# Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation®





**Review Article** 

# Interpreting Evaluating Respiratory Symptoms<sup>™</sup> in COPD Diary Scores in Clinical Trials: Terminology, Methods, and Recommendations

Nancy Kline Leidy, PhD<sup>1</sup> Donald M. Bushnell, MA<sup>1</sup> Chau Thach, PhD<sup>2</sup> Carolina Hache, PhD<sup>2</sup> Florian S. Gutzwiller, MD, MPH<sup>2</sup>

## Abstract

Accurately interpreting scores on patient-reported outcome (PRO) measures is essential to understanding and communicating treatment benefit. Over the years, terminology and methods for developing recommendations for PRO score interpretation in clinical trials have evolved, leading to some confusion in the field. The phrase "minimal clinically important difference (MCID)" has been simplified to "minimal important difference (MID)" and use of responder thresholds to interpret statistically significant treatment effects has increased. Anchor-based derivation methods continue to be the standard, with specific variations preferred by regulatory authorities for drug development programs. In the midst of these changes, the Evaluating Respiratory Symptoms<sup>™</sup> in COPD (E-RS:COPD) was developed and qualified for use as an endpoint in chronic obstructive pulmonary disease (COPD) drug development programs. This paper summarizes the evolution of terminology and method preferences for the development of recommendations for interpreting scores from PRO measures used in clinical trials, and how these changes are reflected in the E-RS:COPD recommendations. The intent is to add clarity to discussions around PRO endpoints and facilitate use of the E-RS:COPD as a key efficacy endpoint in clinical trials of COPD.

Abbreviations: patient-reported outcome, PRO; minimal clinically important difference, MCID; minimal important difference, MID; Evaluating Respiratory Symptoms<sup>TM</sup> in COPD, E-RS:COPD; forced expiratory volume in 1 second ,  $FEV_1$ ; EXAcerbations of COPD Tool, EXACT<sup>®</sup>; European Medicines Agency, EMA; Food and Drug Administration, FDA; patient global impression of change, PGIC; standard deviation, SD; standard error of measurement, SEM; cumulative distribution function, CDF; probability density function, PDF; effect size, EF; St George's Respiratory Questionnaire, SGRQ; patient global impression of severity, PGIS; clinical global impression of severity, CGIS; clinical global impression of change, CGIC; randomized control trials, RCTs; forced vital capacity, FVC; 6-minute walk test, 6MWT; incremental shuttle walk test, ISWT; Breathlessness, Cough and Sputum Scale, BCSS<sup>®</sup>; patient global impression, PGI

**Funding Support:** Evidera | PPD, who employ Nancy K. Leidy and Donald Bushnell, received funding from Novartis to conduct this review. Trial registration: MPEX: NCT00739648; AZ1: NCT00949975; AZ 2:NCT01023516

Date of Acceptance: September 19, 2022 | Published Online Date: September 21, 2022

**Citation:** Leidy NK, Bushnell DM, Thach C, Hache C, Gutzwiller FS. Interpreting Evaluating Respiratory Symptoms<sup>™</sup> in COPD diary scores in clinical trials: terminology, methods, and recommendations. *Chronic Obstr Pulm Dis.* 2022;9(4):576-590. doi: https://doi.org/10.15326/jcopdf.2022.0307

1. Evidera, Bethesda, Maryland, United States

2. Novartis Pharma AG, Basel, Switzerland

Note: This article has an online supplement

#### Address correspondence to:

Nancy Kline Leidy, PhD Evidera 7101 Wisconsin Avenue, Suite 1400 Bethesda, MD 20814 Email: Nancy.Leidy@evidera.com Phone: (301) 654-9729

#### **Keywords:**

Evaluating Respiratory Symptoms<sup>™</sup> in Chronic Obstructive Pulmonary Disease; patient-reported outcome measure; symptomatic relief; interpretation recommendations for PRO measures; PROs and clinical trials

## Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive disease characterized by airflow limitation and lung tissue destruction which are manifested in the cardinal respiratory symptoms of breathlessness, cough, and sputum production. Spirometry is a standard respiratory function test used to diagnose and monitor COPD. While spirometry is useful for evaluating the efficacy of treatments designed to reduce airflow limitation, it does not capture the severity of respiratory symptoms. In fact, studies have found weak correlations between forced expiratory volume in 1 second (FEV<sub>1</sub>) and patient reports of respiratory symptoms.<sup>1,2</sup> These symptoms are disabling for people with COPD, impeding their ability to perform daily activities and interfering with quality of life. Clinicians include symptom severity as a key factor in patient assessment and treatment modification, and recent advances in drug development seek to address symptomatic improvement in COPD patients with novel therapeutics.<sup>3</sup> Clearly, respiratory symptoms should be a key endpoint in trials testing products that may offer symptomatic relief to patients with COPD.

The Evaluating Respiratory Symptoms<sup>™</sup> (E-RS) in COPD (E-RS:COPD) diary was developed to address the need for a precise measure to test the effect of treatment on the severity of respiratory symptoms in stable COPD in drug development programs. The 11 respiratory symptom items comprising the E-RS:COPD are part of the 14-item EXAcerbations of COPD Tool (EXACT<sup>®</sup>).<sup>4</sup> The EXACT quantifies symptom severity of exacerbations, i.e., events characterized by an acute, sustained worsening in COPD beyond normal day-to-day variability. It can be used to quantify symptom severity and duration in trials enrolling patients who are acutely ill (e.g., anti-infective trials) or to capture unreported symptom-defined events or the symptomatic severity of events associated with health care utilization in longer trials, often 6 to 12 months long.<sup>5-7</sup> The EXACT includes 3 non-respiratory symptoms that are features of exacerbations: fatigue, sleep disturbance, and worry or concern. The E-RS uses the 11 respiratory symptom items from the EXACT to yield

a score representing severity of the cardinal respiratory symptoms of COPD: breathlessness, cough and sputum, and chest symptoms. This instrument was designed for use in clinical trials enrolling stable patients with COPD and evaluating the effects of treatment on respiratory symptom severity over the study period, generally 12 weeks.<sup>8,9</sup> Both the EXACT and E-RS underwent qualification review by the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) and each was approved for use as exploratory endpoints, with additional information on score interpretation in clinical trials requested by both agencies.<sup>10-12</sup> Use of these measures as key trial endpoints would be based on specific drug development programs (contexts of use) and empirically supported recommendations for score interpretation.

Like other clinical trial endpoints, patient-reported outcome (PRO) effects are tested using statistical techniques that yield probability estimates, interpreted in terms of chance. Statistical significance does not necessarily equate to a meaningful treatment effect, however. During the evaluation of new drugs, regulatory reviewers, payers, clinicians, and patients each ask: "Does the empirically observed treatment effect provide a meaningful benefit to patients?" In pharmaceutical trials, the interpretation of clinical efficacy endpoints is often based on experiential and historical data, such as the improvement of  $FEV_1$  in COPD.<sup>13</sup> Scores of PRO endpoints are often less intuitive. Empiricallybased recommendations for PRO score interpretation address this issue. Since the E-RS:COPD was developed, terminology and regulator preferences for specific methods have evolved, which may have led to some confusion or misunderstanding regarding the proposed interpretation recommendations for this measure.

The purpose of this paper is to address confusion related to the interpretation of PROs in clinical research by clarifying the history, terminology, and methods associated with the development of score interpretation recommendations (Part 1) and applying this information to the E-RS:COPD (Part 2). For those interested in how terminology and methods have evolved over time, the paper provides this information, using the E-RS:COPD as a case example. For those interested in the interpretation of E-RS:COPD scores in clinical trials, the paper offers a summary of the methods used to derive recommendations in the context of this history, with new analyses related to the proposed thresholds.

### Part 1: Terminology and Methods for Developing Patient-Reported Outcome Score Interpretation Recommendations – Historical Context

# *Minimal Important Difference Terminology and Food and Drug Administration Guidelines*

Terminology and methods for interpreting PRO scores date back to the 1980s when the term "minimal clinically important difference" (MCID) was first proposed to refer to the "smallest difference (change) in the domain score of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."<sup>14</sup> Using a 15-point Likert-type patient global impression of change (PGIC) (-7 "worse" to +7 "better), changes were defined as small, medium, or large, with the small change considered the MCID. In the 1990s, questions were raised about the extent to which a patient rating could be labeled a clinical variable; reference to clinical was dropped, resulting in the phrase "minimal important difference" (MID) to refer to meaningful within-patient change, and later to within-group and between-group differences.15-17 Over the years, MID (within-group change and between-group difference) values were estimated using various anchor-based approaches, including mapping changes in the PRO score of interest to other PRO or non-PRO scores, or to single-item global impressions (ratings) of change. Distribution-based methods (e.g., 1/2 standard deviation [SD] and standard error of measurement [SEM] of baseline scores) or empirical rules (e.g., 8% of the theoretical range) were also used.

In 2006, the FDA published a draft guidance on the use of PROs in drug development.<sup>18</sup> This guidance described the need to interpret group-level change and between-group differences (i.e., MIDs) in clinical trials. Various anchorbased methods for deriving MIDs were discussed, including mapping changes to existing measures and the single-item global impression of change. Distribution-based methods were characterized as supportive and not appropriate as the sole source for guideline interpretation. Meaningful change in an individual's response to treatment was also considered, with the suggestion that patients be categorized as responders to treatment, based on "pre-specified criteria backed by empirical evidence supporting the responder definition as a measure of benefit."<sup>18</sup> The FDA indicated they may request an a priori MID or an a priori responder definition to serve as a benchmark for interpreting PRO trial results.<sup>18</sup>

The final FDA PRO Guidance was released in 2009 without reference to the MID.<sup>19</sup> The guidance discusses the use of responder definitions to understand and communicate treatment benefit, recommending anchor-based methods to derive these values, with distribution-based methods as supportive. Anchors should be "easier to interpret than the PRO measure itself,"<sup>19</sup> with patient ratings of change mentioned as one approach. The final guidance also suggested visually presenting the entire distribution of change scores by anchor categories in cumulative distribution function (CDF) and probability density function (PDF) figures to further understand score interpretation. Responder thresholds may be shown along the curve to assist with interpretation.<sup>19</sup> While not explicitly stated, approaches for estimating responder definitions are the same as methods used to estimate MIDs.<sup>20</sup>

To summarize, MID, or the occasional reference to the MCID, refers to group-level data (within-group change and between-group differences) that can be interpreted as meaningful based on empirical evidence consistent with the context of use. MID values are expressed as numeric summaries (e.g., mean) and are often used to inform power estimates for statistical analyses of between-group differences during trial design. In contrast, a responder definition refers to patient-level data, and is the threshold of within-patient change in the PRO score that can be interpreted as meaningful, again based on empirical evidence consistent with the context of use. Responder analyses are expressed as the number and percentage of responders/non-responders in each treatment group or the odds ratio of clinical benefit. For hypothesis testing to uncover and understand treatment effects, a 2-step process is often used, with group-level comparisons powered using the pre-specified MID estimate and tested statistically, followed by descriptive responder analyses to assist with interpretation when group-level effects are statistically significant.<sup>20,21</sup>

MID and responder threshold estimates for specific PRO measures can solidify or evolve over time as evidence across studies, samples, and methods converge. Because thresholds may vary across different target populations or trial designs (including placebo-controlled or active comparator studies and differing levels of background therapies), it is important to understand an instrument's context of use and confirm or revise estimates as new contexts arise.

#### Methods for Estimating Minimal Important Difference and Responder Thresholds

Historically, distribution- and anchor-based methods have

formed the quantitative foundation for the development of both MID and responder threshold estimates. Distributionbased methods for characterizing the magnitude of grouplevel change include effect size (ES), a standardized metric with generally accepted standards for small (0.2), medium (0.5), and large (0.8) effects,<sup>22</sup> and an early empirically-based suggestion that ½ SD of baseline scores often corresponds to a medium ES of change.<sup>23</sup> As noted previously, distributionbased methods are not used alone, but rather as part of a triangulation of evidence, i.e., an integration of results based on alternative benchmarks to arrive at the best estimate of meaningful change at the group and individual level that are not method bound.<sup>24</sup>

Anchor-based methods use clinically-based criterion variables (e.g., PRO instruments, clinical measures) or global ratings as standards against which score changes in the PRO measure of interest are evaluated. Briefly, change scores on the PRO measure of interest are computed for pre-specified levels of the anchor variable to yield group-level and patient-level threshold estimates. For an anchor to be considered useful, it must be interpretable in and of itself and correlate with the variable/score of interest (a correlation coefficient of least 0.30 is recommended).<sup>25</sup>

Two types of anchor variables are commonly used, with both considered standards in the field. The first type includes validated measures of the same or related construct with known interpretation values, including other, related PRO measures or clinical anchors. For COPD, the St George's Respiratory Questionnaire (SGRQ), a health-status measure, is often used for this purpose, applying the widely used and accepted 4-point threshold for both MID and responder thresholds.

A second type of anchor is a single-item global assessment of severity or change, reported by the patient (patient global impression of severity [PGIS] or change [PGIC]) and/or clinician (clinical global impression of severity [CGIS] or clinical global impression of change [CGIC]). Because PROs quantify patient perception, PGIS/PGIC ratings are generally preferred. Over time, regulator preferences have evolved from construct-related anchors (related PRO or clinical measures) to simpler, single item global ratings, due in part to ease of interpretation. The preferred form of these global ratings has evolved from retrospective recall (asking patients to rate the degree of change they experienced over the course of the trial) to cross-sectional impressions, with change computation (asking patients to rate the construct [e.g., severity] at baseline and end-of-study and deriving the difference between these ratings). The preferred response

scale has also changed over time, from 15-point to 7-point to 5-point and most recently to 4-point scales. The 4-point PGIS scale asks patients to rate the severity of their symptom(s) by selecting 1 of 4 response options, generally labeled None, Mild, Moderate, and Severe.<sup>26</sup> PGIS assessments are completed at baseline and at pre-specified intervals during the trial and/or end of treatment. The difference between the 2 responses serves as the anchor for threshold estimation. Changes of 1 or 2 points (improvement or worsening) are generally considered *minimal*.

Over time, CDF and PDF plots have been requested by regulatory authorities and journal reviewers to provide visual representation of data to further understand and contextualize the proposed estimates.

## Part 2: Development of Interpretation Recommendations for the Evaluating Respiratory Symptoms in COPD

#### Evaluating-Respiratory Symptom Structure

The 11-item E-RS:COPD is part of the 14 item EXACT daily diary. Study participants complete the diary each evening before bedtime, selecting answers that best describe the severity of each symptom on that day using Likert-type scales. The E-RS:COPD Total score is computed by summing responses to each question; scores range from 0-40 with higher scores indicating more severe symptoms. The 3 subscales embedded in the measure capture breathlessness (5 items; score range 0-17), cough and sputum (3 items; score range 0-12).

# **Evaluating Respiratory Symptoms: COPD Context of Use**

Recommendations for interpreting scores for a given outcome measure are directly linked to the context of use. The E-RS:COPD was designed to quantify the severity of patient-reported respiratory symptoms and test treatment efficacy in clinical trials of COPD. As noted in the regulatory qualification submissions, publications, and user manual,<sup>4,8,9</sup> the E-RS:COPD is intended for use in trials of patients with a clinical diagnosis of COPD, 40 years of age or older, current or former smokers, with stable COPD (defined as exacerbationfree within 60 days prior to enrollment). The E-RS:COPD was validated in samples from 4 studies (1 observational validation dataset and 3 randomized control trials [RCTs]) with the following inclusion criteria:

- Clinical diagnosis of COPD
- FEV1/forced vital capacity (FVC) ratio ≤0.70 postbronchodilator
- FEV<sub>1</sub> % predicted  $\leq$  80%
- $\geq$ 40 years of age
- Current or former smoker
- Smoking history  $\geq 10$  pack years
- Symptomatic: breathlessness, cough, sputum, and/or chest congestion
- ≥1 exacerbation in the previous 6 to 12 months, seen in a clinic or emergency department and treated with steroids or antibiotics or requiring a hospital admission
- No exacerbation requiring treatment within 4 to 6 weeks of enrollment

The E-RS:COPD was designed to serve as a primary or secondary efficacy endpoint in phase 2-4 RCTs evaluating the effect of treatments intended to reduce the severity of respiratory symptoms of COPD. These trials are generally 12 weeks in duration; however, the study length, number and nature of treatment arms, and specific outcome assessments and assessment intervals are determined by the sponsor, based on the target product profile, target claims, and related data requirements. Trials simultaneously examining exacerbation outcomes may last 6 to 12 months.

Content validity of the E-RS:COPD was based on qualitative research with patients in the target population and input from clinical experts.<sup>9</sup> Quantitative measurement properties were tested in multiple datasets.<sup>8,9</sup> These analyses provided evidence of score reliability (test-retest, internal consistency) and construct validity. Responsiveness was shown in participants experiencing meaningful improvements in health status (SGRQ for COPD) and exercise capacity (6-minute walk test [6MWT] and incremental shuttle walk test [ISWT]) over time.<sup>8</sup>

#### Methods for Deriving Interpretation Recommendations

With this context of use in mind, the first evidence to guide interpretation of the E-RS:COPD score was based on data from an observational validation study of 188 participants with clinically stable COPD who completed the E-RS:COPD diary over 7 days, as part of the EXACT development study.<sup>9</sup> This dataset was also used for the initial tests of score reliability and validity. Because of the limited 1-week observation period, analyses of change were not possible and so analyses were limited to distribution-based methods.<sup>9</sup> The  $\frac{1}{2}$  SD estimates and previously unreported SEM values (SD x square root [1- $\alpha$ ]) in this sample are shown in Table 1.

Data from 3 separate RCTs in patients with COPD from the target population were used to test the performance of E-RS:COPD and develop empirically-based recommendations for score interpretation.<sup>8</sup> Distribution-based methods included 1/2 SD and SEM at baseline. Anchor-based methods included 3 criterion constructs with published thresholds for meaningful change: 2 PRO anchors (health status [SGRQ≥4 points] and symptoms [Breathlessness, Cough, and Sputum Scale (BCSS<sup>©</sup>) $\geq$ 1 point]) and 2 clinical anchors (exercise capacity: 6MWT≥26 meters or ISWT≥47.5 meters). Change was examined over a period of 12 weeks. Statistical indicators of magnitude of change (ES and mean percentage change) were also computed. Given the conceptual relationship between symptom severity and health status and observed correlations greater than 0.40 between E-RS:COPD Total and E-RS subscale scores and SGRQ scores across the 3 trials (exceeding the recommended 0.30 threshold), the SGRQ was considered the best anchor for developing interpretation recommendations in the absence of patient global impression data in these trials.

A summary of the distribution- and SGRQ anchorbased estimates for E-RS:COPD scores in the 3 trials<sup>8</sup> is shown in Table 2. It is important to note that, consistent with the triangulation approach, results from the other anchor variables (BCSS, 6MWT, and ISWT) were also considered during the development of the E-RS:COPD interpretation recommendations. Results of those analyses are reported elsewhere.<sup>8</sup>

In keeping with the evolving preferences from health authorities for a visual representation of change in PRO scores, CDF and PDF plots were recently developed using the SGRQ as the anchor variable in each of the 3 clinical trial datasets. Results are shown in Figures 1 and 2 (E-RS:COPD Total Score) and in the online supplement (E-RS:COPD subscale scores). Note that group is defined by the wellestablished SGRQ total score change threshold of 4 points, indicating a meaningful improvement in health status.<sup>27</sup> This is comparable to 2 treatment groups with differential health status improvements, with  $\geq$ 4 points indicating meaningful improvement and <4 points demonstrating no meaningful improvement/treatment effect.

# Table 1. Initial Distribution-Based Interpretation Estimates for Evaluating Respiratory Symptoms in COPD Scores<sup>a</sup>

Scale	½ SD	SEM
E-RS:COPD Total Score	3.35	2.32
Breathlessness	1.85	1.38
Cough and Sputum	1.15	1.20
Chest Symptoms	1.05	0.92

<sup>a</sup>Observational validation study of 188 patients with moderate to severe clinically stable COPD.<sup>9</sup>

SD=standard deviation; SEM=standard error of measurement; E-RS:COPD=Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease

# Table 2. Distribution- and Anchor-Based<sup>a</sup> Estimates for Evaluating Respiratory Symptoms in COPD Scores by Trial<sup>b</sup>

Scale (Range) Trial	Baseline Mean (SD)	Distribution-Based: Baseline		Anchor-Based: Change SGRQ≥4 points	Magnitude of Change in E-RS Scores for SGRQ Responders (≥4 points)	
		½ SD	SEM		Effect Size	% Change
Total Score (0–40)					1	
AZ1 <sup>c</sup>	15.9 (6.0)	1.47	3.00	-2.7	0.44	13%
AZ2 <sup>c</sup>	18.2 (5.99)	1.34	3.00	-3.4	0.61	17%
MPex <sup>d</sup>	15.7 (5.93)	1.88	2.97	-2.5	0.41	18%
Breathlessness (0-1	17)					1
AZ1	7.9 (3.17)	0.78	1.59	-1.1	0.35	10%
AZ2	8.9 (3.12)	0.76	1.56	-1.5	0.50	15%
MPex	8.2 (5.93)	1.57	2.97	-1.0	0.27	17%
Cough and Sputum	(0–11)					
AZ1	4.2 (1.56)	0.81	0.78	-0.8	0.53	17%
AZ2	4.7 (1.60)	0.75	0.80	-1.1	0.72	20%
MPex	4.1 (2.08)	1.44	1.04	-0.7	0.54	13%
Chest Symptoms (0-	–12)	•		1		1
AZ12	3.8 (1.93)	0.47	0.97	-0.7	0.39	8%
AZ20	4.6 (1.91)	0.51	0.96	-0.9	0.51	15%
MPex	3.3 (2.08)	0.42	1.04	-0.8	0.42	14%

<sup>a</sup>St George's Respiratory Questionnaire

<sup>b</sup>Summary of data presented in Leidy et al<sup>8</sup>; Sample sizes: MPex Trial (n=235); AZ12 Trial (N=749); AZ20 Trial (N=597).

<sup>c</sup>Dataset from 12-week, parallel group AZ trials testing AZD9668. AZ1: dose-ranging with patients on a baseline treatment of tiotropium. AZ2: tested 1 dose against placebo with patients receiving budesonide/formoterol.

<sup>d</sup>Mpex Pharmaceuticals, Inc., dataset is from a 6-month trial with levofloxacin (MP-376).

SD=standard deviation; SEM= standard error of measurement, SGRQ=St George's Respiratory Questionnaire; E-RS=Evaluating Respiratory Symptoms; AZ=AstraZeneca

The CDF plots show a clear distinction in symptomatic change between these 2 groups along the center of the continuum of symptomatic change (i.e., between 20 and 80 of cumulative percentage), with a substantial difference noted at 2-points, the proposed meaningful change value for the E-RS:COPD Total score. Differentiation is less clear at the extreme ends of the continuum, due in part to fewer patients at these values and the likelihood of health status



Figure 1 a. Cumulative Distribution Function for Change in Evaluating Respiratory Symptoms in COPD Total Score by St George's Respiratory Questionnaire Change Group--AstraZeneca 1 Trial

E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

optimization or deterioration at the highest and lowest ends of symptomatic change, respectively. These results are replicated across the 3 trials.

CDFs show the proportions of responders and nonresponders for every chosen responder threshold. PDFs show the density of participants achieving each change (per anchor category) and provide further insights into the distribution of the change scores. Although essentially the same information, regulatory bodies frequently ask to see both plots. PDFs for E-RS:COPD Total scores by SGRQ group are shown in Figure 2. CDF and PDF figures for the subscales are provided in the supplement (online supplement Figures 2a1 through 4b3). Median E-RS:COPD change in total scores for the responder and non-responder groups (meaningful change; no meaningful change) are highlighted by the green line, with between-group differences of 2.6, 1.8, and 1.6 for the 3 trials, respectively. These values are lower than the distribution- and anchor-based results shown above and in Table 1, suggesting the proposed E-RS:COPD

recommendations may be conservative estimates, i.e., may underestimate treatment benefit in the proposed context of use.

#### Interpretation Recommendations for the Evaluating Respiratory Symptoms in COPD

Based on the totality of results from the original work and new analyses presented, the E-RS:COPD interpretation recommendations provided in Table 3 are appropriate for the intended target population and trial design, i.e., the original/proposed context of use. These values can be used as the MID for sample size estimation and testing and interpreting group-level treatment effects. These values can also be used to categorize study participants as responders or non-responders for understanding and communicating treatment benefit.





E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

### Discussion

PRO measures are often used to test the extent to which a new treatment has a statistically significant and meaningful effect on how patients feel or function. The outcomes must be meaningful to patients, and the measures must be fit for purpose, precise, and interpretable within the context of use. Methods for developing recommendations for interpreting PRO scores in clinical trials have been driven by standards in the psychometric field and, in the drug development arena, influenced by regulatory information needs and preferences.

In COPD trials, the SGRQ, COPD Assessment Test<sup>28,29</sup> and COPD Questionnaire<sup>30</sup> are useful PRO measures for understanding the effects of treatment on patient perceptions of their health status or health-related quality of life. The developers of each of these instruments have provided information to inform trial

design and score interpretation. However, these measures were not designed to precisely quantify and test the effects of treatment on respiratory symptoms of COPD in drug development programs.

The E-RS:COPD has been recognized by the FDA and EMA as a valid and reliable measure of respiratory symptom severity for use in clinical trials of COPD.<sup>10,11</sup> To date, it has been translated into 57 languages and has been used as an endpoint in at least 20 double blind RCTs in the COPD patient population, serving as an exploratory, secondary, and primary endpoint.<sup>31</sup> It has also been used as an exploratory endpoint in Duaklir<sup>®</sup> trials for marketing authorization.<sup>32,33</sup>

As with all measures, the E-RS:COPD is not without limitations. It is a daily diary, asking patients to rate the severity of their respiratory symptoms at the close of each day, which may seem burdensome to some. Although the E-RS:COPD was designed to be simple and easy for patients to complete electronically, care should





E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

be taken to administer the instrument in a manner that will optimize recording accuracy, including easy to use devices, staff and patient training, and periodic monitoring with encouragement. It is also important to note that the E-RS:COPD was designed for use in clinical trials of COPD, with the specific context of use outlined above. The content of the instrument, and its measurement of breathlessness, cough and sputum, and chest symptoms, intuitively transcend the COPD respiratory symptom experience. It is possible that the instrument could be used in studies of populations other than COPD. This would represent a substantial change in the context of use, however, requiring evaluation content validity, quantitative properties, and of interpretation recommendations for this new context.<sup>19</sup> To date, the instrument has been evaluated for use in asthma-COPD overlap,<sup>34</sup> asthma,<sup>35</sup> and idiopathic pulmonary fibrosis,<sup>36</sup> with results suggesting the instrument may be useful in these populations, although further research is needed.

The E-RS:COPD interpretation recommendations are based on construct-related anchor variables supported by distribution-based methods applied to 3 separate RCTs in the target population. The trials pre-dated regulatory preferences for patient global impression (PGI) anchors, and thus did not include these variables. It is wise for sponsors and investigators to include PGI measures in their studies and re-test the proposed values as part of their research program. Thresholds may need to be adjusted for target populations that differ substantively from the original context of use and/or for specific interventions. For example, values may be different for patients with milder disease or trials testing non-pharmaceutical interventions, such as pulmonary rehabilitation. In addition, designing trials and interpreting treatment effects in placebo-controlled and active comparator studies and with different background therapies can be quite different.

The group-level MID can be useful for estimating

Figure 2a. Probability Density Function for Change in Evaluating Respiratory Symptoms in COPD Total Score by St George's Respiratory Questionnaire Change Group--AstraZeneca 1 Trial



E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

sample sizes for placebo-controlled trials and active comparator studies. Results of threshold (responder) analyses, with the derived, pre-specified interpretation recommendations, help readers and reviewers weigh the proportion of patients in each treatment group experiencing meaningful treatment benefit on the outcome of interest, in this case, symptomatic improvement, relative to other trial endpoints and side effects. Interpreting results of individual trials and trial programs are based on a totality of all results to arrive at the best decision for approval, reimbursement, or clinical practice. For drug development programs, threshold values should be tested in phase 2 trials to confirm or adjust them for the specific context of use. Results and proposed thresholds should be discussed with regulatory authorities prior to use in phase 3 trials. Figure 2b. Probability Density Function for Change in Evaluating Respiratory Symptoms in COPD Total Score by St George's Respiratory Questionnaire Change Group--AstraZeneca 2 Trial



E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

## Conclusion

Although terminology and regulatory preferences for methods to derive interpretation recommendations for PRO measures have evolved, the methods themselves have largely remained unchanged. The goal is to identify a value that can be used to power trials to test the MID and categorize study participants as responders or non-responders to assist with interpretation of results. Anchor-based methods are used to derive these values, including clinical variables and PGI scales, with insight from distribution-based metrics. Recommendations for the E-RS:COPD were based on results of clinical anchor- and distribution-based methods, applied to data from 3 RCTs consistent with the instrument's proposed context of use; the new CDF and PDF figures offer further support for the proposed estimates. Results suggest the recommendations for interpreting scores on the E-RS:COPD in the context of use described here are sound and can be used as a starting point for those who design and interpret trials testing the effects of treatment on respiratory symptoms of COPD. Figure 2c. Probability Density Function for Change in Evaluating Respiratory Symptoms in COPD Total Score by St George's Respiratory Questionnaire Change Group--MPex Trial



E-RS=Evaluating Respiratory Symptoms; SGRQ=St George's Respiratory Questionnaire; BL=baseline

# Table 3. Recommendations for Interpreting Evaluating Respiratory Symptoms in COPD Scores in the Proposed/Original Context of Use

E-RS:COPD Scale	Score Range <sup>a</sup>	Meaningful Threshold
Total Score	0–40	≥2.0
Breathlessness	0–17	≥1.0
Cough and Sputum	0–11	≥0.7
Chest Symptoms	0–12	≥0.7

<sup>a</sup>Higher scores indicate more severe symptoms.

E-RS:COPD=Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease

#### Acknowledgments

**Author Contributions:** All authors contributed to the conception and design of the review and participated in the development and critical review of the manuscript. All authors provided final approval for publication submission and are accountable for the accuracy and integrity of this work.

We also extend our appreciation to Camille Cheng of Novartis and Karin Coyne of Evidera for their comments and work on an earlier version of this paper. Finally, we thank Namita Kaushik and Harneet Kaur for assisting with the submission of this manuscript.

#### **Declaration of Interest**

NKL and DMB are employed by Evidera | PPD who received funding from Novartis to conduct the review. FSG, CH, and CT are employed by Novartis. The development of figures in this paper was supported by Novartis, as part of a contract with Evidera for a related body of work.

### References

- 1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. GOLD website. Published 2020. Accessed June 20, 2022. https://goldcopd. org/wp-content/uploads/2019/11/GOLD-2020-REPORTver1.1wms.pdf
- Jones P, Miravitlles M, Van der Molen T, Kulich K. Beyond FEV1 in COPD: a review of patient reported outcomes and their measurement. *Int J Chron Obstruct Pulmon Dis.* 2012;7:697-709. doi: https://doi.org/10.2147/COPD.S32675
- Rowe SM, Jones I, Dransfield MT, et al. Efficacy and safety of the CFTR potentiator Icenticaftor (QBW251) in COPD: results from a phase 2 randomized trial. *Int J Chron Obstruct Pulmon Dis.* 2020;15:2399-2409. doi: https://doi.org/10.2147/COPD.S257474
- Leidy NK, Murray LT. The COPD biomarkers pipeline: patientreported outcome (PRO) measures for clinical trials of COPD: the EXACT and E-RS. *COPD*. 2013;10(3):393-398. doi: https://doi.org/10.3109/15412555.2013.795423
- 5. Leidy NK, Wilcox T, Jones PW, et al. Standardizing measurement of COPD exacerbations: reliability and validity of a patientreported diary. *Am J Respir Crit Care Med.* 2011;183(3):323-329. doi: https://doi.org/10.1164/rccm.201005-07620C
- Jones PW, Chen WH, Wilcox TK, Sethi S, Leidy NK. Characterizing and quantifying the symptomatic features of COPD exacerbations. *Chest.* 2011;139(6):1388-1394. doi: https://doi.org/10.1378/chest.10-1240
- Leidy NK, Wilcox TK, Jones PW, et al. Development of the EXAcerbations of chronic obstructive pulmonary disease tool (EXACT): a patient-reported outcome (PRO) measure. *Value Health.* 2010;13(8):965-975. doi: https://doi.org/10.1111/j.1524-4733.2010.00772.x
- Leidy NK, Murray LT, Monz BU. Measuring respiratory symptoms of COPD: performance of the E-RS in three clinical trials. *Respir Res.* 2014;15(1):124. doi: https://doi.org/10.1186/s12931-014-0124-z
- Leidy NK, Sexton CC, Jones P, Notte S. Measuring respiratory symptoms in clinical trials of COPD: reliability and validity of a daily diary. *Thorax*. 2014;69(5):443-449. doi: https://doi.org/10.1136/thoraxjnl-2013-204428

- 10. European Medicines Agency (EMA). Draft qualification opinion of qualification of exacerbations of chronic pulmonary disease tool (EXACT), and EXACT respiratory symptoms measure (E-RS) for evaluating treatment outcomes in clinical trials in COPD. EMA website. Published April 10, 2015. Accessed June 20, 2022. https://www.ema.europa.eu/en/documents/regulatoryprocedural-guideline/draft-qualification-opinion-qualificationexacerbations-chronic-pulmonary-disease-tool-exact-exact\_ en.pdf
- 11. Food and Drug Administration (FDA). Attachment to guidance on qualification process for drug development tools. Evaluating respiratory symptoms in chronic obstructive pulmonary disease, a patient-reported outcome instrument for the measurement of severity of respiratory symptoms in stable chronic obstructive pulmonary disease: qualification for exploratory use. FDA website. Published March 2016. Accessed June 20, 2022. https://www.fda.gov/media/96271/download
- 12. Food and Drug Administration (FDA). Attachment to guidance on qualification process for drug development tools. Qualification of exacerbations of chronic pulmonary disease tool for measurement of symptoms of acute bacterial exacerbation of chronic bronchitis in patients with chronic obstructive pulmonary disease FDA website. Published January 2014. Accessed June 20, 2022. https://www.fda.gov/media/87409/ download
- Donohue JF. Minimal clinically important differences in COPD lung function. *COPD*. 2005;2(1):111-124. doi: https://doi.org/10.1081/COPD-200053377
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415. doi: https://doi.org/10.1016/0197-2456(89)90005-6
- 15. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol.* 1994;47(1):81-87. doi: https://doi.org/10.1016/0895-4356(94)90036-1
- Schünemann H, Guyatt GH. Commentary—goodbyeM(C) ID! Hello MID, where do you come from? *Health Serv Res.* 2005;40(2):593-597. doi: https://doi.org/10.1111/j.1475-6773.2005.0k375.x
- Wyrwich KW, Norquist JM, Lenderking WR, Acaster S; Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res.* 2013;22(3):475-483. doi: https://doi.org/10.1007/s11136-012-0175-x

- 18. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, FDA Center for Biologics Evaluation and Research, FDACenter for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;11(4):79. doi: https://doi.org/10.1186/1477-7525-4-79
- 19. U.S. Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. FDA website. Published December 2009. Accessed June 20, 2022. http://www.fda.gov/ downloads/Drugs/GuidanceComplianceRegulatoryInformation/ Guidances/UCM193282.pdf
- McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):163-169. doi: https://doi.org/10.1586/erp.11.12
- 21. U.S. Food and Drug Administration (FDA). Patient-focused drug development guidance public workshop. Incorporating clinical outcome assessments into endpoints for regulatory decision making. FDA website. Published December 6, 2019. Accessed June 20, 2022. https://www.fda.gov/media/132505/download
- 22. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. Lawrence Erlbaum Associates; 1988. https://www. utstat.toronto.edu/~brunner/oldclass/378f16/readings/ CohenPower.pdf
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582-592. doi: https://doi.org/10.1097/01.MLR.0000062554.74615.4C
- Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *COPD*. 2005;2(1):157-165. doi: https://doi.org/10.1081/COPD-200050508
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109. doi: https://doi.org/10.1016/j.jclinepi.2007.03.012
- 26. U.S. Food and Drug Administration (FDA). Patient-focused drug development guidance public workshop. Methods to identify what is important to patients and select, develop or modify fit-forpurpose clinical outcomes assessments. FDA website. Published October 15-16, 2018. Accessed June 20, 2022. https://www. fda.gov/media/116277/download
- 27. Jones PW. St. George's respiratory questionnaire: MCID. COPD. 2005;2(1):75-79.
   doi: https://doi.org/10.1081/COPD-200050513

- Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Leidy NK. Development and first validation of the COPD assessment test. *Eur Respir J.* 2009;34(3):648-654. doi: https://doi.org/10.1183/09031936.00102509
- 29. Jones PW, Harding G, Wiklund I, et al. Tests of the responsiveness of the COPD assessment test following acute exacerbation and pulmonary rehabilitation. *Chest.* 2012;142(1):134-140. doi: https://doi.org/10.1378/chest.11-0309
- 30. Canavan JL, Dilaver D, Clark AL, et al. Clinical COPD questionnaire in patients with chronic respiratory disease. *Respirology*. 2014;19(7):1006-1012. doi: https://doi.org/10.1111/resp.12350
- Bushnell DM, Wilson R, Gutzwiller FS. Use of the E-RS:COPD as an outcome measure in clinical trials: a rapid systematic review. *Chron Obstr Pulm Dis.* 2021;8(4):551-571. doi: https://doi.org/10.15326/jcopdf.2021.0235
- 32. European Medicines Agency (EMA). Duaklir Genuair. EMA website. Published May 12, 2014. Accessed June 20, 2022. https://www.ema.europa.eu/en/medicines/human/EPAR/ duaklir-genuair
- 33. European Medicines Agency (EMA). Annex I summary of product characteristics. EMA website. Published November 19, 2014. Updated August 23, 2019. Accessed June 20, 2022. https:// www.ema.europa.eu/en/documents/product-information/ duaklir-genuair-epar-product-information\_en.pdf
- Nelsen LM, Lee LA, Wu W. Reliability, validity and responsiveness of E-RS:COPD in patients with spirometric asthma-COPD overlap. *Respir Res.* 2019;20(1):107. doi: https://doi.org/10.1186/s12931-019-1070-6
- 35. Tabberer M, Von Maltzahn R, Bacci ED. Measuring respiratory symptoms in moderate/severe asthma: evaluation of a respiratory symptom tool, the E-RS®: COPD in asthma populations. *J Patient Rep Outcomes.* 2021;5(1):104. doi: https://doi.org/10.1186/s41687-021-00338-6
- Bacci ED, O'Quinn S, Leidy NK, Murray L, Vernon M. Evaluation of a respiratory symptom diary for clinical studies of idiopathic pulmonary fibrosis. *Respir Med.* 2018;134:130-138. doi: https://doi.org/10.1016/j.rmed.2017.11.011