was defined as the point on the ROC curve closest to 100% sensitivity and specificity (i.e., the point with the shortest difference from the upper left quadrant (0,1)). The Youden index was used to determine the cough severity VAS change score that optimized sensitivity and specificity for predicting global improvements on the PGIC scale.

Distribution-based methods were also used to determine the minimum change needed to exceed the inherent noise of the cough severity VAS. Distribution-based estimates included calculation of one-half of the SD of the cough severity VAS scores at baseline and the standard error (SE) of measurement (SEM), calculated by multiplying the baseline SD by the square root of $(1 - ICC)$.

By triangulating findings from the responsiveness, anchor-based, and distribution-based analyses, a single estimate of the clinically meaningful change threshold for the cough severity VAS was proposed for use, with a specific focus on a threshold to be used for clinical trial research. After estimating this clinically meaningful change threshold, an additional analysis was performed to confirm the converse definition of this threshold (i.e., by assessing PGIC scores among patients who did vs. did not have a cough severity VAS change exceeding this estimated threshold).

Results

Patient population

Baseline characteristics of this patient population ($N=253$) have been previously reported.^{13,23} Patients enrolled in the study were predominantly female (76%), had a median (range) age of 61 (22–79) years, and had a mean (SD) cough duration of 14.5 (11.7) years. The mean (SD) cough severity VAS score at baseline across treatment groups was 57.5 (22.3) mm, with a median (range) of 60 (7–100) mm.

Table 2. Test–retest reliability of VAS scores from baseline to Week 4.

Parameter	N	Baseline cough severity VAS, mean (SD), mm	Week 4 cough severity VAS, mean (SD), mm	Mean change score difference, mm	ICC
Patients with $\leq 10\%$ change in awake cough frequency from baseline to Week 4	32	56.2 (25.8)	46.2 (23.8)	-10.0	0.45
Patients reporting “no change” on PGIC from baseline to Week 4	61	62.0 (20.6)	56.5 (21.0)	-5.5	0.51

ICC, intraclass correlation coefficient; PGIC, patient global impression of change; SD, standard deviation; VAS, visual analog scale.

Test–retest reliability

Patients who were defined as “stable” on the basis of a change in awake cough frequency of $\leq 10\%$ or a PGIC score of 4 (i.e., no change) had some change in mean cough severity VAS scores from baseline to Week 4 (ICCs of 0.45 and 0.51, respectively), suggesting moderate test–retest reliability (Table 2).

Convergent validity

Cough severity VAS scores were significantly correlated with other cough-related measures at baseline including mean weekly CSD and its subscales and the LCQ and its subscales (all $p < 0.0001$; Table 3). Generally, moderate correlations were seen between the cough severity VAS and mean weekly CSD total score ($r = 0.53$) and subscale scores. Low to moderate correlations were also seen with the LCQ total score ($r = -0.41$) and physical, psychological, and social subscale scores.

Known-groups validity

Patients were stratified into tertiles by weekly CSD total scores, LCQ total scores, and awake cough frequency at baseline and Week 4 (Table 4). Mean cough severity VAS scores were significantly different between different known groups for each comparator outcome at both time points, supporting the known-groups validity of the cough severity VAS. Cough severity VAS scores decreased with improvement in mean weekly CSD total scores from baseline to Week 4 ($p < 0.0001$). Similarly, decreased cough severity VAS scores were associated with improvements in cough-related quality of life (LCQ total scores) and awake cough frequency (Table 4).

Responsiveness

Responsiveness of cough severity VAS scores between baseline and Week 4 was assessed using both PGIC and awake cough frequency as criterion variables. The least-squares mean (LSM) changes in cough severity VAS scores were compared between different PGIC categories (Figure 1(a)). Least-squares mean cough severity VAS scores were significantly different between PGIC categories ($F = 64.64$; $p < 0.0001$). Notably, the number of patients with worsening status (PGIC score of ≥ 5) was very small ($n = 9$). Patients with the greatest improvements in PGIC (i.e., a PGIC score of 1 or 2) also had the greatest change in mean cough severity VAS scores from baseline to Week 4 (-37.3 mm), which corresponded with a large effect size ($d = -1.8$; Table 5). Effect sizes (Cohen d) for PGIC scores of 3, 4, 5, and 6 or 7 were -0.8 , -0.3 , 0.4 , and 0.4 , respectively (Table 5).

Similarly, LSM changes in cough severity VAS scores were found to be significantly greater in responders *versus* nonresponders using an awake cough frequency responder threshold of $\geq 30\%$ reduction from baseline ($F = 68.34$; $p < .0001$; Figure 1(b)). Significant differences in LSM (SE) cough severity VAS changes were also observed between awake cough frequency responders *versus* nonresponders using a $\geq 50\%$ -reduction threshold (-36 (2.3) vs -13 (1.7) mm; $F = 77.33$; $p < 0.0001$), a $\geq 70\%$ -reduction threshold (-39 (2.9) vs -16 (1.6) mm; $F = 67.68$; $p < 0.0001$), and a reduction of ≥ 0.30 SD (-29 (2.0) vs -12 (2.1) mm; $F = 58.52$; $p < 0.0001$). Mean change scores and effect sizes were consistently greater for those who were considered responders *versus* nonresponders by awake cough frequency responder thresholds, with effect sizes for responders ranging from -1.3 to -1.8 (Table 5). Patients with a $\geq 30\%$ reduction in awake cough

Table 3. Convergent validity of cough severity VAS with other cough-related measures at baseline.

Cough-related measure	Pearson <i>r</i> with cough severity VAS score	<i>p</i> value
Cough severity diary		
Total score	0.53	<0.0001
Frequency	0.57	<0.0001
Intensity	0.50	<0.0001
Disruption	0.42	<0.0001
Leicester Cough Questionnaire		
Total score	-0.41	<0.0001
Physical	-0.34	<0.0001
Psychological	-0.38	<0.0001
Social	-0.34	<0.0001
VAS, visual analog scale.		

frequency showed a mean cough severity VAS score change of -29.0 ($d = -1.3$), compared with -10.9 ($d = -0.5$) for those with a <30% reduction in awake cough frequency (Table 5).

Score interpretation

The clinically important change threshold was defined using the PGIC scale response categories of “very much improved” or “much improved” (PGIC score of 1 or 2) and “minimally improved” (PGIC score of 3). Among patients with a PGIC score of 1 or 2, mean (SD) changes in cough severity VAS scores at Weeks 4 and 12 were -37.3 (24.8) and -37.6 (24.2) mm, respectively. Patients with a PGIC score of 3 at Weeks 4 and 12 had mean (SD) changes in cough severity VAS scores of -18.9 (22.3) and -15.7 (24.9) mm, respectively.

An ROC curve analysis was then conducted using two definitions of a response on the PGIC: a PGIC score of 1, 2, or 3 (Supplemental Figure A) or a PGIC score of 1 or 2 (Supplemental Figure B). Changes in the cough severity VAS score of 20 or 30 mm (for PGIC score of 1–3 or 1–2, respectively) were found to maximize sensitivity and specificity for predicting improvements on the PGIC scale, as evidenced by the highest

Youden index values being observed at these two thresholds (Table 6).

Results of the distribution-based analyses indicated that the minimum changes in cough severity VAS needed to exceed the inherent baseline noise of the measure were 11.2 and 16.6 mm based on the one-half SD and SEM estimates, respectively.

After triangulating the results from the responsiveness analyses and score-interpretation methods, a clinically meaningful change threshold of ≥ 30 mm was proposed for use in the context of clinical research. This degree of change was observed among patients rating themselves as “very much improved” or “much improved” on the PGIC scale and among patients categorized as responders based on a $\geq 30\%$ change in awake cough frequency; it was also consistent with the threshold identified via the ROC curve analysis. When patients were divided into responder *versus* nonresponder groups based on this threshold and the PGIC categories reported by these 2 subgroups were assessed (Table 7), a greater proportion of patients who were VAS responders (i.e., achieved a ≥ 30 -mm reduction from baseline on the VAS) had a PGIC score of 1, 2, or 3 compared with nonresponders.

Discussion

Previous studies have suggested that the cough severity VAS may be responsive, reproducible, and associated with objective cough measures.^{31,32} This analysis confirms the psychometric characteristics of the cough severity VAS in a population of adults with RCC/UCC, wherein the measure was found to have acceptable test–retest reliability, strong convergent validity, known-groups validity, and responsiveness. A ≥ 30 -mm reduction was proposed as the clinically meaningful change threshold for cough severity VAS after triangulation of multiple anchor-based approaches. A high proportion of patients considered as cough severity VAS responders by this definition were found to report themselves as improved on the PGIC scale, supporting the meaningfulness of this threshold.

The test–retest reliability for cough severity VAS in this analysis (ICCs of 0.45–0.51) was found to be moderate compared with that of previous studies of other cough-related measures (ICCs of

Table 4. Known-groups validity of the cough severity VAS score at baseline and Week 4.

Cough severity, VAS score, mm	Cough Severity Diary total score tertiles						F value	p value
	0-33 percentile		34-66 percentile		67-100 percentile			
	N	Mean (SE)	n	Mean (SE)	n	Mean (SE)		
Baseline	76	46.0 (2.3)	85	57.8 (2.2)	78	68.7 (2.3)	23.8	<0.0001
Week 4	76	19.5 (2.2)	75	31.9 (2.2)	68	59.8 (2.3)	85.4	<0.0001
Leicester Cough Questionnaire total score tertiles								
Score 3-8		Score 9-13		Score 14-21		F value		p value
N	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	
Baseline	29	73.6 (3.9)	150	58.7 (1.7)	74	48.9 (2.5)	14.7	<0.0001
Week 4	8	80.6 (7.3)	72	51.7 (2.4)	156	27.2 (1.7)	53.1	<0.0001
Awake cough frequency tertiles								
0-33 percentile		34-66 percentile		67-100 percentile		F value		P value
n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	
Baseline	83	50.1 (2.4)	84	58.9 (2.4)	84	63.8 (2.4)	8.5	0.0003
Week 4	76	24.0 (2.6)	77	34.9 (2.6)	76	51.0 (2.6)	27.7	<0.0001
SE, standard error; VAS, visual analog scale.								

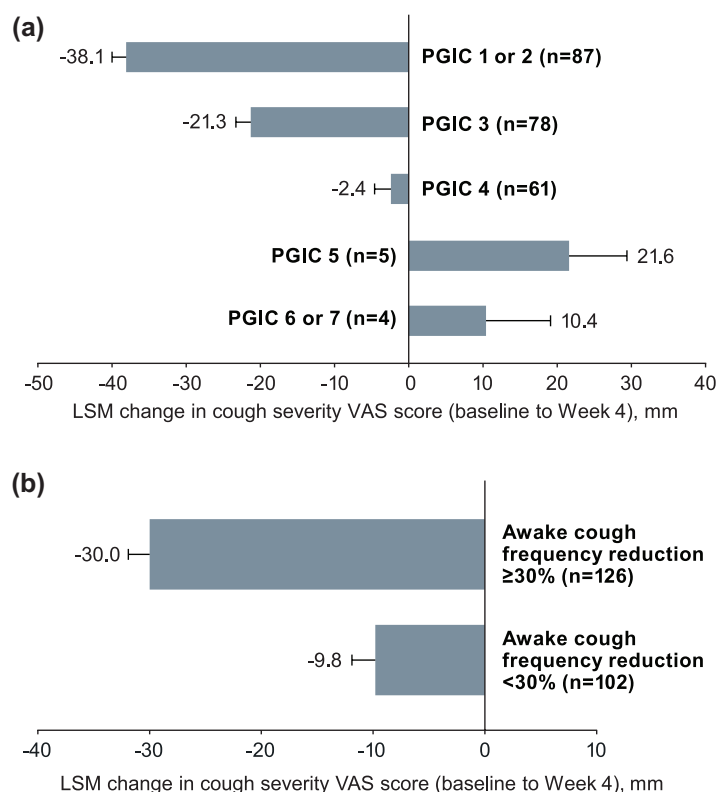


Figure 1. Responsiveness of cough severity VAS change scores by (a) PGIC category at Week 4 and (b) $\geq 30\%$ vs. $< 30\%$ reduction in awake cough frequency at Week 4. Error bars are standard error. LSM, least-squares mean; PGIC, patient global impression of change; VAS, visual analog scale.

0.68–0.96).^{20,22,33,34} This finding was anticipated on the basis of the long duration between test periods (i.e., 4 weeks, compared with 1 to 2 weeks in previous studies), as well as the lack of a gold standard for defining a “stable” population. Although defining stability using an objective measure of awake cough frequency was considered a reasonable approach in this study, strong reliability estimates were not expected because of differences in the concepts measured with these outcomes (i.e., cough frequency vs. perceived cough severity) and previous reports of a lack of a strong correlation between cough frequency and cough-related PROs.³⁵ Given these expectations, the analysis of cough severity VAS changes among patients reporting “no change” on the PGIC scale may have provided a more reliable estimate of the test–retest reliability in this study; however, the ICCs for test–retest reliability were comparable when defining stability by cough frequency and “no change” on the PGIC. Similarly, convergent validity measurements between the VAS and the LCQ and CSD showed stronger correlations in previous studies (Spearman rank correlation or

Pearson r of -0.72 and 0.84 , respectively)^{20,22} compared with that of the current study (Pearson r of -0.41 and 0.53 , respectively). Differences in the patient populations examined may contribute to the differences in test–retest reliability and convergent validity measurements between the current study and previous studies. For example, the inclusion criteria for the population in the current study required a diagnosis of RCC or UCC for ≥ 1 year and baseline cough severity VAS score of ≥ 40 mm. The mean cough duration in this study was much longer than that in Birring *et al.*²⁰ (referenced above), which validated the LCQ in patients with chronic cough (mean cough duration: 14.5 vs. 5.0 years, respectively). Additionally, in previous work assessing the CSD (Vernon *et al.*; referenced above), only half of the patients with chronic cough had a cough duration greater than 1 year.²² This study focused on validation of the cough severity VAS PRO in a patient population with relatively severe cough of long duration. Moreover, unlike other patient populations with chronic cough, it is possible that patients with RCC or UCC with a lengthy history

Table 5. Responsiveness of cough severity VAS by PGIC and awake cough frequency.

Category	N	Baseline cough severity VAS, mean (SD), mm	Week 4 cough severity VAS, mean (SD), mm	Mean change score difference (range), mm ^a	Effect size ^b
PGIC score					
1 or 2	87	56.5 (20.9)	19.2 (15.8)	-37.3 (-84.0, 26.0)	-1.8
3	78	54.1 (23.3)	35.1 (18.9)	-18.9 (-65.0, 35.0)	-0.8
4	61	62.0 (20.6)	56.5 (21.0)	-5.5 (-55.0, 54.0)	-0.3
5	5	77.2 (18.8)	85.4 (14.5)	8.2 (0, 18.0)	0.4
6 or 7	4	56.5 (25.2)	67.6 (33.1)	11.1 (-18.5, 57.0)	0.4
Awake cough frequency					
≥30% reduction					
Responder	126	55.9 (21.8)	26.9 (20.1)	-29.0 (-84.0, 35.0)	-1.3
Nonresponder	102	59.2 (22.6)	48.3 (25.4)	-10.9 (-65.0, 57.0)	-0.5
≥50% reduction					
Responder	79	57.8 (20.8)	21.5 (16.5)	-36.2 (-84.0, 24.0)	-1.7
Nonresponder	149	57.2 (22.9)	44.4 (25.1)	-12.8 (-82.0, 57.0)	-0.6
≥70% reduction					
Responder	52	56.7 (21.4)	18.0 (15.7)	-38.8 (-84.0, 10.0)	-1.8
Nonresponder	176	57.6 (22.4)	41.9 (24.6)	-15.6 (-82.0, 57.0)	-0.7
≥ 0.30 SD reduction					
Responder	118	59.3 (20.8)	29.0 (22.1)	-30.4 (-84.0, 31.0)	-1.5
Nonresponder	110	55.3 (23.4)	44.5 (25.5)	-10.8 (-73.0, 57.0)	-0.5
PGIC, patient global impression of change; SD, standard deviation; VAS, visual analog scale.					
^a Calculated as Week 4 score - baseline score.					
^b Calculated as score difference/SD of baseline score.					

of living with severe cough may experience greater variability in perceived cough severity compared with the impact of cough or disruption to activities captured by other PRO instruments.

Although the ROC curve analyses supported 2 potential definitions of a clinically meaningful threshold for change in cough severity VAS (i.e., ≥ 20 - or ≥ 30 -mm thresholds), a change in cough severity VAS of ≥ 20 mm may be reflective of a minimum meaningful change threshold for use in a clinical practice setting. Of note, the ROC analyses demonstrated that the 20-mm-change

threshold had an equivalent Youden index value to that of the 30-mm-change threshold in patients who reported greater changes on the PGIC scale (i.e., PGIC score of 1 or 2). Moreover, patients reporting themselves as minimally improved on the PGIC (i.e., PGIC score of 3) had mean cough severity VAS improvements between 16 and 19 mm at Weeks 4 and 12. However, because of the high placebo response rate observed in clinical studies in cough,³⁶ a greater threshold of a ≥ 30 -mm change aligned with the highest categories of improvement measured on the PGIC may be warranted. Additionally, the test-retest reliability of

Table 6. Receiver operating characteristic curve analyses for cough severity VAS score thresholds predictive of PGIC scores.

VAS change score threshold, mm	Sensitivity	Specificity	PPV	NPV	Youden index
PGIC score of 1, 2, or 3					
≤-10	0.76	0.64	0.83	0.54	0.41
≤-15	0.68	0.76	0.87	0.50	0.44
≤-18	0.64	0.81	0.89	0.49	0.46
≤-20	0.61	0.86	0.91	0.48	0.47
≤-30	0.48	0.89	0.91	0.42	0.37
≤-40	0.38	0.93	0.93	0.39	0.31
≤-50	0.24	0.99	0.98	0.35	0.22
≤-60	0.12	1.00	1.00	0.32	0.12
PGIC score of 1 or 2					
≤-10	0.83	0.47	0.48	0.82	0.29
≤-20	0.76	0.70	0.59	0.83	0.45
≤-30	0.66	0.79	0.65	0.80	0.45
≤-40	0.55	0.86	0.71	0.77	0.42
≤-50	0.37	0.95	0.80	0.72	0.31
≤-60	0.17	0.97	0.79	0.67	0.15
≤-70	0.07	1.00	1.00	0.65	0.07

NPV, negative predictive value; PGIC, patient global impression of change; PPV, positive predictive value; VAS, visual analog scale.

Table 7. Proportion of cough severity VAS responders by PGIC group at Week 4.

PGIC score category, n (%)	VAS score change of ≥ 30 mm	
	Responder (n = 88)	Nonresponder (n = 147)
1 or 2	57 (65)	30 (20)
3	23 (26)	55 (37)
4	8 (9)	53 (36)
5	0	5 (3)
6 or 7	0	4 (3)

PGIC, patient global impression of change; VAS, visual analog scale.

the VAS was only moderate, and VAS scores can exhibit high variability as evidenced by the distribution-based estimates. Ultimately, these considerations suggest that a more conservative cough severity VAS threshold change of ≥ 30 mm is justified for clinical research settings to reliably assess treatment effects of novel antitussives. Although it may be assumed that the ≥ 20 - and ≥ 30 -mm threshold changes would correspond to 2- and 3-point reductions on 10-point scales often used in a clinical practice setting (e.g., a Borg VAS), further studies are needed to assess whether the clinically meaningful thresholds for change identified in this study translate to these other scales.

These study findings suggest the single-item cough severity VAS is a useful cough-related outcome measure to assess changes in cough severity in patients with chronic cough. The responsiveness of the cough severity VAS in this study was supported by large effect sizes of this measure among patients categorized as PGIC 1 or 2 or PGIC 3. However, it is worth noting that the cough severity VAS is not designed to capture all aspects of the burden of cough, as patients with low cough severity VAS scores may experience a notable burden from their cough due to different factors (e.g., frequency, intensity, disruption). Thus, it is important for studies assessing antitussives in chronic cough clinical trials to include multiple responsive PRO measures (e.g., the LCQ, CSD) to assess different aspects of patients' cough and their responsiveness to treatment.

This was a large study ($N=253$) of a standardized intervention that assessed correlations with the cough severity VAS and multiple related instruments, using both anchor- and distribution-based approaches to derive clinically meaningful thresholds for change. A potential study limitation is lack of generalizability, as the patient population was a well-selected, blinded population of patients in the United States and United Kingdom with RCC or UCC and a cough severity VAS score of ≥ 40 mm at screening. Additional studies examining a broader, more diverse population of patients with chronic cough irrespective of underlying conditions, as well as a patient population with a lower cough severity VAS score at screening, are warranted to assess the generalizability of these findings. Moreover, patients who have chronic cough for many years may become accommodated to their cough over time, which may alter their perceptions regarding their cough severity.

Therefore, it is possible that a cough severity scale comparing current cough severity to a time before the cough emerged (e.g., “how bad do you feel now vs. before you had a cough?”) may be a useful additional marker of cough severity to investigate in future research. Finally, the use of VAS measures for PROs is susceptible to inherent limitations, including the ceiling effect (which can conceal variation in intensity perceptions) and end-of-scale bias (where respondents are more or less likely to utilize extreme ends of the scale). Respondents may also be unable to make fine distinctions along a 100-mm scale; however, these limitations should not prevent the application or interpretation of VAS as a simple clinical tool.^{37,38}

These findings support the cough severity VAS as a reliable, valid, and responsive measure of cough symptom severity in patients with RCC or UCC that can discriminate clinically meaningful changes in cough severity in clinical trials.

Acknowledgements

Medical writing and editorial assistance were provided under the direction of the authors by Nathan Rodeberg, PhD, and Jenna Lewis, MA, ELS, of MedThink SciCom. This assistance was funded by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA.

Author contributions

All listed authors meet the criteria for authorship set forth by the International Committee of Medical Journal Editors and have significantly contributed to, seen, and approved the final submitted version of the manuscript.

Conflict of interest statement

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: AMN, CLR, DM, and JS are employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, and may hold stock or stock options in Merck & Co., Inc., Kenilworth, NJ, USA. EDB is an employee of Evidera, which provides consulting and other research services to pharmaceutical, medical device, and related organizations. In her salaried position, 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. MV reports nonfinancial support from Merck and personal fees from Evidera during the conduct of the study. SSB reports grants from Merck; personal fees for advisory board work from Bayer, Bellus, GSK, Menlo, Merck, Nacion, Sanofi, and Shionogi; and reimbursement for travel expenses from Boehringer Ingelheim.

Funding

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

ORCID iD

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

Supplemental material

Supplemental material for this article is available online.

References

1. Irwin RS, French CL, Chang AB, *et al.* Classification of cough as a symptom in adults and management algorithms: CHEST guideline and expert panel report. *Chest* 2018; 153: 196–209.
2. Morice AH, Millqvist E, Bieksiene K, *et al.* ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. *Eur Respir J* 2020; 55: 1901136.
3. 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.
4. Arinze JT, de Roos EW, Karimi L, *et al.* Prevalence and incidence of, and risk factors for chronic cough in the adult population: the Rotterdam Study. *ERJ Open Res* 2020; 6: 00300-2019.
5. Çolak Y, Nordestgaard BG, Laursen LC, *et al.* Risk factors for chronic cough among 14,669 individuals from the general population. *Chest* 2017; 152: 563–573.
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. French CL, Crawford SL, Bova C, *et al.* Change in psychological, physiological, and situational factors in adults after treatment of chronic cough. *Chest* 2017; 152: 547–562.
8. Kang S-Y, Won H-K, Lee SM, *et al.* Impact of cough and unmet needs in chronic cough: a survey of patients in Korea. *Lung* 2019; 197: 635–639.
9. Koskela HO, Lähti AM and Pekkanen J. Risk factors for repetitive doctor's consultations due to cough: a cross-sectional study in a Finnish employed population. *BMJ Open* 2019; 9: e030945.
10. Koskela HO, Lähti AM and Purokivi MK. Long-term prognosis of chronic cough: a prospective, observational cohort study. *BMC Pulm Med* 2017; 17: 146.
11. Yousaf N, Montinero W, Birring SS, *et al.* The long term outcome of patients with unexplained chronic cough. *Respir Med* 2013; 107: 408–412.
12. McGarvey L and Gibson PG. What is chronic cough? Terminology. *J Allergy Clin Immunol Pract* 2019; 7: 1711–1714.
13. Smith JA, Kitt MM, Morice AH, *et al.* Gefapixant, a P2X3 receptor antagonist, for the treatment of refractory or unexplained chronic cough: a randomised, double-blind, controlled, parallel-group, phase 2b trial. *Lancet Respir Med* 2020; 8: 775–785.
14. Smith JA, Kitt MM, Butera P, *et al.* Gefapixant in two randomised dose-escalation studies in chronic cough. *Eur Respir J* 2020; 55: 1901615.
15. Smith J, Allman D, Badri H, *et al.* The neurokinin-1 receptor antagonist orvepitant is a novel antitussive therapy for chronic refractory cough: results from a phase 2 pilot study (VOLCANO-1). *Chest* 2020; 157: 111–118.
16. Vernon M, Kline Leidy N, Nacson A, *et al.* Measuring cough severity: perspectives from the literature and from patients with chronic cough. *Cough* 2009; 5: 5.
17. Birring SS and Spinou A. How best to measure cough clinically. *Curr Opin Pharmacol* 2015; 22: 37–40.
18. Spinou A and Birring SS. An update on measurement and monitoring of cough: what are the important study endpoints? *J Thorac Dis* 2014; 6(Suppl. 7): S728–S734.
19. Cho PSP, Birring SS, Fletcher HV, *et al.* Methods of cough assessment. *J Allergy Clin Immunol Pract* 2019; 7: 1715–1723.
20. 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.
21. French CT, Irwin RS, Fletcher KE, *et al.* Evaluation of a cough-specific quality-of-life questionnaire. *Chest* 2002; 121: 1123–1131.
22. 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.
23. Martin Nguyen A, Bacci E, Dicpinigaitis P, *et al.* Quantitative measurement properties and score interpretation of the Cough Severity Diary in patients with chronic cough. *Ther Adv Respir Dis* 2020; 14: 1753466620915155.
24. Birring SS, Brew J, Kilbourn A, *et al.* Rococo study: a real-world evaluation of an over-the-counter medicine in acute cough (a multicentre, randomised, controlled study). *BMJ Open* 2017; 7: e014112.
25. Lee KK, Matos S, Evans DH, *et al.* A longitudinal assessment of acute cough. *Am J Respir Crit Care Med* 2013; 187: 991–997.
26. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.
27. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012; 24: 69–71.
28. Martin Nguyen A, Muccino D and Birring S, *et al.* Defining minimal clinically important differences (MCID) in chronic cough: analyses of objective cough counts from a phase 2 randomized controlled trial. *J Allergy Clin Immunol* 2019; 143(Suppl.): AB169.
29. McGarvey L, Birring S, Morice A, *et al.* Late breaking abstract-two phase 3 randomized clinical trials of gefapixant, a P2X3 receptor antagonist, in refractory or unexplained chronic cough (COUGH-1 and COUGH-2). *Eur Respir J* 2020; 56(Suppl. 64): 3800.
30. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
31. Decalmer SC, Webster D, Kelsall AA, *et al.* Chronic cough: how do cough reflex sensitivity and subjective assessments correlate with

- objective cough counts during ambulatory monitoring? *Thorax* 2007; 62: 329–334.
32. Brightling CE, Monterio W, Green RH, *et al.* Induced sputum and other outcome measures in chronic obstructive pulmonary disease: safety and repeatability. *Respir Med* 2001; 95: 999–1002.
33. Koo H-K, Jeong I, Kim J-H, *et al.* Development and validation of the COugh Assessment Test (COAT). *Respirology* 2019; 24: 551–557.
34. Zhan W, Zhang L, Jiang M, *et al.* A new simple score of chronic cough: cough evaluation test. *BMC Pulm Med* 2020; 20: 68.
35. 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.
36. Eccles R. The powerful placebo effect in cough: relevance to treatment and clinical trials. *Lung* 2020; 198: 13–21.
37. Bird SB and Dickson EW. Clinically significant changes in pain along the visual analog scale. *Ann Emerg Med* 2001; 38: 639–643.
38. Garza AG and Wyrwich KW. Health utility measures and the standard gamble. *Acad Emerg Med* 2003; 10: 360–363.

Visit SAGE journals online
[journals.sagepub.com/
home/tar](https://journals.sagepub.com/home/tar)

 SAGE journals