

Original Research

Baseline Severity as Predictor of Change in St George's Respiratory Questionnaire Scores in Trials of Long-acting Bronchodilators with COPD Patients

Paul W. Jones, PhD, FRCP, FERS^{1,2} Heather Gelhorn, PhD³ Niklas Karlsson, PhD⁴ Shailendra Menjoge, PhD⁵
Hana Müllerova, PhD² Stephen I. Rennard, MD^{6,7} Ruth Tal-Singer, PhD⁸ Hilary Wilson, PhD³
Debora Merrill, MBA⁹ Maggie Tabberer, MSc²

Abstract

Background: In trials of long-acting bronchodilators, health status is an important trial outcome, however the influence of baseline severity on response measured by St George's Respiratory Questionnaire (SGRQ) is not known. We have compared SGRQ changes between patients with chronic obstructive pulmonary disease (COPD) of mild-moderate severity or dyspnea (Global initiative for chronic Obstructive Lung disease [GOLD] grades 1 and 2; modified Medical Research Council [mMRC] grades 1 and 2) to those with severe-very severe severity or dyspnea (GOLD grades 3 and 4; mMRC grades 3 and 4).

Methods: Combined individual patient data from the COPD Biomarkers Qualification Consortium database (trials of long-acting bronchodilators) were used comprising of patients from short-term (≤ 1 -year duration; $n=10802$) and medium-term (2-4 years' duration; $n=8963$) studies. A repeated measures analysis of variance (ANOVA) was used to determine the effects of baseline severity (GOLD/mMRC) on SGRQ response to treatment. All treatment arms were combined.

Results: In short-term studies, milder patients showed a greater response than those with more severe disease in terms of GOLD grade (partial $\text{Eta}^2=0.03$, $p<0.0001$) and mMRC grade (partial $\text{Eta}^2=0.05$, $p<0.0001$). Similar results were seen in the medium-term studies (partial $\text{Eta}^2=0.02$, $p<0.0001$; mMRC: partial $\text{Eta}^2=0.05$, $p<0.0001$).

Conclusions: Patients with less severe airflow limitation and less severe dyspnea showed larger improvements in SGRQ score than more severely obstructed or dyspneic patients. Although these severity influences are small (2%-5% of the variance in SGRQ score), they do suggest that pre-specified separate analyses are warranted to test for differences in response, based on baseline severity.

Abbreviations: St George's Respiratory Questionnaire, **SGRQ**; chronic obstructive pulmonary disease, **COPD**; Global initiative for chronic Lung Disease, **GOLD**; modified Medical Research Council dyspnea scale, **mMRC**; analysis of variance, **ANOVA**; long-acting bronchodilators, **LABs**; health-related quality of life, **HRQL**; COPD Biomarkers Qualification Consortium, **CBQC**; long-acting anti-muscarinic, **LAMA**; long-acting beta2-agonist, **LABA**; randomized controlled trials, **RCTs**; standard deviation, **SD**; body mass index, **BMI**; forced expiratory volume in 1 second, **FEV₁**; cardiovascular disease, **CVD**

Funding Support: Funding for this COPD Biomarkers Qualification Consortium working group was provided by AstraZeneca, Boehringer-Ingelheim, GlaxoSmithKline, Novartis and Pfizer.

Date of Acceptance: February 1, 2017

Citation: Jones PW, Gelhorn H, Karlsson N, et al. Baseline severity as predictor of change in St George's Respiratory Questionnaire scores in trials of long-acting bronchodilators with COPD patients. *Chronic Obstr Pulm Dis.* 2017;4(2):132-140. doi: <https://doi.org/10.15326/jcopdf.4.2.2017.0129>

1 Institute of Infection and Immunity, St George's University of London, United Kingdom

2 Research and Development, GlaxoSmithKline, Uxbridge, United Kingdom

3 Evidera, Bethesda, Maryland

4 Research and Development, AstraZeneca, Gothenburg, Sweden

5 Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut

6 Division of Pulmonary, Critical Care, Sleep and Allergy, Nebraska Medical Center, Omaha

7 Research and Development, AstraZeneca, Cambridge, United Kingdom

8 Research and Development, GlaxoSmithKline, King of Prussia, Pennsylvania

9 COPD Foundation, Washington, D.C.

Address correspondence to:

Paul W. Jones, PhD, FRCP, FERS
Institute of Infection and Immunity
St George's University of London
London, United Kingdom
Email: pjones@sgul.ac.uk
Phone: +44 (0)20 8990 9000

Keywords:

St George's Respiratory Questionnaire; SGRQ; COPD Biomarkers Qualification Consortium; CBQC; predictors; forced expiratory volume in 1 second; FEV₁; modified Medical Research Council dyspnea scale; mMRC

Introduction

Long-acting bronchodilators (LABs) form the mainstay of treatment for chronic obstructive pulmonary disease (COPD).¹ They are appropriate for symptomatic patients of all degrees of severity, although there is little information about which patients benefit most in terms of health-related quality of life (HRQL). Unlike trials of treatments to reduce exacerbations, clinical trials for symptomatic treatments tend to recruit patients with a wide range of severity. To test whether there were differences in change in St George's Respiratory Questionnaire (SGRQ) scores that were dependent on baseline severity, we used the COPD Biomarkers Qualification Consortium (CBQC) database of clinical trials in COPD to compare response to SGRQ between patients with mild-moderate airflow limitation (Global initiative for chronic Obstructive Lung Disease [GOLD] grades 1 and 2) and those who were severe-very severe (GOLD grades 3 and 4).¹ In addition, we compared

the treatment response in patients with less severe breathlessness (modified Medical Research Council [mMRC] grades 1 and 2) compared to those with mMRC grades 3 and 4. Data from randomized trials of long-acting anti-muscarinics (LAMAs) and long-acting beta2-agonists (LABAs) were combined together with any placebo arms.

Methods

Identification of Clinical Trials

The CBQC database of clinical trials in COPD was used for this analysis.² This contains 18 randomized controlled trials (RCTs) and 3 observational studies which include SGRQ data. Only RCTs that included a treatment arm with placebo or long-acting bronchodilator (long-acting beta2-agonists and long-acting anti-muscarinic agents) were used in this analysis (N=17). Details of the RCTs included in the analysis are described elsewhere,³ and further details are contained in the online data supplement to that publication. All patients who had a baseline SGRQ score were included in these analyses. All analyses used fully anonymized, pooled data, so ethics and institutional review board approval were not required. Appropriate approvals and written informed consent had been obtained originally for all the included studies.

The 17 RCTs were categorized by study duration: short-term (≤ 1 -year duration) comprising 14 trials with 10802 participants (placebo=3670; bronchodilator=7132), and medium-term (2-4 years' duration) comprising 3 trials with 8963 participants (placebo=4184; bronchodilator=4779). The 2 types of trials were analyzed separately to test for consistency of findings across studies of different duration. Only 1 of the medium-term trials lasted for 4 years so the analysis was confined to the first 3 years of that trial. The patients were categorized into lower and higher airway function (GOLD grades 1 and 2 versus GOLD grades 3 and 4) and those with less or more breathlessness (mMRC grades 1 and 2 versus mMRC grades 3 and 4). Grouping of severity categories to provide 2 subgroups rather than 4 individual groups, was required to ensure a more even distribution of patients in each subgroup, since there were relatively few patients in the least and most severe groups, whether GOLD grade, or mMRC grade.

Statistical Analysis

For each trial dataset (i.e., short-term or medium-term),

demographic characteristics are presented as mean and standard deviation (SD). Comparisons between groups were performed using t-tests for continuous variables and chi-square tests for categorical variables.

To assess how the degree of limitation in airway function (GOLD grades 1 and 2 versus GOLD grades 3 and 4) and breathlessness (mMRC grades 1 and 2 versus mMRC grades 3 and 4) at baseline affects the SGRQ total score over the course of treatment with LABs, a repeated measures analysis of variance (ANOVA) was conducted. Separate models were run for GOLD (grades 1 and 2 versus grades 3 and 4) and mMRC (grades 1 and 2 versus grades 3 and 4). The dependent variable for these models was SGRQ total score and each model had 1 *between participants* factor (either GOLD or mMRC groups), and 1 *within participants* factor (time). For short-term trials, the study visits (time) included baseline and 1, 3, 6 and 12 months. For the medium-term trials, the study visits included baseline and 6, 12, 24, and 36 months. The main effect for time, GOLD or mMRC, and the interaction term were all evaluated; however, the primary hypothesis was centered on the interaction between GOLD/mMRC and time. Baseline covariates included baseline SGRQ score, age, sex, body mass index (BMI), World Health Organization socio-economic status of the patient's country (low, medium or high) and smoking status (former or current). In addition, baseline mMRC grade was included for models that tested the effect of GOLD grade on SGRQ response and baseline GOLD grade was included in models that tested the effect of mMRC grade. The following hypotheses were evaluated for the ANOVA models:

1. There will be a main effect for GOLD and mMRC, with more severe groups (GOLD grades 3 and 4 and mMRC grades 3 and 4) reporting higher SGRQ scores (HRQL) relative to the less severe groups (GOLD grades 1 and 2 and mMRC grades 1 and 2).

2. The response to treatment will be different (in either direction) between patients with more or less severe disease (as defined).

3. There will be a main effect for time, indicating that SGRQ scores change over the course of treatment with LABs in both the short-term (up to 12 months) and medium-term (up to 36 months). Given that COPD is a progressive condition, treatment with LABs often results in an initial improvement in symptoms, and subsequent deterioration,^{4,5} a non-linear effect of time was modeled.

4. There will be no significant interaction between

GOLD or mMRC with time, indicating that the rate of change in SGRQ with LAB treatment does not vary based on the degree of limitation in airway function (GOLD grades 1 and 2 versus GOLD grades 3 and 4) and breathlessness (mMRC grades 1 and 2 versus mMRC grades 3 and 4).

The size of the databases may have resulted in some of the analyses being overpowered, resulting in small differences attaining statistical significance. In view of the number of study participants and the possible number of comparisons, significance was accepted at $p < 0.01$. Attention should be paid to the size of the effect, rather than the p -value. For this reason, the partial-Eta squared effect size is reported for all effects, and the following rules of thumb were utilized to facilitate interpretation: small (0.01), medium (0.06), and large (0.14).^{6,7}

Results

Baseline Characteristics

Short-term Trials - Demographics

By the nature of the categorization at baseline, mean forced expiratory volume in 1 second (FEV₁) in GOLD grades 1 and 2 and mMRC grades 1 and 2 patients was higher (62.3% and 51.1% predicted) than in patients in GOLD grades 3 and 4 and mMRC grades 3 and 4 (37.6% and 44.2% predicted respectively), Table 1. Similarly, SGRQ scores were worse (higher) in patients in GOLD grades 3 and 4, and those that reported greater breathlessness on the mMRC scale (GOLD grades 3 and 4=49.9 versus GOLD grades 1 and 2=41.1; mMRC grades 3 and 4=61.1 versus mMRC grades 1 and 2=45.8; $p < 0.0001$), Table 1. Patients in the more severe groups (GOLD grades 3 and 4 and mMRC grades 3 and 4) were more likely to have experienced an exacerbation in the 12 months leading up to the baseline study visit relative to the less severe groups (GOLD grades 1 and 2 and mMRC grades 1 and 2) ($p < 0.0001$).

The FEV₁ percent predicted was lower in GOLD grades 3 and 4 (37.6%) compared with mMRC grades 3 and 4 (44.2 %). There were proportionally more men in GOLD grades 3 and 4 (73.8%) than in mMRC grades 3 and 4 (62.4%). Otherwise, the 2 methods of assessing severity identified patients of a similar age and BMI and broadly similar proportions of patients in terms of: socio-economic status, smoking status, history of exacerbations and hospitalization (Table 1).

Table 1. Baseline Demographics of Participants in Short-Term Trials

	Overall Sample ^a (N=10802)	GOLD 1 & 2 (N=4638)	GOLD 3 & 4 (N=5845)	<i>p</i> -value ^b	mMRC 1 & 2 (N=4768)	mMRC 3 & 4 (N=1035)	<i>p</i> -value ^c
Age , mean (SD)	63.9 (8.9)	63.9 (9.1)	63.9 (8.6)	0.9292	63.4 (8.8)	63.6 (9.0)	0.6155
Sex , n (%) male	7645 (70.8%)	3126 (67.4%)	4311 (73.8%)	<.0001	3363 (70.5%)	646 (62.4%)	<.0001
Baseline FEV₁ % Predicted , mean (SD)	48.8 (14.9)	62.3 (8.7)	37.6 (8.2)	<.0001	51.1 (14.6)	44.2 (15.3)	<.0001
Baseline SGRQ Score , mean (SD)	46.1 (18.2)	41.1 (17.8)	49.9 (17.6)	<.0001	45.8 (16.7)	61.1 (16.2)	<.0001
Smoking Status , n (%)							
Current	4340 (40.2%)	1986 (42.8%)	2214 (37.9%)	<.0001 ^d	2089 (43.8%)	453 (43.8%)	0.9790 ^d
Past	6456 (59.8%)	2651 (57.2%)	3627 (62.1%)		2679 (56.2%)	582 (56.2%)	
Income							
High Income	7541 (69.8%)	3313 (71.4%)	3977 (68.0%)	0.0002 ^d	2999 (62.9%)	741 (71.6%)	<.00010 ^d
Low/Middle Income	3261 (30.2%)	1325 (28.6%)	1868 (32.0%)		1769 (37.1%)	294 (28.4%)	
Body Mass Index , mean (SD)	26.3 (5.3)	26.8 (5.2)	25.7 (5.3)	<.0001	26.3 (5.2)	27.2 (6.7)	<.0001
Exacerbation in 12 Months Prior , n (%)							
0	5441 (76.3%)	3217 (88.9%)	2186 (64.6%)	<.0001 ^d	3070 (71.8%)	548 (59.4%)	<.0001 ^d
1	1127 (15.8%)	293 (8.1%)	778 (23.0%)		797 (18.7%)	219 (23.8%)	
2+	561 (7.9%)	110 (3.0%)	420 (12.4%)		406 (9.5%)	155 (16.8%)	
Hospitalized Exacerbation in 12 Months Prior , n (%)							
0	5959 (92.6%)	3408 (94.6%)	2470 (90.2%)	<.0001 ^d	3478 (92.5%)	625 (88.9%)	0.0025 ^d
1	460 (7.1%)	188 (5.2%)	257 (9.4%)		272 (7.2%)	72 (10.2%)	
2+	18 (0.3%)	5 (0.1%)	12 (0.4%)		12 (0.3%)	6 (0.9%)	

^aNote that GOLD and mMRC were not available on all participants, so the analysis population for GOLD and mMRC was limited to only participants who had these variables

^bComparison [T-test (continuous) or chi-square (categorical)] of GOLD grades 1&2 vs. GOLD grades 3&4

^cComparison [T-test (continuous) or chi-square (categorical)] of mMRC grades 1&2 vs. mMRC grades 3&4

^dGlobal *p*-value

Note: 1. T-test for BMI was calculated with unequal variances assumed.

2. Not every participant had information on smoking status, BMI, mMRC Score, exacerbations, or CVD history.

GOLD=Global initiative for chronic Lung Disease; mMRC=modified Medical Research Council dyspnea scale; SD=standard deviation; FEV₁=forced expiratory volume in 1 second; SGRQ=St George's Respiratory Questionnaire; BMI=body mass index; CVD=cardiovascular disease

Medium-term Trials - Demographics

The pattern of baseline findings in the medium-term trials was similar to those seen with the short-term trials, i.e., the differences in FEV₁ and SGRQ between the GOLD and mMRC subgroups reflected patient categorization at baseline (Table 2). The more severe groups (GOLD grades 3 and 4 and mMRC grades 3 and 4) were more likely to have experienced an exacerbation in the 12 months leading up to the baseline visit compared with the less severe groups. The 2 methods of severity assessment identified patients with similar characteristics, with the exception that SGRQ scores were worse in patients with mMRC grades 3 and 4 than in GOLD grades 3 and 4 (60.0 versus 50.3), Table 2. The FEV₁ percent predicted was almost identical in

patients in GOLD grades 3 and 4 and mMRC grades 3 and 4 (36.9% versus 36.8%), Table 2.

SGRQ Response in GOLD Grades 1 and 2 versus GOLD Grades 3 and 4 Groups

In the short-term dataset, the repeated measures ANOVA indicated that the main effect for airflow limitation was significant ($F_{1, 30677}=1042.4$, $p<0.0001$, partial $\eta^2=0.03$), with patients with greater airflow limitation reporting worse scores for SGRQ. Post hoc tests at each time point show that the improvement in SGRQ score from baseline in the GOLD grades 3 and 4 group was significantly less than in patients in GOLD grades 1 and 2 (Figure 1). The average difference was approximately 3 units. Both the linear ($F_{1, 30677}=401.2$,

Table 2. Baseline Demographics in Medium-term Trials

	Overall Sample ^a (N=8963)	GOLD 1 & 2 (N=2669)	GOLD 3 & 4 (N=5100)	<i>p</i> -value ^b	mMRC 1 & 2 (N=2197)	mMRC 3 & 4 (N=713)	<i>p</i> -value ^c
Age , mean (SD)	64.7 (8.4)	64.7 (8.5)	64.7 (8.2)	0.9955	64.9 (8.2)	65.6 (8.3)	0.0597
Sex , n (%) male	6773 (75.6%)	1962 (73.5%)	3977 (78.0%)	<.0001	1685 (76.7%)	540 (75.7%)	0.5999
Baseline FEV₁ % Predicted , mean (SD)	45.8 (12.7)	57.6 (5.5)	36.9 (8.1)	<.0001	43.9 (12.1)	36.8 (11.8)	<.0001
Baseline SGRQ Score , mean (SD)	47.4 (17.2)	43.6 (17.0)	50.3 (16.5)	<.0001	47.1 (16.1)	60.0 (14.4)	<.0001
Smoking Status , n (%)							
Current	3130 (34.9%)	996 (37.3%)	1728 (33.9%)	0.0026 ^d	989 (45.0%)	271 (38.0%)	0.0010 ^d
Past	5833 (65.1%)	1673 (62.7%)	3372 (66.1%)		1208 (55.0%)	442 (62.0%)	
Income							
High Income	6115 (68.2%)	1807 (67.7%)	3503 (68.7%)	0.3763 ^d	1589 (72.3%)	461 (64.7%)	<.0001 ^d
Low/Middle Income	2848 (31.8%)	862 (32.3%)	1597 (31.3%)		608 (27.7%)	252 (35.3%)	
Body Mass Index , mean (SD)	25.8 (5.1)	26.1 (4.8)	25.1 (5.0)	<.0001	25.5 (4.8)	25.7 (5.9)	0.4970
Exacerbation in 12 Months Prior , n (%)							
0	1190 (38.4%)	446 (48.8%)	742 (34.4%)	<.0001 ^d	887 (40.4%)	211 (29.6%)	<.0001 ^d
1	751 (24.2%)	227 (24.8%)	522 (24.2%)		563 (25.6%)	142 (19.9%)	
2+	1159 (37.4%)	241 (26.4%)	891 (41.3%)		747 (34.0%)	360 (50.5%)	
Hospitalized Exacerbation in 12 Months Prior , n (%)							
0	7200 (82.2%)	2262 (86.3%)	3944 (79.3%)	<.0001 ^d	1778 (82.8%)	514 (74.9%)	<.0001 ^d
1	1152 (13.1%)	280 (10.7%)	746 (15.0%)		303 (14.1%)	120 (17.5%)	
2+	411 (4.7%)	80 (3.1%)	286 (5.7%)		67 (3.1%)	52 (7.6%)	

^aNote that GOLD and mMRC were not available on all participants, so the analysis population for GOLD and mMRC was limited to only participants who had these variables

^bComparison [T-test (continuous) or chi-square (categorical)] of GOLD grades 1&2 vs. GOLD grades 3&4

^cComparison [T-test (continuous) or chi-square (categorical)] of mMRC grades 1&2 vs. mMRC grades 3&4

^dGlobal *p*-value

Note: 1. T-test for BMI was calculated with unequal variances assumed.

2. Not every participant had information on smoking status, BMI, mMRC Score, exacerbations, or CVD history.

GOLD=Global initiative for chronic Lung Disease; mMRC=modified Medical Research Council dyspnea scale; SD=standard deviation; FEV₁=forced expiratory volume in 1 second; SGRQ=St George's Respiratory Questionnaire; BMI=body mass index; CVD=cardiovascular disease

$p < 0.0001$, partial $\eta^2 = 0.01$) and quadratic effects for time ($F_{1, 30677} = 296.0$, $p < 0.0001$, partial $\eta^2 = 0.01$) were also significant, but both accounted for a very small proportion of the model variance. While the plots in Figure 1 suggest that the SGRQ differences appeared to widen over time, and the interaction between GOLD status and time was statistically significant, the effect size was negligible ($F_{1, 30677} = 4.56$, $p = 0.03$, partial $\eta^2 = 0.0001$).

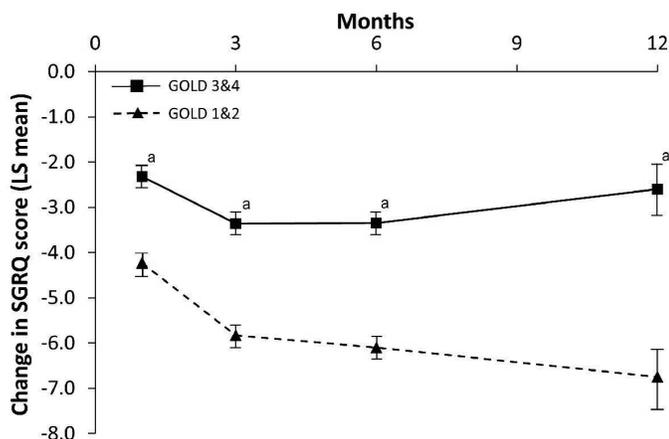
The pattern seen in the medium-term trials was similar: the GOLD main effect ($F_{1, 27872} = 453.1$, $p < 0.0001$, partial $\eta^2 = 0.02$), linear effect of time ($F_{1, 27872} = 142.9$, $p < 0.0001$, partial $\eta^2 = 0.005$), quadratic effect of time ($F_{1, 27872} = 114.2$, $p < 0.0001$, partial $\eta^2 = 0.004$), and GOLD x time ($F_{1, 27872} = 1.50$, $p = 0.22$ partial $\eta^2 = 0.0001$).

Again, at 3 out of 4 time points, the response in GOLD grades 3 and 4 was significantly smaller than in GOLD grades 1 and 2 (Figure 2). Importantly, by 24 months, patients in GOLD grades 3 and 4 had returned to base line, whereas those in GOLD grades 1 and 2 remained significantly better than at baseline, even at 36 months ($p = 0.0018$).

SGRQ Response Between mMRC Grades 1 and 2 versus mMRC Grades 3 and 4 Groups

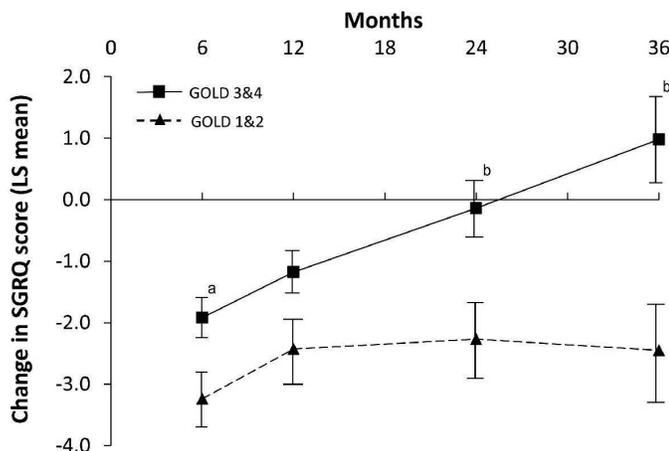
In the short-term dataset, the repeated measures ANOVA indicated that the main effect for mMRC was significant ($F_{1, 20606} = 1205.0$, $p < 0.0001$, partial $\eta^2 = 0.05$). The post hoc analysis of change from baseline showed that at 1, 3 and 6 months, patients with mMRC grades 1 and 2 had

Figure 1. Least Square Means for Change in SGRQ Score from Baseline in Short-Term Trials: Gold Grades 1&2 and 3&4



Data are presented as means and standard error bars. Significance at each time point was tested using post hoc tests ^a $p < 0.001$. Samples sizes for each data point ranged from 1273 (GOLD 1&2 at 12 months) to 5845 (GOLD 3&4 at Baseline).

Figure 2. Least Square Means for Change in SGRQ Score From Baseline in Medium-Term Trials: Gold Grades 1&2 and 3&4

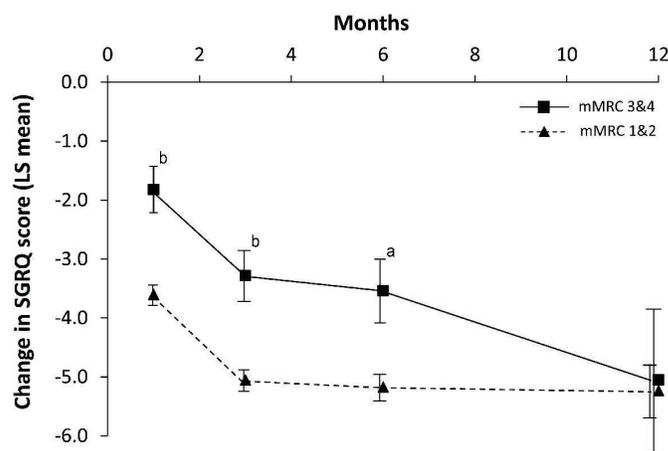


Data are presented as means and standard error bars. Significance at each time point was tested using post hoc tests ^a $p < 0.05$; ^b $p < 0.01$. The sample size for each data point ranged from 1703 (GOLD 1&2 at 36 months) to 5100 (GOLD 3&4 at baseline).

a significantly larger improvement in SGRQ than those with mMRC grades 3 and 4 (Figure 3). At 12 months, the scores converged, but it will be noted that the standard errors for mMRC grades 3 and 4 widened considerably from those at 6 months. Both the linear ($F_{1, 20606}=296.9, p < 0.0001$, partial $\eta^2=0.01$) and quadratic effects for time ($F_{1, 20606}=247.5, p < 0.0001$, partial $\eta^2=0.01$) were significant but with small effect sizes. The interaction between mMRC and time was not significant ($F_{1, 20606}=0.83, p=0.36$, partial $\eta^2=0.00004$), indicating that the rate of change in SGRQ over time did not vary by baseline mMRC status.

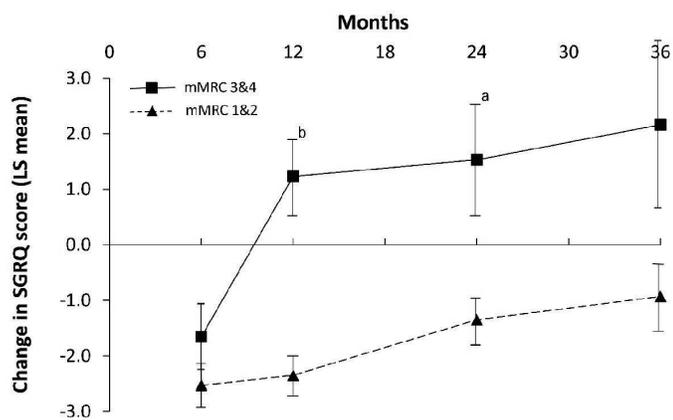
The ANOVA results in the medium-term dataset showed a similar difference: $F_{1, 8939}=437.6, p < 0.0001$, partial $\eta^2=0.05$). Figure 4 shows that there was quite a large and consistent difference between the 2 groups of patients and patients with mMRC grades 1 and 2 remained significantly better than at baseline at 24 months ($p=0.0017$), but not at 36 months ($p=0.06$). In contrast, patients with mMRC grades 3 and 4 were numerically, but not significantly, worse than baseline from 12 months onwards. The linear effect for time was significant ($F_{1, 8939}=33.4, p < 0.0001$, partial $\eta^2=0.004$), as was the quadratic effect ($F_{1, 8939}=29.6, p < 0.0001$,

Figure 3. Least Square Means for Change in SGRQ Score from Baseline in Short-Term Trials: mMRC Grades 1&2 and 3&4



Data are presented as means and standard error bars. Significance at each time point was tested using post hoc tests ^a $p < 0.01$; ^b $p < 0.001$. The sample size for each data point ranged from 395 (mMRC 3&4 at 12 months) to 4768 (mMRC 1&2 at baseline).

Figure 4. Least Square Means for Change in SGRQ Score From Baseline in Medium-Term Trials: mMRC Grade 1&2 and 3&4



Data are presented as means and standard error bars. Significance at each time point was tested using post hoc tests ^a $p < 0.01$; ^b $p < 0.001$. The sample size for each data point ranged from 195 (mMRC 3&4 at 36 months) to 2197 (mMRC 1&2 at baseline)

partial $\eta^2 = 0.003$), but the term for mMRC \times time was not ($F_{1, 8939} = 1.00$, $p = 0.32$, partial $\eta^2 = 0.0001$, showing that there was no significant difference between groups in terms of change over time.

Discussion

This analysis shows that in both short- and medium-term trials, patients in GOLD grades 1 and 2 showed a greater improvement in SGRQ score than those in GOLD grades 3 and 4. These differences were seen at nearly all time points. In short-term trials, the improvement in SGRQ score seen in GOLD grades 1 and 2 was approximately double that observed in GOLD grades 3 and 4. Visually the responses appear to widen over time (Figure 1), although the effect size was negligible, despite the relatively large sample size. In the medium-term trials, benefit was maintained over 3 years in GOLD grades 1 and 2 patients, but was lost by 24 months in GOLD grades 3 and 4 (Figure 2).

Using mMRC score as a measure of baseline severity, the findings were similar. In the short-term trials, the improvements were smaller in patients with worse dyspnea (i.e., mMRC grades 3 and 4). In short-term trials, there appears to be convergence at 12 months, but the

large 95% confidence intervals in mMRC grades 3 and 4 patients, due to smaller numbers of trial participants contributing data at this time point, lend uncertainty to this observation. In medium-term trials, the patients with less dyspnea (mMRC grades 1 and 2) lose some benefit over time, but even after 36 months remain a little better when compared to baseline, whereas in mMRC grades 3 and 4 groups, benefit was lost at 12 months and for the subsequent 2 years. It is not clear why there is a difference at 12 months between the 2 groups of trials, but it may be due to inclusion criteria and participant demographics. The exacerbation rate and hospitalization rate in medium-term trials were both higher than in the short-term trials (Tables 1 and 2) and there is an association between exacerbation rate and worsening of SGRQ score.⁸

It should be appreciated that this analysis did not compare bronchodilator treatment with placebo, but it looked at the SGRQ changes during the trial regardless of treatment arm, so some patients will have received active treatment, while others would have had placebo. As shown in a companion paper,⁹ in the context of a clinical trial, receiving placebo appears to be an active, albeit less effective, treatment that results in an improvement in SGRQ score that is progressive at onset and is sustained for many months. In that paper, we showed that while there was a clear difference in size of improvement in score between patients in low-medium versus high socio-economic countries, whether receiving active treatment or placebo, the difference between active treatment and placebo arms was minimally affected by socio-economic status. In this analysis, while there was a significant difference between patients who were more or less severe at baseline, the proportion of variance in SGRQ score explained by differences in baseline GOLD or mMRC grade was small and most of the partial η^2 values would be judged to be small.^{6,7}

The same limitations apply to these analyses as they do to most other analyses from the CBQC database. While the database is made up of individual patient data from a large number of patients and trials, the comparisons are post hoc and largely indirect. The trials had different inclusion criteria and will, because of issues of sampling from a population, have included patients with different baseline characteristics. That having been said, the findings are quite consistent and show that, at a study population level, patients with more severe airflow limitation or more severe breathlessness, show a smaller SGRQ response to long-acting bronchodilators

than less severe patients.

Taken overall, the conclusions from this analysis have implications for clinical trial design and analysis and suggest that pre-specified subgroup analysis by baseline severity should be undertaken. These analyses may also have implications for clinical practice, since they suggest that patients with less severe disease may respond better, and for longer, to a therapeutic intervention whether active treatment or just observation and monitoring. This should, perhaps, encourage doctors to treat milder patients more aggressively, since these analyses suggest that such patients may experience the greatest health status gain.

Acknowledgments

The authors would like to thank Debbie Merrill, COPD Foundation, for managing the review process, the

COPD Biomarkers Qualification Consortium for their role in aggregating the data, Thomas Martin of Novartis and Katja Rüdell, formerly of Pfizer for their review and oversight through the CBQC Steering Committee. They also acknowledge the assistance provided by Kate Hollingworth of Continuous Improvement Ltd in copyediting and formatting the manuscript; this was funded by the COPD Foundation.

Declaration of Interest

PWJ, NK, SM, HM, SIR, RTS and MT are employees and shareholders of the pharmaceutical companies who funded this analysis. HG and HW participated in this project as employees of Evidera, a company which performs work for hire for multiple pharmaceutical and device companies in outcomes research. DM has nothing to declare.

References

1. Global Initiative for Chronic Lung Disease (GOLD) Global Strategy for the Diagnosis, Management and Prevention of COPD. GOLD website. <http://www.goldcopd.org> Published 2016. Accessed January 2017.
2. Casaburi R, Celli B, Crapo J, et al. The COPD Biomarker Qualification Consortium (CBQC). *COPD*. 2013;10(3):367-377. doi: <https://doi.org/10.3109/15412555.2012.752807>
3. Tabberer M, Benson VS, Gelhorn H, et al for the COPD Biomarkers Qualification Consortium. The COPD Biomarkers Qualification Consortium database: Baseline characteristics of the St George's Respiratory Questionnaire dataset. *Chronic Obstr Pulm Dis*. 2017;4(2): 112-123. doi: <https://doi.org/10.15326/jcopdf.4.2.2017.0129>
4. Calverley PM, Anderson JA, Celli B, et al for the TORCH investigators. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med*. 2007; 356:775-789. doi: <https://doi.org/10.1056/NEJMoa063070>
5. Tashkin DP, Celli B, Senn S, et al, for the UPLIFT Study Investigators: A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med*. 2008; 359:1543-1544. doi: <https://doi.org/10.1056/NEJMoa0805800>
6. Miles J, Shevlin M. *Applying Regression and Correlation: A Guide for Students and Researchers*. London: Sage; 2001.
7. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1998.
8. Spencer S, Calverley MA, Burge S, Jones W. Impact of preventing exacerbations on deterioration of health status in COPD. *Eur Respir J*. 2004;23(5):698-702. doi: <https://doi.org/10.1183/09031936.04.00121404>
9. Jones PW, Gelhorn H, Wilson H, et al for the COPD Biomarkers Qualification Consortium. Socio-economic status as a determinant of health status treatment response in COPD trials. *Chronic Obstr Pulm Dis*. 2017;4(2): 150-158. doi: <https://doi.org/10.15326/jcopdf.4.2.2017.0132>