BMJ Open

Shortness of Breath with Daily Activities questionnaire: validation and responder thresholds in patients with chronic obstructive pulmonary disease

Michael L Watkins,[1] Teresa K Wilcox,[2] Maggie Tabberer,[3] Jean M Brooks,[3] James F Donohue,[4] Antonio Anzueto,[5] Wen-Hung Chen,[2] Courtney Crim[1]

For numbered affiliations see end of article.

Correspondence to
Dr Michael Watkins;
michael.l.watkins@gsk.com

ABSTRACT

Objectives: To test the reliability, validity and responsiveness of the 13-item Shortness of Breath with Daily Activities (SOBDA) questionnaire, and determine the threshold for response and minimal important difference (MID).

Design: 6 week, randomised, double-blind, placebo-controlled study.

Setting: 40 centres in the USA between 29 October 2009 and 1 July 2010.

Primary and secondary outcome measures: 547 patients with chronic obstructive pulmonary disease (COPD) were enrolled and 418 entered the 2-week run-in period. Data from the run-in period were collected to test internal consistency, test–retest reliability, convergent validity and known-groups validity of the SOBDA. Three hundred and sixty six patients were randomised 2:2:1 to fluticasone propionate/salmeterol 250/50 μg, salmeterol 50 μg or placebo, twice daily. Results from the SOBDA questionnaire, Patient Global Assessment of Change Question, modified Medical Research Council Dyspnoea Scale (mMRC), Clinician Global Impression of Dysponea Severity (CGI-S), Clinician Global Impression of Change Question and Chronic Respiratory Disease Questionnaire self-administered standardised version (CRQ-SAS) were evaluated; spirometry and safety parameters were measured. Study endpoints were selected to investigate the cross-sectional and longitudinal validity of the SOBDA questionnaire in relation to the clinical criteria.

Results: Internal consistency of the SOBDA questionnaire (Cronbach α) was 0.89. Test–retest reliability (intraclass correlation) was 0.94. The SOBDA weekly scores correlated with the patient-reported and clinician-reported mMRC, CGI-S and CRQ-SAS dyspnoea domain scores (0.29, 0.24, 0.24 and −0.68, respectively). The SOBDA weekly scores differentiated between the responders and the non-responders as rated by the patients and the clinicians. Anchor-based and supportive distribution-based analyses produced a range of the potential values for the threshold for the responders and MID.

Conclusions: The 13-item SOBDA questionnaire is reliable, valid and responsive to change in patients with COPD. On using anchor-based methods, the proposed

ARTICLE SUMMARY

Strengths and limitations of the study
- This study reconfirmed the initial psychometric validation observed in the non-interventional study (A2-4398-003).[2]
- Only patients with modified Medical Research Council Dyspnoea Scale ≥2 were included in the patient population, thereby restricting the shortness of breath severity range.
- Approximately half of the patients did not answer the last Patient Global Assessment of Change question thereby possibly affecting some validity assessments.

responder threshold shows a −0.1 to −0.2 score change. A specific threshold value will be identified as more data are generated from future clinical trials.

Trial registration: NCT00984659; GlaxoSmithKline study number: ASQ112989.

INTRODUCTION

Dyspnoea, sometimes referred to as 'shortness of breath' or 'breathlessness' by the patient, is a common and significant symptom of patients with chronic obstructive pulmonary disease (COPD). In one survey of 3000 patients with COPD, 56% were found to have breathlessness during normal physical activities and 42% reported breathlessness while doing household chores.[1]

Capturing the effect of a treatment intervention on dyspnoea from the patient's perspective is therefore an important objective in order to demonstrate treatment effectiveness. While patient-reported aspects of COPD have been assessed using currently available instruments, most do not adequately address the concept of dyspnoea in patients with COPD for use in clinical

trials, due to the limited assessment of psychometric properties during development of the questionnaire or inconsistent clinical validity in use. In addition, there are no currently available instruments for assessing COPD-related dyspnoea that can support a specific label claim for a medicinal product in the USA.

The Shortness of Breath with Daily Activities (SOBDA) questionnaire is a daily diary questionnaire developed to quantify a patient's perception of dyspnoea related to daily activities and how this changes over time during treatment.[2][3] Development of the SOBDA questionnaire followed the Patient-Reported Outcomes Guidance for drug development issued by the US Food and Drug Administration[4] and included the creation of an endpoint rationale and the development of a conceptual framework.[3] Qualitative research, including individual interviews and patient focus groups, was used to develop potential questions (item pool), item format and response options, which were subject to clinical and translation expert review. Further cognitive interviews with patients were conducted to test content validity.[3] The item pool was tested in a non-interventional study, and the number of items was appropriately reduced to produce the final SOBDA questionnaire.[2] Initial psychometric validation from this non-interventional study showed excellent internal consistency and test–retest reliability.[2]

The objectives of this study were to (1) confirm reliability and validity, (2) evaluate the responsiveness, (3) define the threshold for the responders and also the minimal important difference (MID) of the final SOBDA questionnaire in patients with COPD. The threshold for response was established by comparing the SOBDA change scores for the responders and the non-responders, defined according to a range of established patient-completed and clinician-completed assessments. The study included active treatments to ensure that some patients would be classified as 'responders' based on the established clinical measures.

METHODS
Patients
Male and female patients ≥40 years of age with an established clinical history of COPD in accordance with the American Thoracic Society/European Respiratory Society definitions[5] were recruited. At screening, the patients were required to have a postsalbutamol forced expiratory volume in 1 s (FEV_1) ≤70% of the predicted normal and FEV_1/forced vital capacity (FVC) ratio of <0.70; to be a current or former smoker with a history of at least 10 pack-years; and to demonstrate evidence of dyspnoea as assessed by a patient-reported modified Medical Research Council Dyspnoea Scale (mMRC) score ≥2. The study protocol was institutional review board approved and all the patients provided written informed consent before enrolment.

Study design
This randomised, double-blind, placebo-controlled study was conducted at 40 centres in the USA from 29 October 2009 to 01 July 2010 (trial registration: NCT00984659; GlaxoSmithKline study number: ASQ112989). Patients attended three clinic visits. At screening visit 1, eligible patients entered a 2-week run-in period during which short-acting bronchodilator rescue medications (salbutamol and/or ipratropium) were permitted. At visit 2, eligible patients were randomised (2:2:1) to receive fluticasone propionate/salmeterol combination (FSC) 250/50 µg, salmeterol (SAL) 50 µg or placebo, all administered twice daily via a DISKUS inhaler, for 6 weeks. The FSC and SAL active treatments were included to potentially induce a change in the degree of the patients' symptoms of dyspnoea, which would allow the responsiveness of the SOBDA questionnaire to be assessed. The final dose of study medication was taken on the day before visit 3 (week 6). In the event of a patient not completing the week 6 visit, attempts were made for the patient to attend an early withdrawal visit that included the week 6 assessments.

All non-COPD medications, including pre-existing selective β-blocker therapy, could be continued if their dose remained constant. The concurrent use of inhaled or oral corticosteroids, long-term oxygen therapy, long-acting bronchodilators and theophylline were the exclusion criteria within the study protocol.

Measurements and assessments
Patient-completed measures: SOBDA questionnaire
The 13-item SOBDA questionnaire (box 1) was completed in an electronic diary (e-diary) each evening

> **Box 1** 13-Item Shortness of Breath with Daily Activities (SOBDA) questionnaire
>
> *How short of breath were you when*
> ▶ You put on long pants or stockings?
> ▶ You put on your shoes (sandals)?
> ▶ You washed yourself?
> ▶ You reached above your head to put things away?
> ▶ You cleaned or fixed something at floor level?
> ▶ You put things away in the cupboard or shelf at chest level?
> ▶ You put things away in the cupboard or shelf at knee level?
> ▶ You prepared food or a meal?
> ▶ You picked up light objects off the floor?
> ▶ You carried objects at your side like bags or baskets?
> ▶ You walked at a slow pace?
> ▶ You walked up 3 stairs?
> ▶ You walked up 8 stairs?
> *Response options included*
> ▶ I did not do the activity today
> ▶ Not at all
> ▶ Slightly
> ▶ Moderately
> ▶ Severely
> ▶ So severely that I did not do the activity today

immediately before bedtime, which allowed the patient to reflect on and capture the current day's activities.[2] [3] All items followed the same format: 'How breathless were you when [completing the specified activity]?' Individual item responses are completed on a scale from 'not at all' to 'so short of breath that I did not do the activity'. Items 1–4, 6, 8, 9, 11 and 12 are scored from 1 ('not at all'), 2 ('slightly'), 3 ('moderately') to 4 ('severely' or 'so severely that I did not do the activity today'), and items 5, 7, 10 and 13 are scored from 1 ('not at all' and 'slightly'), 3 ('moderately') and 4 ('severely' or 'so severely that I did not do the activity today'). Patients were also given an option of 'did not do' for activities they did not perform for other reasons. In scoring the questionnaire, these responses were regarded as missing data. Owing to the design of the e-diary, it was not possible for the patients to skip individual questions within the diary, although a full day of data could be missed if the patient did not access the diary within the time window allowed.

Analyses were conducted aggregating daily data over weekly time periods to account for day-to-day variability and the fact that not all the activities were performed every day. A daily SOBDA score was computed across the 13 items as a mean score ranging from 1 to 4, if at least 7 items had non-missing scores. A weekly mean SOBDA score was then computed as the mean of the daily mean scores in a 7-day period, if at least 4 of 7 days had non-missing SOBDA daily scores. The baseline SOBDA weekly score for each patient was calculated as the mean value during the week before randomisation.

Patient-completed measures: other
Additional questions were completed via an e-diary, daily or weekly. Daily questions included any form of contact with healthcare professionals, frequency of rescue medication use and completion of a Global Assessment of Shortness of Breath question: 'Overall, were you short of breath during your activities today?' Patients responded to this question on a 5-point scale from '1=not at all' to '5=extremely'. Every 7 days, patients responded to a Patient Global Assessment of Change (PGAC) question that asked, 'Compared with last week (7 days ago), how was your shortness of breath today?' on a scale of '1=much worse' to '5=much better', with 3='no change'.

Patients completed the mMRC at each clinic visit and the 20-item Chronic Respiratory Disease Questionnaire self-administered standardised version (CRQ-SAS) at visit 2 and week 6/early withdrawal.

Clinician-completed assessments
A Clinician Global Impression of Dysponea Severity (CGI-S) question to assess dyspnoea severity on a scale of 1 (mild) to 4 (very severe) was completed at visit 2 and week 6/early withdrawal. A CGI of Change (CGI-C) question to assess change in dyspnoea on a scale of 1 (much worse) to 5 (much better), with 3 being no change, was completed at week 6/early withdrawal.

Clinicians rated the patient's dyspnoea on the 5-point mMRC scale at each clinic visit.

Spirometry
Spirometry (FEV_1 and FVC) was performed at all clinic visits after the questionnaires were completed. FEV_1 responders were defined as patients who had a change of ≥100 mL from visit 2 to week 6/early withdrawal, whereas FEV_1 non-responders were those patients with a change of <100 mL. Bronchodilator reversibility testing was also performed 30 min postsalbutamol (360 µg) at screening. The predicted FEV_1 values were calculated according to the National Health and Nutrition Examination Survey III reference values.[6]

Safety
Safety was assessed by the reported adverse events (AEs) and COPD exacerbations.

Statistical analyses
Sample size and powering
Sample size calculations were based on the evaluation of the responsiveness of the SOBDA questionnaire[2] [3] and allowed for comparison of the SOBDA change scores for the responders and the non-responders. Calculations assumed 90% power, a two-sided 5% significance level and a standardised between-groups effect size of 0.5 (defined as the difference between the responders and the non-responders divided by the SD of the difference). The sample size was increased to allow exploratory comparisons of the SOBDA scores between treatment arms. Assuming that 90% of randomised patients would provide sufficient data for this comparison and a randomisation ratio of 2:2:1, approximately 350 patients were planned for randomisation in order to provide 320 evaluable patients.

Analyses for the internal consistency, test–retest reliability in a stable population, convergent validity and known-groups validity were based on the data collected from the run-in population. This population consisted of randomised and non-randomised patients who had completed visit 2. The responsiveness to change of the SOBDA questionnaire was based on data collected from the modified intent-to-treat (mITT) population, defined as all patients who were randomised to treatment and who received at least one dose of the study drug, and were analysed according to the treatment actually received if this was different from the randomised treatment assignment.

Internal consistency
To confirm the reliability and validity of the SOBDA questionnaire,[2] the internal consistency of the instrument was assessed and summary scores were compared with other endpoints collected.

The internal consistency of the SOBDA score was assessed for patients with a non-missing score for each item at day 1 of the run-in period by using Cronbach's formula for coefficient α (scale from 0 to 1.0); a value of

0.70 or greater is recognised as indicating acceptable internal consistency for an instrument.[7] Pearson's correlation coefficient and intraclass correlation coefficient (ICC) were used to evaluate test–retest reliability, comparing the SOBDA weekly scores for patients who reported no change on their weekly PGAC assessment during weeks 1 and 2 of the run-in period.

Convergent and known-groups validity

The SOBDA weekly scores were compared with the other relevant study measures to establish the convergent and known-groups validity of the instrument. Convergent validity was assessed by examining the Spearman rank order correlation coefficient between the baseline SOBDA weekly score and the mMRC (patient and clinician) ratings and CGI-S ratings at visit 2. Pearson's correlation coefficient between the baseline SOBDA weekly scores and the CRQ-SAS dyspnoea domain score at visit 2 was also assessed. Known-groups validity, demonstrating that groups of patients who are known to be different report different SOBDA scores, was assessed by comparisons of SOBDA weekly scores between groups of patients based on the mMRC (patient and clinician) ratings and CGI-S ratings collected at visit 2 using analysis of covariance (ANCOVA) models adjusted for age, gender and $FEV_1\%$ predicted measured during the screening visit.

Responsiveness

Responsiveness of the SOBDA questionnaire was assessed by comparing score changes between the responders and the non-responders on the PGAC, CGI-C and CRQ-SAS dyspnoea domain and mMRC. Responders by PGAC and CGI-C were defined as patients with a rating of 'better' or 'much better', and non-responders were defined as patients with a response of 'much worse,' 'worse' or 'no change', on their respective scales. A CRQ-SAS dyspnoea domain responder was defined as a patient with a score increase of 0.5 unit or more between visit 2 and week 6/early withdrawal, and a non-responder was defined as a patient who had a decrease in score or an increase of less than 0.5 unit. A responder by mMRC was defined as a patient who had a score decrease of 1 unit or more between visit 2 and week 6/early withdrawal, and a non-responder was defined as a patient who had the same score or an increase in score.

Changes from the previous week to the current week's SOBDA score during the six-week study treatment period were compared for the responders and the non-responders (defined according to the corresponding weekly PGAC assessment) using ANCOVA, adjusted for age, gender and the baseline SOBDA weekly score. In addition, changes in the mean SOBDA scores during the last week of the treatment were compared for the responders and the non-responders based on the PGAC, CGI-C and CRQ-SAS dyspnoea domain, clinician-completed mMRC and patient-completed mMRC using ANCOVA adjusted for age, sex and the baseline SOBDA weekly score.

Defining the threshold for SOBDA responders and MID

A preliminary MID for SOBDA mean score change within a participant was also determined. This threshold for the SOBDA response allowed a comparison of the proportions of responders in different intervention groups or treatment categories. Anchor-based methods using the PGAC, CGI-C and CRQ-SAS dyspnoea domain scores were used to establish the threshold for the SOBDA responders and the MID, by calculating SOBDA weekly change scores (for PGAC) and changes in SOBDA weekly scores from baseline to the last week of treatment (for the PGAC, CGI-C and CRQ-SAS dyspnoea domain and FEV_1) in the response category or prespecified grouping of 'better' for each anchor. Cumulative distribution plots based on these anchors were also used to determine the MID.

Post hoc supportive analyses using distribution-based approaches were also conducted after completion of the a priori specified anchor-based analyses to further supplement estimation of a responder threshold. A method described by Revicki *et al*[8] was used to estimate the response threshold by calculating 0.2 and 0.3 times the SD of the SOBDA scores at baseline. In addition, thresholds were calculated by the SE of the measurements method.[9]

RESULTS

A total of 547 patients were screened and 418 completed week 2 (screening visit 1) and week 0 (randomisation, visit 2) assessments; 52 patients were not eligible for randomisation. 366 patients met the inclusion criteria and were randomised; however, one patient refused to take study medication; thus, 365 patients received treatment and were included in the mITT (figure 1). Patients were predominantly white (90%), male (57%) with a mean age of 61.1 years (SD, 9.7 years) and a mean body mass index of 28.3 kg/m² (table 1). The majority (62%) of patients were current smokers with an extensive smoking history (mean pack-years, 54.9). The mean postsalbutamol % predicted FEV_1 was 49.9%, indicative of a population with a severe airflow obstruction.

A total of 29 patients withdrew from the study (FSC 9%, SAL 7% and placebo 8%), 13 because of an AE (FSC 5%, SAL 2% and placebo 4%).

Reliability and validity
Internal consistency

Cronbach's α value for SOBDA was 0.89 (n=344). Test–retest reliability was assessed between weeks 1 and 2 of the run-in period for the 152 patients reporting no change on the second weekly PGAC assessment: Pearson's correlation coefficients and ICC were both 0.94, with a mean difference between weeks 1 and 2 of 0.01 on the 4-point SOBDA scale.

Convergent validity

The relationship of the SOBDA weekly scores to patient-reported and clinical assessments of the dyspnoea severity or constructs hypothetically related to the dyspnoea
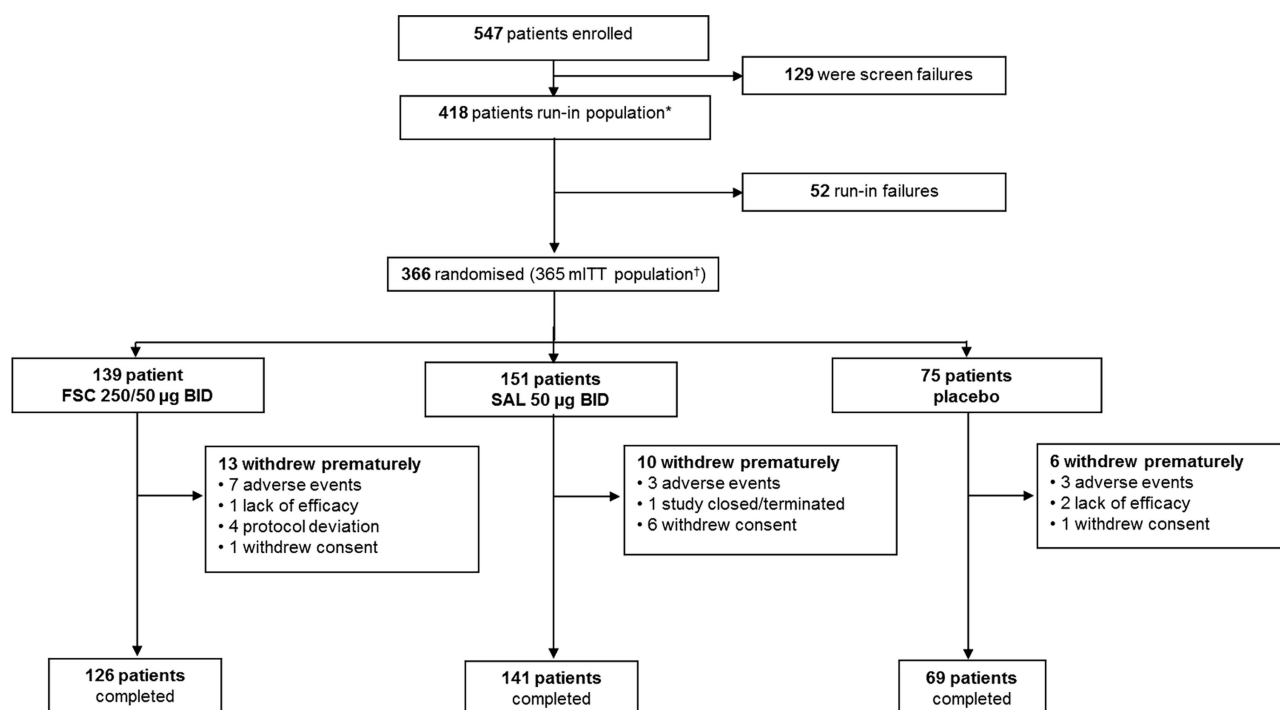
Figure 1 Patient disposition. *Patients who completed visits 1 and 2 including those not randomised. †Patients randomised to treatment and received at least one dose of the study drug. One additional patient was randomised but not treated. BID, twice daily; FSC, fluticasone propionate/salmeterol combination; mITT, modified intent-to-treat; SAL, salmeterol.

severity was examined to assess convergent validity. The Spearman rank-order correlations between the baseline SOBDA weekly scores and mMRC scores were 0.29 (patient reported) and 0.24 (clinician reported), and it was 0.24 for CGI-S. Pearson's correlation between the baseline SOBDA weekly scores and the CRQ-SAS dyspnoea domain score was −0.68 (higher scores in CRQ-SAS, contrary to SOBDA, indicate less dyspnoea and hence the correlation is negative).

Known-groups validity

The known-groups validity was evaluated by determining the extent to which the baseline SOBDA weekly scores differentiated between patients with varying levels of dyspnoea severity as rated on the patient-reported and clinician-reported

mMRC and CGI-S collected at visit 2. The least-squares mean SOBDA weekly scores were increased as the CGI-S and mMRC clinician/patient ratings increased (table 2).

Responsiveness

The SOBDA weekly scores were lower in the PGAC responders than in the non-responders, indicating less dyspnoea with daily activities. Differences between the SOBDA weekly change scores for the PGAC responders and non-responders were statistically significant for each weekly comparison with the exception of week 6 (table 3A).

Changes in the SOBDA weekly score between baseline and the last treatment week were statistically significantly larger for the CGI-C and CRQ-SAS dyspnoea domain responders than for the non-responders (p<0.001). This

Table 1 Demographic and clinical characteristics

	Not randomised	Placebo	FSC 250/50 μg	SAL 50 μg	Total
N	52	75	139	152	418
Age, year (mean (SD))	63.8 (9.6)	62.8 (9.8)	60.2 (9.5)	60.1 (9.6)	61.1 (9.7)
Male, n (%)	25 (48)	46 (61)	79 (57)	89 (59)	239 (57)
White, n (%)	44 (85)	65 (87)	127 (91)	140 (92)	376 (90)
Current smoker, n (%)	29 (57)	46 (61)	84 (60)	99 (65)	258 (62)
Body mass index, mean (SD)	28.3 (6.9)	26.6 (6.1)	29.0 (7.3)	28.5 (6.2)	28.3 (6.7)
Postbronchodilator FEV_1% predicted mean (SD)	50.3 (15.1)	49.4 (13.1)	49.5 (13.7)	50.2 (13.8)	49.9 (13.8)
FEV_1/FVC % (mean (SD))	55.7 (35.2)	51.6 (11.4)	53.7 (11.4)	52.2 (10.9)	53.0 (16.1)
Percentage of reversibility (mean (SD))	8.6 (14.4)	16.7 (19.2)	14.5 (18.5)	11.7 (13.9)	13.1 (16.8)

The 'not randomised' column reflects those patients who completed the visit 1 and 2 assessments but were not eligible to be randomised. The 'total' column reflects the run-in population, defined as patients who completed visits 1 and 2, including those who were not randomised. FEV_1, forced expiratory volume in 1 s; FSC, fluticasone propionate/salmeterol combination; FVC, forced vital capacity; SAL, salmeterol.

Table 2 Known groups validity: the least-squares mean baseline SOBDA weekly score by mMRC and CGI-S response categories at visit 2

Response categories	Patient-completed mMRC (n), LS mean SOBDA score (SE)	Clinician-completed mMRC (n), LS mean SOBDA score (SE)	CGI-S (n), LS mean SOBDA score (SE)
0	12, 1.92 (0.19)	–	–
0–1	–	12, 1.78 (0.20)	–
1	103, 1.94 (0.07)	–	19, 1.87 (0.16)
2	138, 2.20 (0.06)	200, 2.08 (0.05)	236, 2.11 (0.05)
3	65, 2.26 (0.08)	117, 2.28 (0.06)	78, 2.33 (0.08)
4	22, 2.73 (0.14)	10, 2.73 (0.22)	5, 2.72 (0.31)

Owing to the small number of 0 and 1 responses in the clinician-completed mMRC, these two categories were combined.
CGI-S, Clinician Global Impression of Dyspnoea Severity; LS, least squares; mMRC, modified Medical Research Council; SOBDA, Shortness of Breath with Daily Activities.

was not seen with the patient-completed or clinician-completed mMRC or PGAC defined responders, although changes in the last treatment week SOBDA scores were numerically larger for the responders versus the non-responders (table 3B).

Threshold for SOBDA responders and MID

Patients classified as 'better' based on the CGI-C, CRQ-SAS dyspnoea domain (change of >0 to 0.5 units) or FEV_1 (change of >50 to <100 mL) had a mean change in the SOBDA score of –0.25, –0.13 or –0.16, respectively, at the last treatment week compared with the baseline. Patients who rated their dyspnoea as 'better' on the PGAC assessments had a mean change in the SOBDA score of –0.26 at week 1, –0.08 at weeks 2, 3 and 5, –0.10 at week 4 and –0.05 at week 6.

By using the method described by Revicki et al,[8] thresholds of –0.14 and –0.21 were calculated by using 0.2 and 0.3 times the SD of the SOBDA scores at baseline. In addition, a similar threshold of –0.17 was identified by the SE of the measurements method.[9]

Exploratory efficacy analyses
SOBDA treatment group differences

After adjusting for age, sex and the SOBDA baseline score, the difference between the FSC and placebo was –0.09 (95% CI –0.23 to 0.05) and between SAL and placebo, it was 0.03 (95% CI –0.11 to 0.16).

CRQ-SAS

The greatest mean changes for dyspnoea and fatigue were observed in the FSC group (0.4 and 0.3, respectively). The mean changes from the baseline in emotional function were similar between the placebo and the two treatment groups (0.2 and 0.1), as were those for mastery (0.2 for placebo, 0.3 for SAL and 0.4 for FSC). The SAL and FSC groups reported a change of 'better' or 'much better' (56% and 65%, respectively) compared with the placebo group (53%). Thirty-four per cent of the patients receiving placebo were rated as responders, whereas 37% of SAL patients and 46% of FSC patients were responders using this measure.

Spirometry

The mean changes in FEV_1 in the placebo, SAL and FSC groups were 1, 61 and 138 mL, respectively. Forty nine per cent of patients receiving FSC were considered as responders, while 38% of patients receiving SAL and 25% of patients receiving placebo were responders. The majority of the patients in the FSC (62%) and SAL (55%) groups reported a change of 'better' or 'much better', and less than half of the patients in the placebo group (38%) reported this change.

Safety

AEs were reported for the 37 patients (27%) in the FSC group, 34 patients (23%) in the SAL group and 14 patients (19%) in the placebo group. COPD exacerbation, dyspnoea, headache and respiratory tract infection were the most commonly reported AEs with no other individual AEs occurring in ≥3% of patients in any group.

Twelve patients experienced serious AEs (SAEs; FSC, 3 [2%] patients; SAL, 5 [3%] patients; placebo, 4 [5%] patients); three of these SAEs were considered possibly related to study medication (SAL, 1 patient; placebo, 2 patients). One fatal SAE of respiratory failure occurred for a patient receiving FSC during the study, but was not considered to be related to the FSC treatment by the study investigator.

DISCUSSION

SOBDA was developed to address the need for a robust and psychometrically sound patient-reported outcomes questionnaire for use in clinical research that would specifically capture dyspnoea experienced with daily activities as perceived by patients with COPD. The available questionnaires have limited assessment of the psychometric properties, inconsistent clinical validity and/or are not dyspnoea-specific. The CRQ-SAS[10–12] and SGRQ[13 14] questionnaires, for example, measure multiple dimensions that are much broader than dyspnoea with activity, which is the specific aim of the current SOBDA questionnaire. The mMRC questionnaire has been used to discriminate between levels of dyspnoea associated with exercise, but

Table 3 Change in the SOBDA (A) weekly score by PGAC responders, and (B)last treatment week score by assessment responders at visit 3

(A)

	Day 8	Day 15	Day 22	Day 29	Day 36	Day 43
PGAC responders (n)	105	91	83	62	77	31
PGAC non-responders (n)	188	212	216	223	200	88
LS mean difference in SOBDA scores between groups (95% CI)	0.24 (0.18 to 0.31)	0.12 (0.06 to 0.19)	0.11 (0.06 to 0.16)	0.11 (0.06 to 0.17)	0.13 (0.08 to 0.18)	0.06 (−0.03 to 0.15)
p Value*	<0.001	<0.001	<0.001	<0.001	<0.001	NS

(B)

	CGI-C	CRQ-SAS dyspnoea domain	Clinician-completed mMRC	Patient-completed mMRC	PGAC
Responders (n)	120	117	91	92	45
Non-responders (n)	181	184	210	209	106
LS mean difference in SOBDA scores between groups (95% CI)	0.24 (0.14 to 0.34)	0.30 (0.21 to 0.40)	0.03 (−0.08 to 0.15)	0.08 (−0.02 to 0.19)	0.08 (−0.07 to 0.23)
p Value†	<0.001	<0.001	NS	NS	NS

*Comparison of SOBDA scores (non-responders minus responders) based on analysis of covariance adjusted for age, gender and previous week's SOBDA score.
†Comparison of SOBDA scores (non-responders minus responders) based on Analysis of Covariance adjusted for age, gender and baseline SOBDA weekly score.
CGI-C, Clinician Global Impression of Change; CRQ-SAS, Chronic Respiratory Disease Questionnaire self-administered standardised version; LS, least squares; mMRC, modified Medical Research Council; NS, not significant; PGAC, Patient Global Assessment of Change; SOBDA, Shortness of Breath with Daily Activities.

shows very limited response to change in clinical trials due to the limited number of categories for response.

This study confirms that the SOBDA questionnaire has sound psychometric properties. The SOBDA weekly scores had an internal consistency reliability of Cronbach's α value of 0.89, which surpassed the established threshold goal of >0.7.[7] SOBDA also had good test–retest reliability (ICC=0.94), exceeding the threshold goal of >0.60, in patients reporting no change in their breathlessness as measured by the PGAC.[15]

The convergent validity assessed through the Spearman rank order correlations was reasonable, although it was lower than expected for CGI-C and mMRC. This may have been due to the narrow range of responses given by clinicians: most patients were rated as '2' or '3' by clinicians on both scales. The narrow range of clinician mMRC ratings reflects the inclusion criteria requiring patients to have an mMRC ≥2 at study entry. The CRQ-SAS dyspnoea scale, which measures the concept most similar to SOBDA, showed the highest correlation with the SOBDA questionnaire and is supportive of SOBDA's construct validity.

The SOBDA weekly scores in the study population demonstrated good known-groups validity through a series of analyses. The scores were differentiated between dyspnoea severity as rated by clinicians and patients. As expected, discrimination based on patient ratings was better than that based on clinician ratings. Known-groups validity was also confirmed when comparing SOBDA with CGI-S.

Assessment of responsiveness of the SOBDA questionnaire was conducted independent of treatment allocation. Good separation in the SOBDA weekly scores was observed between the PGAC groups at day 8 as indicated by significant differences between the scores for the responders and the non-responders. Less separation was observed between the PGAC groups throughout the later weeks of the 6-week treatment period compared with week 1. This diminished separation may be partially explained by the way the PGAC score was derived, that is, each week's PGAC score was based on scores from the previous week. This is also not an unexpected trend as any improvement in dyspnoea would be expected to occur or be perceptible to patients soon after initiating therapy, with continued improvement being less noticeable over time. The particularly diminished responsiveness observed at week 6 was potentially due to approximately half of the patients not providing a response to the PGAC at day 43 or at the last visit. Changes from baseline in the SOBDA last treatment week scores were statistically significant between responders and non-responders using the CGI-C and CRQ-SAS dyspnoea domain, but not the mMRC. This again may be due to the narrow range of the mMRC ratings.

The thresholds for the SOBDA responders and MID were explored by using anchor-based and distribution-based methods. Anchor-based methods were used to establish a preliminary MID range for SOBDA mean score changes within a patient, which would also be considered

as the threshold for the SOBDA responders to allow a comparison of the proportions of responders in different categories (eg, different interventions or treatments). The evaluation of data around the MID was based on the change from the baseline in the SOBDA score for those patients who endorsed or had the clinician endorse for them, (depending on the anchor), the response category 'better' for the global assessments or the prespecified grouping of meaningful improvement on other measures (PGAC, CGI-C, CRQ-SAS and FEV_1). Based on these anchors, a preliminary response threshold for the SOBDA questionnaire is a −0.1 to −0.2 score change. This is further supported by distribution-based estimations of the MID using methods described by Revicki *et al*[8] and Wyrwich *et al*.[9] Thus, a threshold of −0.1 to −0.2 for the score range of 1–4, supported by anchor-based and distribution-based methods, seems reasonable at this stage of questionnaire development. This MID estimation is also consistent in scale with that of the CRQ-SAS in which the MID is 0.5 on a 7-point Likert scale.[16]

Exploratory analysis by treatment group suggested that the proportion of patients crossing the −0.1 and −0.2 thresholds was numerically greater for the SAL group compared with the placebo, and numerically greater for the FSC group compared with the SAL group. As the study was designed only to validate SOBDA and cannot reliably demonstrate differences between treatment groups, these changes from baseline in the SOBDA weekly score at last treatment can only be regarded as exploratory. Even after adjusting for age, gender and the baseline SOBDA weekly score, the mean change in the score for each treatment group when compared with placebo did not meet the MID of −0.1 or −0.2.

This study had some limitations. Only patients with mMRC ≥2 were included in the study, which restricted the ranges of the dyspnoea severity. The effects of exacerbation and possible cultural differences on the study results were not evaluated. Finally, approximately half of the patients did not answer the last PGAC question despite completing other final visit assessments. These limitations could have had an effect on some of the results of our study, although we do not feel that there would be any change to the overall conclusions.

In summary, this study demonstrates that the 13-item SOBDA questionnaire is reliable, valid and responsive to change in patients with COPD. At this stage of questionnaire development, a change score of −0.1 to −0.2 is the most appropriate estimation for determining a threshold for treatment response. A specific value will be identified as more data are generated from future clinical trials.

Author affiliations
[1]Respiratory Medicines Development Center, GlaxoSmithKline, Research Triangle Park, North Carolina, USA
[2]Evidera (formerly United BioSource Corporation), Bethesda, Maryland, USA
[3]Respiratory Medicines Development Centre, GlaxoSmithKline, Uxbridge, UK
[4]Division of Pulmonary and Critical Care Medicine, University of North Carolina, Chapel Hill, North Carolina, USA
[5]South Texas Veterans Health Care System and University of Texas Health Science Center, San Antonio, Texas, USA

REFERENCES

1. Rennard S, Decramer M, Calverley PM, *et al*. Impact of COPD in North America and Europe in 2000: subjects' perspective of Confronting COPD International Survey. *Eur Respir J* 2002;20:799–805.
2. Wilcox TK, Chen W-H, Howard KA, *et al*. Reliability and validity of the Shortness of Breath with Daily Activities (SOBDA) questionnaire: a new outcome measure for evaluating dyspnea in COPD. American Thoracic Society – 106th International Conference. 2010;181:A5429.
3. Howard K, Berry P, Petrillo J, *et al*. Development of the Shortness of Breath with Daily Activities questionnaire (SOBDA). *Value Health* 2012;15:1042–50.
4. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf (accessed 4 Apr 2013).
5. Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;23:932–46.
6. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999;159:179–87.

7. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd edn. New York: McGraw-Hill, Inc., 1994.
8. Revicki DA, Cella D, Hays RD, *et al*. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006;4:70.
9. Wyrwich K, Tierney W, Wolinsky F. Further evidence supporting a SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52:861–73.
10. Aaron SD, Vandemheen KL, Clinch JJ, *et al*. Measurement of short-term changes in dyspnoea and disease-specific quality of life following an acute COPD exacerbation. *Chest* 2002;121:688–96.
11. Aaron SD, Vandemheen KL, Hebert P, *et al*. Outpatient oral prednisone after emergency treatment of chronic obstructive pulmonary disease. *N Engl J Med* 2003;348:2618–25.

12. Parker CM, Voduc N, Aaron SD, *et al*. Physiological changes during symptom recovery from moderate exacerbations of COPD. *Eur Respir J* 2005;26:420–8.
13. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991;85(Suppl. B):2531.
14. Jones PW, Quirk FH, Baveystock CM, *et al*. A self-complete measure of health status for chronic airflow limitation. The St George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145:1321–7.
15. Hays R, Anderson R, Revicki DA. Assessing reliability and validity of measurement in clinical trials. In: Staquet MJ, Hays RD, Fayers PM, eds. *Quality of life assessment in clinical trials*. New York, NY: Oxford University Press, 1998: 167–82.
16. Schunemann HJ, Puhan M, Goldstein R, *et al*. Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). *COPD* 2005;2:81–9.