Evaluating symptoms in chronic obstructive pulmonary disease: validation of the Breathlessness, Cough and Sputum Scale ©

N. K. Leidy, J. K. Schmier, M. K. C. Jones, J. Lloyd and K. Rocchiccioli


Abstract This study examined the reliability and validity of a new measure for evaluating symptoms in patients with chronic obstructive pulmonary disease (COPD): the Breathlessness, Cough and Sputum Scale (BCSS©). Designed as a daily diary, the BCSS is a patient-reported outcome measure that asks patients to rate the severity of the three symptoms, each on a 5-point scale; higher scores indicate more severe symptoms. Item scores are summed to yield a total score. Secondary analysis of data were from two multinational trials (n=629; 765).

The BCSS item and total scores exhibited evidence of internal consistency (α = 0.70 daily; 0.95 to 0.99 over time) and reproducibility (ICC = 0.77 to 0.88). Correlations (r) with pulmonary function (FEV₁ % predicted, PEF) were -0.01 (n.s.) to -0.36 (P < 0.001). Correlations with the St George’s Respiratory Questionnaire total and SF-36 Physical Functioning subscale were 0.44 to 0.59 (P < 0.001). Breathlessness and total scores differentiated patients by disease severity (P < 0.01) and rescue medication use (P < 0.01). Cough, sputum, and total scores increased with sputum volume (r=0.27, 0.30, 0.31; P < 0.001). Patients for whom treatment was moderately or highly effective reported significant improvements in BCSS scores (P < 0.0001).

Results suggest that the BCSS is a reliable, valid, and responsive patient-reported outcome measure of symptom severity in patients with COPD.

© 2003 Elsevier Science Ltd

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) refers to a cluster of diseases (including chronic bronchitis and emphysema) in which patients experience progressive reduction in expiratory airflow with symptoms of breathlessness, cough, and sputum production. Although pulmonary function is an important objective marker of the degree of airway obstruction, the patient’s self-report of the severity of these symptoms is key to successfully monitoring disease activity, adjusting treatment, and evaluating outcomes (1). There are surprisingly few instruments to assess symptom severity and symptom outcomes in this population. The Baseline Dyspnoea Index (BDI)/Transition Dyspnoea Index (TDI) (2), modified Medical Research Council (MRC) Dyspnoea Scale (3), Oxygen Cost Diagram (4) and Borg Scale (5) have been used to evaluate breathlessness, but do not address cough or sputum. The length and 3-month recall period of the Chronic Lung Disease Severity Index (6) is not conducive to evaluating treatment outcomes in short-term clinical trials. Condition-specific health-related quality of life (HRQL) measures, such as the St George’s Respiratory Questionnaire (SGRQ) (7) or the Chronic Respiratory Questionnaire (CRQ) (8), assess the impact of symptoms on daily functioning and well-being rather than symptom severity per se. To date, there has been no published, validated (reliable, valid, responsive) measure to assess day-to-day symptom severity in COPD.

The purpose of this study was to assess the reliability, validity, and responsiveness of the Breathlessness, Cough and Sputum Scale (BCSS©). The BCSS is a patient-reported outcome measure that was designed as a daily
diary in which subjects are asked to record the severity of three symptoms of COPD: breathlessness, cough, and sputum (Appendix A). Each symptom is represented by a single item which is evaluated on a 5-point Likert-type scale ranging from 0 to 4, with higher scores indicating a more severe manifestation of the symptom. Total score is expressed as the sum of three item scores, with a range of 0 to 12. Weekly and period (baseline, treatment, follow-up) scores can be computed by aggregating daily scores over time.

**METHODS**

**Design**

This study used secondary analyses of blinded data from two multicentre, multinational, randomized, double-blind, placebo-controlled clinical trials evaluating the safety and efficacy of AR-C68397AA (Viozan™, sibenadet HCl), a novel dual dopamine D2-receptor/β2-adrenoreceptor agonist for use in patients with COPD (9–14). Both trials involved a 7- to 14-day baseline period followed by 4 (trial 1) or 6 (trial 2) weeks of treatment. The data analysis plan was formulated by the first two authors (NKL, JKS) based on the study protocols before receiving the data; these authors performed the analyses and were blind to treatment throughout the evaluation.

**Study subjects**

Patient selection criteria were similar in both trials. All patients were between 45 and 70 years of age with a medical diagnosis of COPD = 2 years, a smoking history, and a per cent predicted forced expiratory volume in one second (FEV1% predicted) between 20% and 70%. Exclusion criteria included a complicating comorbid condition or need for domiciliary oxygen. In trial 2, patients must have recorded a BCSS total score of 2 or more on 7 consecutive days during the baseline period.

The research protocols were approved by the appropriate institutional review boards, and all patients gave written informed consent prior to enrolment in the trial.

**Measures**

**Clinical assessments**

Baseline FEV1, % predicted, peak expiratory flow (PEF), rescue medication use, and sputum volume were used as clinical indicators for assessing the construct validity of the BCSS. Evening PEF values (Mini-Wright® PEF meter) and rescue medication use during the previous 24 hours were recorded by patients each day in a diary. For the purposes of these analyses, these values were expressed as the mean daily PEF value and mean number of actuations per day during the baseline period. Sputum volume was gathered in trial 1, during the last 24 hours of the baseline period.

**Symptoms and Health-Related Quality of Life**

The modified Borg scale was used to assess breathlessness prior to spirometry at baseline in trial 1 (15). Specifically, patients were asked to describe their level of dyspnoea on a scale ranging from 0, indicating ‘nothing at all’, to 10, indicating ‘maximal impairment’.

The SGRQ, a condition-specific HRQL measure, was used to evaluate the validity of the BCSS. The instrument consists of 76 items and three subscales (Symptoms, Activity, and Impacts [on daily life]) that can be aggregated to yield a total score. Seven of the items address activity-related dyspnoea (Part 2, Section 2), and four items address cough (Part 2, Section 3, 4). Subscale and total scores range from 0 to 100, with higher scores indicating poorer HRQL (16). The SGRQ has been used widely in research involving patients with COPD (7,16–24). It has been suggested that a change of 4.0 or greater in SGRQ total score be considered clinically meaningful (16).

The Medical Outcomes Study Short-Form 36 Health Survey (SF-36), a generic HRQL measure, was also used to validate the BCSS (data gathered in trial 1 only). This 36-item measure comprises eight subscales that can be aggregated into two summary scores: the Physical Component Summary (PCS) and the Mental Component Summary (MCS) (24,25). Scores range from 0 to 100, with higher scores indicating better HRQL. The SF-36 is widely used to evaluate burden of illness and outcomes of treatment in a variety of health conditions, including COPD (2,26–33), and is often used to assess construct validity of newly developed patient outcome measures.

**Analyses**

Unless otherwise specified, baseline data were used in the analyses.

**Missing data and distributional characteristics**

The rate of missing data and the distributional properties of the BCSS scores were examined first. Of particular interest were floor (the number and percentage of patients reporting the lowest possible score on the measure) and ceiling (number and percentage of patients reporting the highest possible score; worst symptoms possible) effects. These values indicate the extent to which an instrument is capable of detecting symptomatic improvement or deterioration (respectively) over time or with treatment.

**Reliability**

The measurement error associated with the BCSS was examined by assessing its internal consistency and reproducibility (test–retest reliability). The internal
evaluating symptoms in COPD

consistency reliability of the daily BCSS total score and total and item scores aggregated over time were assessed using the Cronbach’s formula for coefficient alpha (α).

Reproducibility of scores from baseline to treatment was assessed using the intraclass correlation coefficient (random-effects model), with Pearson product-moment correlations and mean score differences used to further assess score stability. These analyses were performed across treatment groups in a subsample of patients for whom both physician and patient agreed that the treatment produced no change or was only mildly effective.

Validity

Pearson product-moment correlations between BCSS scores and FEV1 % predicted, evening PEF, sputum volume (trial 2) and HRQL (SGRQ and SF-36 [trial 1]) were used to assess the construct validity of the BCSS. Correlations between the BCSS and the Symptom subscale of the SGRQ, dyspnoea- and cough-specific items from the SGRQ, and the Borg scale (trial 1) were used to assess the measure’s concurrent, convergent and divergent validity.

The BCSS was also subjected to two tests of discriminant validity. Analysis of variance (ANOVA) with post hoc Scheffé test was used to test the extent to which BCSS scores varied by level of disease severity (35,36). The Student’s t-test was used to compare BCSS scores for patients in the lower 25th and upper 25th percentile of rescue medication use. The primary concern in these analyses was the performance of the Breathlessness item of the BCSS, although the discriminant validity of the Cough, Sputum and total scores was also assessed.

Responsiveness

The responsiveness (sensitivity) of the BCSS to within-group change (symptom improvement) and between-group differences (improved vs. stable) was also examined using the paired t-test and Student’s t-test, respectively. Data were analysed across treatment group in a subsample of patients categorized as improved based on patient and physician agreement that treatment was moderately or highly effective. The method for identifying stable patients was described previously.

Exploratory analyses

Three alternative methods for assessing symptom outcomes in COPD using BCSS diary data were also examined: symptom-free days (a score of 0 on BCSS total score), symptom relief (‘good’) days (0 or 1 on each BCSS item), and easy (‘EZ’) breathing days (Breathlessness score of 0 or 1, regardless of other symptoms). ‘Good’ and ‘EZ’ breathing days would offer two new methods for evaluating symptom outcomes in COPD. Because results from the two trials were comparable for these exploratory measures, only results from analyses of data from trial 1 are presented.

Significance testing and interpretation

All statistical analyses were conducted using Statistical Analysis Software (SAS) version 6.12 (SAS Institute, Cary, NC), a significance level of 0.05, and 2-tailed tests. Because this was a psychometric evaluation, no adjustments were made for multiple comparisons. Correlation coefficients (r) and effect size estimates for within- and between-group differences were interpreted using guidelines proposed by Cohen (36), in which ± 0.10 is considered small, ± 0.30 moderate, and ± 0.50 large. To assist in interpretation of change scores, data were examined for convergence between mean change in BCSS total and item scores, per cent change, statistical effect size and change in HRQL.

RESULTS

Sample

Demographic and clinical characteristics of the two study populations used in this analysis are presented in Table I. Across trial populations, most subjects were male, mean FEV1 % predicted was less than 45%, and mean duration of COPD was greater than 5 years.

Missing data and distributional characteristics

Missing data for patients who did not withdraw from the trial was less than 5% across all days. Mean baseline BCSS total scores were 4.14 ± 2.20 and 5.09 ± 1.99 in the two studies. The distribution was near normal, with median values of 4 and 5 and modes of 3 and 6, respectively. Although the distribution fell on the low half of the 13-point scale, floor effects were small (< 2%), indicating symptomatic improvement could be detected. The small ceiling effects (< 1% of the patients reported the highest score possible) also indicate there is room for score movement with increases in symptom severity.

Of the three items, Breathlessness mean scores were the highest (trial I: 1.7 ± 0.9; trial 2: 2.0 ± 0.7), followed by Cough (1.4 ± 0.9; 1.7 ± 0.9) and Sputum (1.10 ± 0.88; 1.37 ± 0.87). Floor effects were lowest for the Breathlessness item (8%; 1%), followed by Cough (9%; 4%) and Sputum (19%; 12%).

Reliability

Internal consistency levels (α) of the day-to-day BCSS total score during the baseline and treatment periods ranged from 0.70 to 0.76 in trial I and from 0.68 to 0.80.
in trial 2 (data not shown). The range of α levels for BCSS scores over the baseline period ranged from 0.94 to 0.97 (Table 2).

Intraclass correlation coefficients for the BCSS total score ranged from 0.77 to 0.78 (Table 2). Pearson product-moment correlation coefficients were similar. BCSS total scores changed from 4.1 (± 2.2) during the baseline period to 3.7 (± 2.0) during treatment in trial 1 (9% change, effect size of 0.12) and from 5.0 (± 1.9) at baseline to 4.9 (± 1.9) during treatment in trial 2 (3% change; effect size = −0.07).

Validity

Pearson product-moment correlation coefficients testing the construct and concurrent validity of the BCSS are shown in Tables 3 and 4. Correlations between the BCSS scores, FEV₁ % predicted and PEF were in the predicted direction and small to moderate in magnitude, with the Breathlessness score the strongest correlate (Table 3). Correlations between the BCSS and sputum volume were also in the predicted direction, with Sputum score the strongest correlate. Sputum volumes by BCSS item responses for the same 24-hour period are shown in Figure 1a,b,c. Patients recording high scores on the Breathlessness, Cough, or Sputum items during this time period (scores of 3 or 4) produced significantly more sputum than those reporting little or no symptom difficulty (scores of 0 or 1; P < 0.05, P < 0.0001, and P < 0.0001, respectively).

Correlations between BCSS total and item scores and SGRQ total, Activity, and Impacts scores, tests of construct validity, were moderate and statistically significant, with the strongest correlations between BCSS and SGRQ total scores (Table 4). The relationships between BCSS total and item scores and the generic SF-36 were also in the predicted direction, small to moderate in magnitude, and statistically significant.

Correlations between BCSS scores and Symptoms subscale, dyspnoea- and cough-specific scores from the SGRQ and the modified Borg scale (trial 1) indicate concurrent validity and provide data on the convergent and divergent validity of the three items comprising the BCSS. Correlations between the BCSS and the SGRQ Symptoms subscale were statistically significant, moderate to large in magnitude, and consistent across the two trials (Table 3). The correlation between the Breathlessness score and Borg scale score was also moderate (Table 3); the fact that this relationship was stronger than the correlation between the Borg scale score and either Cough or Sputum scores is evidence of the convergent and divergent validity of the BCSS.
Similar results were observed with dyspnoea- and cough-specific items from the SGRQ (Table 3). Correlations between the dyspnoea-specific item scores from the SGRQ and BCSS Breathlessness item were moderate and stronger than correlations between these SGRQ item scores and the BCSS Cough or Sputum scores. Similarly, correlations between cough-specific item scores from the SGRQ and BCSS Cough scores were stronger than correlations between the SGRQ Cough item scores and BCSS Breathlessness score. SGRQ Cough scores also correlated with the BCSS Sputum score.

Mean BCSS total scores increased as disease severity (FEV\(_1\), % predicted, American Thoracic Society severity classification (34) increased (Figure 2). The Breathlessness score exhibited the strongest evidence of discriminant validity, differentiating among the four severity groups in both trials (F=7.03, R\(^2\)=0.04, P < 0.001; F=20.61, R\(^2\)=0.10, P < 0.001, respectively; data not shown). Results were similar using the European Respiratory Society severity classification (35), with mean BCSS total scores increasing as disease severity increased (from 3.6 to 4.3 in trial 1 [F=2.53, R\(^2\)=0.01, P < 0.09] and from 4.6 to 5.3 in trial 2 [F=6.45, R\(^2\)=0.02, P < 0.01]).

The BCSS total score was also able to discriminate between high and low rescue medication use in both data sets (Figure 3). The BCSS Breathlessness score differentiated the two groups (t=3.26, P < 0.001; t=10.28, P < 0.0001) as did Sputum score (t=2.91, P < 0.01) and Sputum score (t=3.65, P < 0.001). The BCSS Cough scores differentiated the groups in trial 2 only (t=2.91, P < 0.01; data not shown).

### Table 2. Reliability estimates for the BCSS

<table>
<thead>
<tr>
<th>BCSS item:</th>
<th>Internal consistency*</th>
<th>Reproducibility†</th>
<th>Mean difference (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1 (n=558)</td>
<td>Trial 2 (n=574)</td>
<td>Trial 1 (n=433)</td>
</tr>
<tr>
<td>BCSS total</td>
<td>0.97</td>
<td>0.97</td>
<td>0.78</td>
</tr>
<tr>
<td>Breathlessness (B)</td>
<td>0.95</td>
<td>0.94</td>
<td>0.78</td>
</tr>
<tr>
<td>Cough (C)</td>
<td>0.96</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td>Sputum (S)</td>
<td>0.97</td>
<td>0.97</td>
<td>0.77</td>
</tr>
</tbody>
</table>

* Estimates based on the Cronbach α computed across daily BCSS scores over the baseline period; † test–retest reliability of the BCSS from baseline to treatment in patients for whom both the clinician and patient reported minimal to no change with treatment.

### Table 3. Pearson product-moment correlations between BCSS score, pulmonary function and symptom scores in trials 1 and 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>BCSS total</th>
<th>Breathlessness</th>
<th>Cough</th>
<th>Sputum</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV(_1), % predicted</td>
<td>629</td>
<td>-0.13*</td>
<td>-0.25*</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>PEF</td>
<td>629</td>
<td>-0.31*</td>
<td>-0.36*</td>
<td>-0.19*</td>
<td>-0.22*</td>
</tr>
<tr>
<td>Sputum volume†</td>
<td>618</td>
<td>0.31*</td>
<td>0.16*</td>
<td>0.27*</td>
<td>0.30*</td>
</tr>
<tr>
<td>Borg scale (resting)</td>
<td>626</td>
<td>0.33*</td>
<td>0.48*</td>
<td>0.13*</td>
<td>0.20*</td>
</tr>
<tr>
<td>SGRQ dyspnoea items</td>
<td>330</td>
<td>0.18*</td>
<td>0.38*</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>SGRQ cough items</td>
<td>329</td>
<td>0.41*</td>
<td>0.29*</td>
<td>0.35*</td>
<td>0.36*</td>
</tr>
</tbody>
</table>

* Baseline data were used in these analyses. Sample sizes vary due to missing data; the SGRQ was not completed in all countries due to lack of a validated measure in these languages. † P < 0.001; ‡ daily BCSS scores for the 24 hours corresponding to sputum collection were used in these analyses; ‡ P < 0.05; ‡ P < 0.01.
Responsiveness

Significant improvements in BCSS total and item scores were observed for patients categorized as having improved during the course of treatment (Figure 4). Patients in the 'improved' group reported significant improvement in BCSS total and item scores from baseline to the end of treatment, with scores significantly better than those of the stable group. With the exception of the Sputum score in trial 1, all between-group differences (improved vs. stable patients) were statistically significant.

To facilitate the interpretation of changes in BCSS scores, data were examined for points of convergence between mean change, per cent change, and statistical effect size for BCSS scores and change in HRQL (Table 5). In trial 1, the mean score change of -1·1 on the BCSS total score corresponded to 28% improvement in symptom severity and a moderate to large effect size. For patients with HRQL data, the mean improvement in SGRQ total score exceeded the guideline of 4·0, indicating a clinically meaningful change. In trial 2, the mean change in BCSS total score of -0·9 corresponded to an 18% improvement and a moderate to large effect size, and the mean change in SGRQ total score exceeded the guideline for clinically meaningful change.

Exploratory analyses

Patients who improved during trial 1 experienced significantly more symptom-free days than patients who were stable during the 28-day treatment period (5·7 ± 8·5 vs. 2·1 ± 5·9 days; P < 0·05). These patients also experienced significantly more symptom-relief ('good') days (16·1 ± 11·2 vs. 10·6 ± 10·8) and 'EZ' breathing days (18·7 ± 10·6 vs. 13·4 ± 11·3; P < 0·05).
Table 4. Pearson product-moment correlations between BCSS and health-related quality of life scores in trials 1 and 2

<table>
<thead>
<tr>
<th>Trial</th>
<th>n</th>
<th>BCSS total</th>
<th>Breathlessness</th>
<th>Cough</th>
<th>Sputum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1 SGRQ*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>328</td>
<td>0.44†</td>
<td>0.59†</td>
<td>0.21†</td>
<td>0.59†</td>
</tr>
<tr>
<td>Activity</td>
<td>328</td>
<td>0.24†</td>
<td>0.48†</td>
<td>0.04</td>
<td>0.48†</td>
</tr>
<tr>
<td>Impact</td>
<td>328</td>
<td>0.40†</td>
<td>0.54†</td>
<td>0.18†</td>
<td>0.54†</td>
</tr>
<tr>
<td>Symptoms</td>
<td>330</td>
<td>0.54†</td>
<td>0.45†</td>
<td>0.11†</td>
<td>0.45†</td>
</tr>
<tr>
<td>SF-36†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCS</td>
<td>314</td>
<td>-0.42†</td>
<td>-0.54†</td>
<td>-0.22†</td>
<td>-0.54†</td>
</tr>
<tr>
<td>MCS</td>
<td>314</td>
<td>-0.16†</td>
<td>-0.21†</td>
<td>-0.04</td>
<td>-0.21†</td>
</tr>
<tr>
<td>Trial 2 SGRQ*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>654</td>
<td>0.49†</td>
<td>0.55†</td>
<td>0.29†</td>
<td>0.36†</td>
</tr>
<tr>
<td>Activity</td>
<td>654</td>
<td>0.35†</td>
<td>0.52†</td>
<td>0.14†</td>
<td>0.21†</td>
</tr>
<tr>
<td>Impact</td>
<td>654</td>
<td>0.42†</td>
<td>0.47†</td>
<td>0.25†</td>
<td>0.31†</td>
</tr>
<tr>
<td>Symptoms</td>
<td>665</td>
<td>0.56†</td>
<td>0.43†</td>
<td>0.46†</td>
<td>0.48†</td>
</tr>
</tbody>
</table>

Baseline data were used in these analyses. Sample sizes vary due to missing data; the SGRQ was not completed in all countries due to lack of a validated measure in these languages; † higher scores indicate better HRQL; ‡ P < 0.001; † P < 0.01; ‡ higher scores indicate poorer HRQL; PCS = Physical Component Score, MCS = Mental Component Score.

Table 4: Pearson product-moment correlations between BCSS and health-related quality of life scores in trials 1 and 2.

**DISCUSSION**

This secondary analysis examined the distributional properties, reliability, validity, and responsiveness of the BCSS, a new three-item measure for evaluating the severity of three symptoms associated with COPD: breathlessness, cough and sputum. The distribution of BCSS scores was near normal, with low floor and ceiling effects, indicating the measure is capable of detecting improvement or deterioration in symptoms. Internal consistency reliability levels of the daily BCSS total score exceeded the recommended criterion of 0.70 considered sufficient for predictive or construct validation research (13). Internal consistency levels over time exceeded the 0.80 criterion for experimental research and the standard of 0.95 when individual-level data are being used for decision-making purposes (37).
Figure 3. Mean BCSS total score of patients classified by rescue medication use.

Figure 4. Mean BCSS change score by change group (improved vs. stable): within-group and between-group differences.

Reproducibility estimates for the BCSS suggest scores are replicable with values meeting or exceeding values reported for measures such as the Symptoms subscale of the SGRQ (21,38,39).

The validity of the BCSS was evaluated by examining the relationship between the item and total scores and FEV\(_1\) % predicted, PEF, sputum volume, modified Borg scale, and indicators of HRQL, including the specific...
EVALUATING SYMPTOMS IN COPD

impact of cough and dyspnoea. Results suggest the BCSS item scores are valid indicators of the severity of patient breathlessness, cough and sputum, respectively. Correlations between FEV₁ % predicted and the BCSS Breathlessness score were small to moderate in magnitude and consistent with correlations between FEV₁ % predicted and BDI and CRQ Dyspnoea subscale score (r=0.31 and 0.35, respectively) reported by Hajiro et al. (21) and with the MRC dyspnoea scale score (Spearman p=0.37) reported by Eltayara et al. (3).

One of the advantages of the BCSS is the simultaneous inclusion of breathlessness, cough and sputum assessments. Results suggest BCSS scores are valid indicators of the severity of cough and sputum in patients with COPD. Scores for both items were correlated with sputum volume. Patients who reported 'severe' cough and sputum (scores of 3 or 4) produced significantly more sputum than those reporting little or no symptom difficulty (scores of 0 or 1). Scores on these items also correlated with cough-specific HRQL items from the SGRQ.

Correlations between the BCSS total scores and indicators of HRQL suggest that the BCSS total score is a valid summary indicator of symptom severity. Correlations between the SGRQ Symptom and BCSS total scores were similar to reported correlations between this SGRQ subscale and resting Borg (r=0.43) (40) and the Oxygen Cost Diagram (r=0.46) (41) and larger than correlations between SGRQ symptoms score and daily assessments of dyspnoea (r=0.27), sputum (r=0.30) and bronchitic symptoms (r=0.26) (42). The BCSS total score also increased as the degree of airway obstruction increased and discriminated between high and low rescue medication users. These associations suggest that the measure is sensitive to variation in disease severity.

Results of this study also suggest that the BCSS total and item scores are sensitive to within-group change and between-group differences. Patients in the ‘improved’ group experienced 5 more ‘good’ days and over 5 more ‘EZ’ breathing days during the 28-day treatment period than those in the stable group. This suggests that the daily diary approach may offer a new method for understanding symptomatic improvement in the COPD population in terms of number of ‘good’ and ‘EZ’ breathing days gained with treatment. Based on per cent change, effect size, mean HRQL changes and the number of ‘good’ days observed in the improved patients, the mean change of ±1 point on the BCSS total score appears to be a substantial improvement in symptom severity for patients with moderate to severe COPD.

The BCSS was designed to evaluate the severity of three symptoms associated with COPD: breathlessness, cough, and sputum. Results of this study suggest that the BCSS is reliable, valid and responsive to change. Information about an instrument’s reliability and validity accrues over time, and confidence in the instrument increases as it is used in repeated investigations by various scientists and in diverse conditions and settings. Results of these secondary analyses indicate that further use of this patient-reported outcome measure as a clinical tool and outcome measure in intervention studies involving patients with COPD is warranted.

ACKNOWLEDGMENTS

The BCSS® was developed by Jane Lloyd and Karen Rocchiccioli of AstraZeneca. Copyright for the BCSS is owned by AstraZeneca. Nancy Kline Leidy and Jordana Schmier were funded by AstraZeneca to design and perform the analyses described. M. Kathryn C. Jones, Jane Lloyd and Karen Rocchiccioli are employees of AstraZeneca. The authors thank Vivian Alderfer, Chris Blango, Mitch Goldman, Trish Hepburn, Libby Keating, and Libby Pethick of AstraZeneca for their insightful comments on the data analysis plan, study report, and manuscript; Christine Thompson for SAS programming; and Dr Leah Kleinman of MEDTAP International Inc. for comments on an earlier version of the manuscript.
REFERENCES


Appendix A. Breathlessness, Cough and Sputum Scale (BCSS®)

The BCSS was used in clinical trials as part of a patient daily diary card and is highlighted in bold below in this sample instruction page from the diary card.

**PLEASE COMPLETE IN THE EVENING (Prior to going to bed)**

**PLEASE ENTER DAY, e.g. Mon, Tues, Wed**

**PLEASE RECORD THE DATE (Day/Month)**

**HOW MUCH DIFFICULTY DID YOU HAVE BREATHING TODAY?**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None – unaware of any difficulty</td>
</tr>
<tr>
<td>1</td>
<td>Mild – noticeable during strenuous activity (e.g. running)</td>
</tr>
<tr>
<td>2</td>
<td>Moderate – noticeable during light activity (e.g. bedmaking)</td>
</tr>
<tr>
<td>3</td>
<td>Marked – noticeable when washing or dressing</td>
</tr>
<tr>
<td>4</td>
<td>Severe – almost constant, present even when resting</td>
</tr>
</tbody>
</table>

**HOW WAS YOUR COUGH TODAY?**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None – unaware of coughing</td>
</tr>
<tr>
<td>1</td>
<td>Rare – cough now and then</td>
</tr>
<tr>
<td>2</td>
<td>Occasional – less than hourly</td>
</tr>
<tr>
<td>3</td>
<td>Frequent – one or more times an hour</td>
</tr>
<tr>
<td>4</td>
<td>Almost constant – never free of cough or need to cough</td>
</tr>
</tbody>
</table>

**HOW MUCH TROUBLE WAS YOUR SPUTUM TODAY?**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None – unaware of any difficulty</td>
</tr>
<tr>
<td>1</td>
<td>Mild – rarely caused problem</td>
</tr>
<tr>
<td>2</td>
<td>Moderate – noticeable as a problem</td>
</tr>
<tr>
<td>3</td>
<td>Marked – caused a great deal of inconvenience</td>
</tr>
<tr>
<td>4</td>
<td>Severe – an almost constant problem</td>
</tr>
</tbody>
</table>