Contemporary Clinical Trials 32 (2011) 818–828

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Application of latent growth and growth mixture modeling to identify and characterize differential responders to treatment for COPD

Donald E. Stull [a,*], Ingela Wiklund [b], Rupert Gale [c], Gorana Capkun-Niggli [d], Katherine Houghton [a], Paul Jones [e]

[a] RTI Health Solutions, The Pavilion, Towers Business Park, Wilmslow Road, Didsbury, Manchester, M20 2LS, UK
[b] United BioSource Corporation, 26-28 Hammersmith Grove, London, W6 7HA, UK
[c] Novartis Horsham Research Center, Wimblehurst Road, Horsham, West Sussex RH12 5AB, UK
[d] Novartis Pharma AG, Square 3, Postfach CH-4002 Basel, Switzerland
[e] St. George's, University of London, Cranmer Terrace, SW170RE London, UK

ARTICLE INFO

Article history:
Received 2 March 2011
Received in revised form 26 May 2011
Accepted 28 June 2011
Available online 6 July 2011

Keywords:
COPD
Longitudinal analysis
Growth mixture model
Latent growth model
Clinical trials

ABSTRACT

Objective: To explore the utility of applying growth mixture models (GMMs) in secondary analyses of clinical trials to identify sources of variability in data reported by patients with COPD.

Methods: Analyses were performed on data from two 6-month clinical trials comparing indacaterol and open-label tiotropium or blinded salmeterol and the first six months of a 12-month trial comparing indacaterol and blinded formoterol. Latent growth model (LGM) analyses were conducted to explore the response of the SGRQ Symptoms score from baseline to six months and GMM analyses were evaluated as a method to identify latent classes of differential responders.

Results: Variability in SGRQ Symptom scores was found suggesting subsets of patients with differential response to treatment. GMM analyses found subsets of non-responders in all trials. When the responders were analyzed separately from non-responders, there were increased treatment effects (e.g., symptoms score improvement over six months for whole groups: indacaterol = 8–12 units, tiotropium = 7 units, salmeterol = 9 units, formoterol = 11 units. Responder subgroup improvement: indacaterol = 9–21 units, tiotropium = 7 units, salmeterol = 10 units, formoterol = 20 units). Responders had significantly different baseline SGRQ Symptom scores, smoking history, age, and mMRC dyspnea scores than non-responders.

Conclusions: Patients with COPD represent a heterogeneous population in terms of their reporting of symptoms and response to treatment. GMM analyses are able to identify subgroups of responders and non-responders. Application of this methodology could be of value on other endpoints in COPD and in other disease areas.

1. Introduction

It is very common to observe heterogeneity in treatment response in clinical trials and interest in identifying and categorizing this has grown in recent years [1–4]. A challenge facing researchers is how best to examine whether there are patterns to this variability and the sources of the variability. For example, subpopulations may exist within a larger heterogeneous population that show a differential growth trajectory which is masked when group means as a whole are considered. In terms of treatment, identifying and categorizing differential responders have implications for trial development and clinical management, with the potential for the information to facilitate the targeting of patients most likely

* Corresponding author. Tel.: +44 161 447 6031; fax: +44 161 434 8232.
E-mail addresses: dstull@rti.org (D.E. Stull), ingela.wiklund@unitedbiosource.com (I. Wiklund), rupert.gale@novartis.com (R. Gale), gorana.capkun-niggli@novartis.com (G. Capkun-Niggli), khoughton@rti.org (K. Houghton), pjones@sgul.ac.uk (P. Jones).

1551-7144 © 2011 Elsevier Inc. Open access under CC BY-NC-ND license.
doi:10.1016/j.cct.2011.06.004

to benefit. Previous studies have shown that latent growth modeling (LGM) and growth mixture modeling (GMM) may yield more precise parameter estimates than traditional analysis methods and explain more of the variance in an outcome [5,6]. These methods have been implemented using clinical trial results in the treatment of Alzheimer's disease [2] and depression [3,4], with GMMs successfully identifying subgroups of responders with differential growth trajectories. Therefore, analyzing group means as a whole may mask important differences in treatment response between unobserved subgroups. Although studies have established that there is variability in treatment response for chronic obstructive pulmonary disease (COPD) patients [7,8], very little is known about such differential response and which patients are most likely to benefit most/least [9].

COPD is a progressive, life-threatening, chronic lung disorder, that is characterized by airflow limitation that is poorly reversible, airway and systemic inflammation, structural changes (airway remodeling, emphysema), and mucociliary dysfunction [10]. The World Health Organization (WHO) has estimated that about 210 million people have COPD and predicts that COPD will become the third leading cause of death by 2030 [11].

The therapeutic goals for patients with COPD include relief of symptoms, improving health status, preventing and treating exacerbations, slowing the progression of disease and reducing mortality [12,13]. Viewed from the patient's perspective, symptoms affecting health status, activities of daily living, survival, and exacerbations are relevant outcomes [14–16]. Measurement of pulmonary function alone may not adequately reflect the burden of COPD on patients and the effectiveness of therapeutic interventions [17]. Self-report of symptoms and health status, therefore, is essential for evaluating treatment outcome [18–22].

Current guidelines for the management of COPD recommend the use of long-acting inhaled bronchodilators for patients with symptomatic moderate and severe disease with the addition of inhaled corticosteroids (ICS) for patients who experience repeated exacerbations [4]. Indacaterol is a novel, once-daily inhaled long-acting β_2-agonist (LABA) for the treatment of COPD and studies [23–28] have reported that it increases lung function alongside improving patient-reported symptoms and health status. The clinical trials in the development program for indacaterol provided data that allowed an opportunity to carry out an exploratory analysis to ascertain if new techniques could be applied to identify unobserved subgroups of patients exhibiting differential response to outcomes drawn directly from patient experience — self-reported symptoms. The randomized controlled studies included twice-daily LABAs (salmeterol and formoterol), and the once-daily anticholinergic tiotropium, which were evaluated using a range of lung-function and clinical outcome endpoints including patient-reported COPD symptoms [29–31].

The present analysis applied LGM and GMM methodology to examine variability in symptom response data reported by patients with COPD to identify subgroups of patients whose self-reported symptoms exhibited greater or lesser change. Patients showing different symptom responses were characterized into different responder subgroups to understand better how they differed on baseline characteristics.

2. Methods

2.1. Datasets analyzed

Two 6-month trials and one 12-month trial were used for the present analyses. All trials were multi-national, multi-center, double blind, double dummy, placebo-controlled trials with an active comparator. The INHANCE study compared indacaterol 150 μg and 300 μg with placebo and open-label tiotropium 18 μg for six months, randomizing patients at 1:1:1:1 [31]. The INLIGHT-2 study compared indacaterol 150 μg with placebo and blinded salmeterol 50 μg for six months, randomizing patients at 1:1:1 [29]. The third study, INVOLVE, was a 12-month trial that compared indacaterol 300 μg and 600 μg with placebo and blinded formoterol 12 μg, randomizing patients at 1:1:1:1 [30]. Only data from the first six months of INVOLVE were used to keep the analyses comparable across the three trials.

Since the objective of this study was the application of novel analytic methods, all treatment arms (including placebo and the unlicensed 600 μg dose of indacaterol) were included. Details of the trials are reported in the publications referenced above but briefly all trials had the following inclusion criteria: male and female patients aged ≥40 years; diagnosis of moderate-to-severe COPD as classified by GOLD guidelines (2005); smoking history of at least 20 pack-years; post-bronchodilator forced expiratory volume in 1 s (FEV_1) <80% and ≥30% of the predicted normal value; and post-bronchodilator FEV_1/FVC (forced vital capacity) <70%. All patients were supplied with the short-acting β_2-agonist salbutamol/albuterol for use as rescue medication.

2.2. Assessments

Measurement of lung function (FEV_1) was the primary endpoint in the trials but symptoms based measures were also included: The St George's Respiratory Questionnaire (SGRQ) was used to measure COPD health status. In addition, breathlessness was captured with the modified Medical Research Council dyspnea scale (mMRC) and the Baseline and Transition Dyspnea Indices (BDI/TDI). In the present analyses only data from the SGRQ and mMRC were used for all trials, from baseline, week 12 and six months. These measures are described below.

2.2.1. Modified Medical Research Council dyspnea scale (mMRC)

The clinician rated the degree of the participant's dyspnea on the mMRC dyspnea scale [32]. The mMRC is a five point scale based on degrees of physical activities that may lead to dyspnea, ranging from 0 (“no breathlessness except with strenuous exercise”) to 4 (“too breathless to leave the house or breathless when dressing or undressing”).

2.2.2. St George's Respiratory Questionnaire (SGRQ)

The SGRQ is a validated patient-reported outcome (PRO) measure of health status in diseases of chronic airflow limitation and has been widely used in clinical trials in COPD [33,34]. It contains 50 items divided into three subscales: “Symptoms”, “Activity”, and “Impacts”. A score for each subscale was calculated, alongside “Total” score. In

each case the lowest possible value was zero and the highest 100. Higher values correspond to greater impairment in health status. The Symptoms domain was selected for this exploratory analysis given the importance of symptoms to patients, and because active treatments tend to show the greatest level of improvement in this domain [28,35,36].

3. Statistical analysis

Overview: Descriptive statistics are presented for the INHANCE, INLIGHT-2 and INVOLVE studies separately. The effect of treatment over time and effect sizes with regards to the clinical assessments and patient-reported symptoms using the SGRQ Symptoms domain scores are presented separately for each study. In addition, two types of longitudinal analyses were conducted based on the data at baseline, week 12, and six months for each study: latent growth modeling and growth mixture modeling. In both sets of analyses, the models included the following covariates to be as similar as possible to the original clinical trial analyses: COPD severity, smoking status, baseline FEV_1, age, and gender.

Latent growth models (LGM) were used to explore the response of the SGRQ Symptom domain score from baseline to six months across the three assessment points (baseline, 12 weeks, and six months) controlling for key covariates. Latent growth models use structural equation modeling techniques to model trajectories of change, assess effects of treatment, and consider the relationships among multiple outcome variables and multiple time points simultaneously. In LGM analysis, changes in scores are analyzed at the individual level, by generating an intercept and slope of change for each individual, thus modeling individual variability in treatment response. Intra- and inter-individual changes are assessed using all available data points and modeling measurement error, which is especially important in analysis of patient-reported outcomes data. The consequence is more precise parameter estimates while using data from all available time points [5]. More detailed descriptions of these models are presented elsewhere [5], but a brief description is provided here.

An intercept and slope of change in the SGRQ Symptom score is based on the manifest composite score for this domain at each assessment point. Thus, each growth curve is characterized by two latent variables: a variable for the first time point of the curve, labeled as "intercept", and a variable for changes in the scores over time, labeled "slope". Note that the intercept variable is not equivalent to the value of the initial observation for a patient (i.e., baseline score), but rather the value of the growth curve at the start of the trial for each patient, adjusted for covariates. Thus, the intercept and slope variables have means and variances reflecting the mean intercepts and mean slopes of change and the variability in individual intercepts and slopes [5].

Typically in an LGM, each intercept variable is allowed to be correlated with the corresponding latent slope variable. That is, an underlying hypothesis that is tested is that a change in the slope variable is related to the value of the baseline variable. Thus, this parameter is estimated to see if those who start with a higher (worse) SGRQ Symptom score show greater improvements. Dyspnea, as measured by the mMRC, was modeled as a time-varying covariate to contribute to controlling the analysis for the level of breathlessness at each assessment, which could in turn influence SGRQ Symptom score. For each treatment group, a mean intercept and slope is calculated (along with a variance around that mean) which allows the analyst to compare intercepts but, more importantly, to compare differences in slope of change in the SGRQ Symptom scores by treatment groups over the trial.

Growth mixture models (GMM) analyses (longitudinal factor mixture models) were used to see if latent classes (unknown subgroups) existed in the trial data, allowing identification of differential responders in the studies [6]. Mixture modeling refers to modeling with categorical latent variables that represent subpopulations where class membership is not known but is inferred from the data [6,37,38]. Mixture models are designed for data from heterogeneous samples likely to consist of multiple latent classes, each of which is relatively homogeneous.

Mixture modeling was used to assign subjects to their most likely class and to obtain estimates of the model parameters for each class. Growth mixture models were conducted using Mplus version 6.0. In a preliminary run, the software determined a number of latent classes (k) within each treatment arm. A second run was performed with different start values to reduce the chance of local maxima and to determine whether the same number of classes would be extracted, and if the class assignment was consistent. Following this, analyses were performed with $k-1$ and/or $k+1$ classes to evaluate which number of classes extracted yielded the best model fit. The Bayesian Information Criteria (BIC) and Sample-size Adjusted BIC (SABIC) were the primary fit statistics used to determine the number of classes that best fit the data, along with entropy, which gives an indication of the accuracy of latent class assignment for each respondent [39,40], and the size of the smallest latent class that was extracted. Very small classes may represent chance findings and thus give a false indication of the number of latent classes within the heterogeneous data. In case a model did not converge (i.e. the model was not able to run) for a specified number of latent classes, the remaining models were compared and the one with the best combination of BIC, SABIC and entropy value was chosen. Smaller values of BIC and SABIC are preferred when choosing the number of latent classes [40]. Although there is no conventional level for the threshold value for entropy, values closer to 1 indicate greater accuracy of latent class assignment. Multiple random starts were used to avoid solutions based on local maxima, which would give a false sense of the number of latent classes in the data. In addition, visual inspection of the latent class trajectories provided insight into the reasonableness of the numbers of latent classes to be considered. Once classes were extracted, class ordering and separation were explored as a final check on the appropriateness of the number of latent classes extracted. Review of the information from fit statistics, smallest class size, and visual inspection of the slope trajectories by latent class was used to make final decisions about the numbers of latent classes to be extracted. The latent class assignment for each patient was then merged with the original trial data to explore baseline characteristics of each class in *post hoc* analyses.

Table 1
Baseline demographic and clinical characteristics by study and treatment arm.

	INHANCE				INLIGHT-2			INVOLVE			
	IND150 (N = 416)	IND300 (N = 416)	TIOT (N = 415)	PLAC (N = 418)	IND150 (N = 330)	SALM50 (N = 333)	PLAC (N = 335)	IND300 (N = 437)	IND600 (N = 425)	FORM (N = 434)	PLAC (N = 432)
Age, years: mean (SD)	63.4 (9.4)	63.3 (9.3)	64.0 (8.8)	63.6 (8.9)	63.2 (8.7)	63.4 (9.2)	63.9 (8.6)	63.9 (8.6)	62.9 (8.7)	63.6 (8.5)	63.2 (8.3)
Male sex: N (%)	259 (62.3)	263 (63.2)	269 (64.8)	255 (61.0)	238 (72.1)	249 (74.8)	258 (77.0)	351 (80.3)	327 (76.9)	348 (80.2)	352 (81.5)
BMI (kg/m^2): mean (SD)	27.0 (6.3)	26.7 (6.0)	26.9 (6.3)	26.3 (5.6)	25.9 (5.2)	25.7 (5.3)	25.3 (5.2)	26.2 (4.9)	26.5 (5.1)	26.3 (4.8)	26.8 (5.1)
ICS users: N (%)	159 (38.2)	155 (37.3)	145 (34.9)	165 (39.5)	149 (45.2)	152 (45.6)	135 (40.3)	243 (55.6)	226 (53.2)	221 (50.9)	224 (51.9)
Smoking history, pack-years: mean (SD)	48.3 (23.4)	50.8 (27.7)	50.0 (25.1)	49.7 (23.9)	39.6 (17.1)	40.0 (16.7)	41.0 (18.9)	48.6 (41.5)	53.6 (67.2)	49.0 (60.8)	53.3 (69.9)
FEV_1, L: mean (SD)	1.3 (0.5)	1.4 (0.5)	1.3 (0.5)	1.3 (0.5)	1.3 (0.5)	1.3 (0.5)	1.3 (0.5)	1. 5 (0.5)	1. 5 (0.5)	1. 5 (0.5)	1.5 (0.5)
FEV_1,% predicted: mean (SD)[a]	56.1 (14.5)	56.3 (14.5)	53.9 (15.6)	56.1 (14.3)	53.9 (14.0)	53.1 (13.6)	53.0 (14.2)	52.8 (13.6)	51.6 (13.2)	52.9 (14.2)	52.9 (14.1)
FEV_1/FVC, %: mean (SD)[a]	53.0 (10.0)	52.6 (10.1)	52.7 (10.1)	53.4 (10.1)	53.5 (10.0)	52.2 (9.9)	52.7 (11.0)	51.1 (10.7)	51.1 (10.6)	51.3 (10.5)	52.1 (10.6)
Reversibility (β2-agonist),%: mean (SD)[a]	15.6 (15.4)	15.2 (15.4)	15.6 (17.6)	15.5 (18.0)	11.7 (15.3)	11.0 (13.9)	12.7 (16.4)	11.7 (12.7)	13.7 (14.5)	11.8 (12.7)	12.7 (13.1)
Reversibility (anticholinergic),%: mean (SD)[b]	15.3 (15.4)	15.9 (21.9)	14.8 (16.1)	15.9 (18.3)	12.6 (14.4)	13.3 (14.3)	11.6 (13.9)	15.0 (12.5)	14.1 (14.5)	13.6 (14.6)	13.6 (13.4)

Note: IND = indacaterol; TIOT = tiotropium; SALM = salmeterol; PLAC = placebo.
[a] Measured 30 min after albuterol 360 μg inhalation.
[b] Measured 1 h after ipratropium 42 μg inhalation.

4. Results

Results for all three trials are presented in detail in this section.

4.1. Sample description

The patients' characteristics at baseline are shown in Table 1 and indicate that the patient populations in the three studies were similar, although a slightly higher proportion of patients in INVOLVE were male and were using ICS at baseline. None of the variables used in these analyses had distributions that were non-normal (statistics not shown). For example, skewness ranged from −0.35 to 0.92; kurtosis ranged from 1.1 to 3.8.

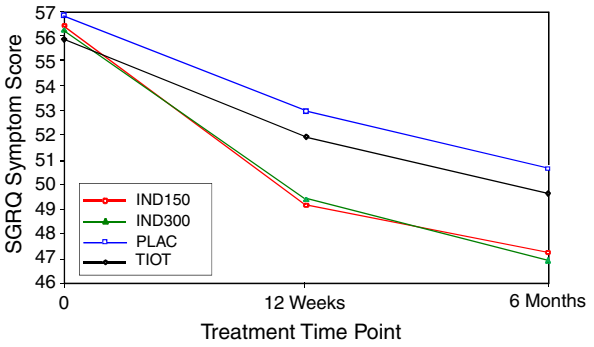
4.2. Latent growth models

Figs. 1, 2, and 3 show the slopes for the growth curve models for INHANCE, INLIGHT-2, and INVOLVE for the SGRQ Symptom domain scores, with the mMRC dyspnea as a time-varying covariate, controlling for gender, age, baseline COPD severity and FEV_1, and smoking status. The growth curves for SGRQ Symptom domain scores were linear for each trial, showing a decrease (improvement) for all groups over the course of INHANCE and for each active treatment group in INLIGHT-2 and INVOLVE (Figs. 1, 2, and 3, respectively). We assessed quadratic curves, but either the models did not converge or there was no improvement in fit, thus indicating that a straight linear slope adequately fit the data.

The results showed a greater overall improvement at six months for both doses of indacaterol over placebo in INHANCE, with a significant advantage at the end of the study for indacaterol 300 μg over tiotropium ($p<0.01$; Fig. 1). The slopes of change (i.e. trajectory from baseline to end of study) for both indacaterol 150 μg and 300 μg were significantly different from that of tiotropium ($t=-3.73$, $p<0.001$ and $t=-5.23$, $p<0.001$, respectively), indicating a greater overall rate of improvement with indacaterol than tiotropium.

There was a significant overall advantage of indacaterol 150 μg over placebo at weeks 12 and six months in INLIGHT-2 ($t=-3.84$, $p<0.001$; $t=-3.46$; $p<0.001$, respectively; Fig. 2). Salmeterol also showed a significant advantage over placebo at weeks 12 and six months ($t=-3.03$, $p<0.01$; $t=-2.73$, $p<0.01$, respectively). There was no significant difference in slopes of change between indacaterol 150 μg and salmeterol, though both were significantly different from placebo ($t=-7.98$, $p<0.001$ and $t=-7.95$, $p<0.001$, respectively; Fig. 2).

There were no significant differences in SGRQ Symptom scores between either dose of indacaterol (300 μg or 600 μg) or formoterol relative to placebo at six months in INVOLVE. This appears to have been a function of the amount of variability in each slope. Likewise, there were no significant differences between either dose of indacaterol and formoterol. However, significant advantages of indacaterol 300 μg, 600 μg and formoterol were found relative to placebo in the slopes of change (i.e., improvement over six months: ind 300 μg vs. placebo $t=-4.95$; ind 600 μg vs. placebo $t=-7.03$; formoterol vs. placebo $t=-5.35$; $p<0.001$ for all comparisons; see Fig. 3). That is, the slopes of change for the



NOTE: Patient age, sex, GOLD stage of COPD severity, smoking status, and baseline FEV_1 are covariates

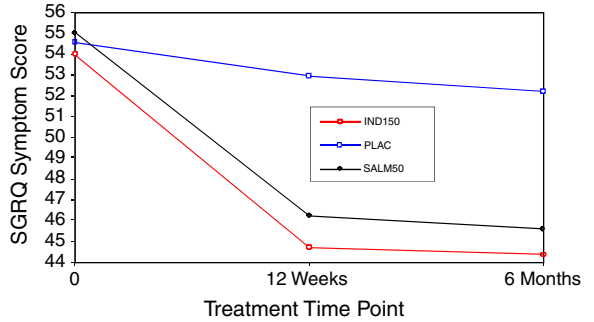
Fig. 1. Growth curves of SGRQ symptoms with mMRC dyspnea as a time-varying covariate.

three treatment doses were significantly different from placebo.

4.3. Growth mixture models

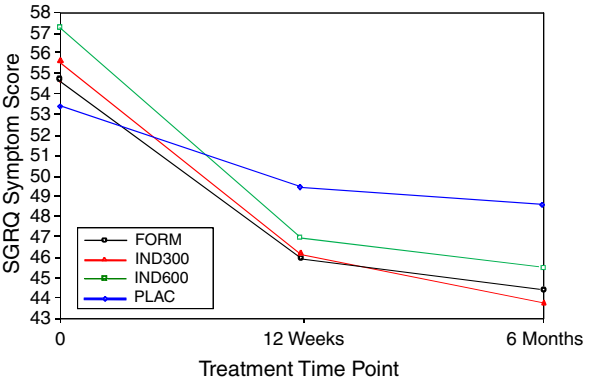
There was substantial variability around SGRQ scores (intercepts and slopes) in all trials (see Table 2), as indicated by standard deviations for the intercepts that were in excess of 30% of the magnitude of the mean scores, and with standard deviations that were often similar in magnitude of the means of the slopes. Although no convention appears to exist about how much variability should exist to recommend conducting mixture models, the high level of variability around the mean intercepts and growth trajectories within the present data may indicate subsets of patients who respond differently to treatment, but are masked when analyzing group means as a whole. Thus, it was deemed appropriate to apply GMM analysis to see if it would identify subgroups of differential responders hidden in this variability in SGRQ Symptom scores.

Table 3 presents the fit information used to make decisions about the acceptable number of latent classes extracted. For INHANCE and INLIGHT-2, the BIC and SABIC were marginally smaller for the 3-class solution, however, entropy was larger for the 2-class solution, especially for



NOTE: Patient age, sex, GOLD stage of COPD severity, smoking status, and baseline FEV_1 are covariates

Fig. 2. Growth curves of SGRQ symptoms with mMRC dyspnea as a time-varying covariate — INLIGHT-2.



NOTE: Patient age, sex, GOLD stage of COPD severity, smoking status, and baseline FEV_1 are covariates

Fig. 3. Growth curves of SGRQ symptoms with mMRC dyspnea as a time-varying covariate — INVOLVE.

INLIGHT-2. The smallest class size was slightly better for the 2-class solution for INHANCE. However, visual inspection of the growth trajectories of the 2- and 3-class solutions suggested clearer class distinctions and greater class separation (i.e., clearer differences in baseline scores and slopes) for the 2-class solutions. Thus, 2-class solutions were selected for these two trials. In contrast, for INVOLVE the results were somewhat mixed. The BIC and SABIC were clearly better for the 3-class solution, but entropy and smallest class size was better for the 2-class solution. However, visual inspection of the intercepts and slopes for the 2- versus 3-class solutions clearly favored to 3-class solution. There was much clearer separation of classes in terms of intercepts and slopes for the 3-class solution, whereas the 2-class solution resulted in intercepts and slopes that were not easily interpretable.

For INHANCE and INLIGHT-2, there were two distinct subsets of patients, based on response to the SGRQ Symptoms domain: one smaller subset of patients had high (poor) SGRQ Symptoms scores at baseline and these changed little, or in some instances deteriorated (Figs. 4 and 5; also see Table 4 for mean SGRQ Symptom scores by latent class). This subgroup is referred to as 'non-responders' due to the lack of improvement (i.e., decrease) in SGRQ Symptom scores. These non-responders – one slope per treatment arm – accounted for 21.5% of the entire trial sample in INHANCE and 18.2% in INLIGHT-2.

In INHANCE the subset of non-responders had substantially higher baseline SGRQ symptom scores than responders and changed very little over the six months (Fig. 4). Non-responders showed a slight decrease (improvement) in mean SGRQ symptom scores at six months (ranging from −1 to −4 [indicating improvement] for treatment and +1 [indicating deterioration] for placebo — see Table 4), while SGRQ symptom scores for responders were substantially lower (better) when non-responders were modeled separately (ranging from −8 to −12 for treatment and −8 for placebo — see Table 4). With non-responders modeled separately, tiotropium and placebo showed similar overall slopes of change in SGRQ Symptom scores and both showed less overall improvement in SGRQ Symptom scores than either indacaterol dose (see Table 4).

Table 2
LGM mean intercept and change, by treatment group: SGRQ symptom domain.

	INHANCE		INLIGHT-2		INVOLVE	
	Intercept (SD)	Slope: baseline–six months (SD)	Intercept (SD)	Slope: baseline–six months (SD)	Intercept (SD)	Slope: baseline–six months (SD)
IND150 (N = 416)	54.1 (19.5)	−13.5 (5.6)				
IND300 (N = 416)	52.0 (18.9)	−11.7 (6.0)				
PLAC (N = 418)	51.1 (18.5)	−5.1 (4.8)				
TIOT (N = 415)	52.3 (20.6)	−6.8 (5.2)				
IND150 (N = 330)			45.7 (18.4)	−6.1 (8.1)		
PLAC (N = 335)			48.0 (19.3)	−3.1 (6.1)		
SALM50 (N = 333)			47.6 (19.1)	−8.0 (5.5)		
IND300 (N = 437)					48.0 (17.5)	−7.6 (2.1)
IND600 (N = 425)					49.1 (17.8)	−9.9 (2.6)
FORM (N = 434)					52.7 (18.4)	−10.9 (3.0)
PLAC (N = 432)					52.9 (19.9)	−2.6 (2.4)

Note: IND = indacaterol; TIOT = tiotropium; SALM = salmeterol; PLAC = placebo.

As with INHANCE, the subset of non-responders in INLIGHT-2 emerging from the GMMs (the three slopes near the top of Fig. 5, one slope per treatment arm) showed a slight worsening in their scores from week 12 to month six and had SGRQ Symptom scores at baseline and across the trial that were substantially higher (poorer) than did the main treatment group from which they emerged (difference from baseline to six months ranged from −1 to +4 for treatment and +3 for placebo — see Table 4). Mean SGRQ Symptom scores for indacaterol 150 μg and salmeterol were substantially lower (better) when responders were modeled separately to non-responders (ranging from −11 to −13 for treatment and −3 for placebo — see Table 4).

GMM results for the INVOLVE study were, in some respects, different from those of INHANCE and INLIGHT-2. In particular, the GMMs for INVOLVE were the only analyses in which three sets of latent classes emerged (see Fig. 6). The three classes identified by the GMM analyses exhibited three different patterns of trajectories and were labeled 'responders', 'partial-responders' or 'non-responders'.

As shown in Fig. 6, non-responders had slightly higher SGRQ symptom scores at baseline compared to responders and showed a slight worsening in their scores from week 12 to month six. The difference from baseline to six months for non-responders ranged from +1 to +3 for treatment and was −2 for placebo. When non-responders and partial-responders were modeled separately, symptom scores were substantially greater for the indacaterol and formoterol responder groups — improvement from baseline to six months ranged from −14 to −17 for treatment and was −5 for placebo (see Table 4). Partial-responders had substantially lower baseline scores than responders or non-responders and exhibited less overall change in SGRQ Symptom scores compared with the overall treatment groups seen in the LGMs (ranging from −7 to −8 for treatment and −7 for placebo; see Figs. 3 and 6 and Table 4).

Table 3
Model fit information for the number of latent classes extracted.

	INHANCE	INLIGHT-2	INVOLVE
2-Class solution			
BIC	36404.9	23646.3	41224.8
SABIC	36284.2	23544.6	41078.6
Smallest class size	4.8%	5.3%	5.0%
Entropy	0.9	0.9	0.9
3-Class solution			
BIC	36449.1	23669.0	38447.5
SABIC	36283.9	23532.4	38285.5
Smallest class size	4.1%	6.5%	3.9%
Entropy	0.8	0.8	0.8

4.4. Characterizing the latent classes

Once the latent classes of differential responders were identified, the next step involved describing their baseline characteristics to see if and how the classes differed. Responders and non-responders (and partial-responders in INVOLVE) were then examined to identify if there were characteristics of each that were unique (see Table 5). Comparisons were made between responder groups on baseline respiratory characteristics and patient age. All comparisons used $p \leq 0.05$. Patient race was examined, but no tests of differences were calculated. Note that although multiple comparisons were made, no adjustments were made for multiplicity as these *post hoc* analyses were largely exploratory. It should be noted that half of the comparisons were statistically significant at or below 0.05.

Responders compared to non-responders had a lower (i.e., better) score on the SGRQ Symptom domain at baseline, lower (better) mMRC dyspnea score, were less likely to be a current smoker, and were older than the non-responders. For INVOLVE, compared with partial-responders, responders had a higher mean baseline SGRQ Symptom score, were significantly more likely to be current smokers, were significantly younger and had significantly worse baseline dyspnea.

5. Discussion

In this exploratory analysis, analytic methods were used to identify unobserved subgroups of patients with differential growth trajectories to those of other subgroups. This approach has the potential to increase understanding of treatment effects and identify patients more or less likely to

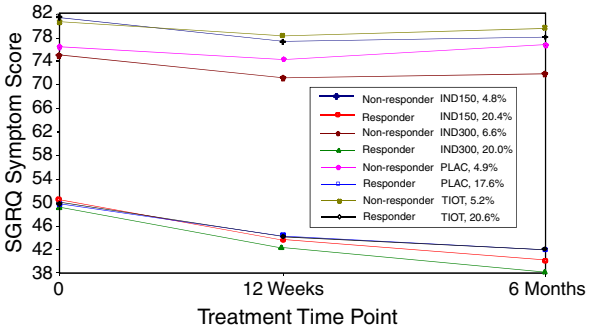


Fig. 4. Growth mixture models of SGRQ symptoms with mMRC dyspnea as a time-varying covariate — INHANCE.

benefit from treatment. Characterizing such patients can potentially inform the development of future trials and increase treatment effectiveness by tailoring programs for relevant populations. Therefore, we evaluated the utility of GMMs in identifying unobserved subgroups by analyzing SGRQ Symptom scores over the course of three COPD trials. Once identified, these subsets of patients were examined to yield a clearer picture of those patients who are most likely to show a positive response to treatment.

The use of GMM is a relatively recent analytic tool for examining heterogeneity in responses in clinical trial data, particularly when examining patterns of change. The LGM technique in the present evaluation resulted in linear growth curves for two 6-month trials and the first six months of a 12-month trial. These results indicated a decrease (improvement) in SGRQ Symptom scores for all active treatment groups.

The GMM analyses examined the variability in responses to the SGRQ Symptoms domain and identified two distinct subsets of patients in both 6-month trials that exhibited differential treatment response. While the proportion of non-responders within each trial was not exceptionally large, each subgroup was large enough and their SGRQ Symptoms scores different enough from that of the responders to have a noticeable effect on the slopes of change, when combined with their respective main treatment groups. The subsets of non-responder patients had high (poor) SGRQ Symptoms scores at baseline. Over the 6-month study period of INHANCE and INLIGHT-2, these scores showed little change and pulled the mean scores of the main treatment groups toward a higher (worse) value. Consequently, this subset negatively affected the overall group scores. That is, in the LGMs there was less apparent, overall improvement in the mean SGRQ Symptom scores and slopes of change in the treatment groups when non-responders were not identified separately, but were part of their respective, original treatment arms. Modeling the non-responders separately resulted in a greater apparent treatment effect on self-reported symptoms for all active treatments in both trials, and a differentiation for indacaterol against placebo in both 6-month studies and the comparator in INHANCE.

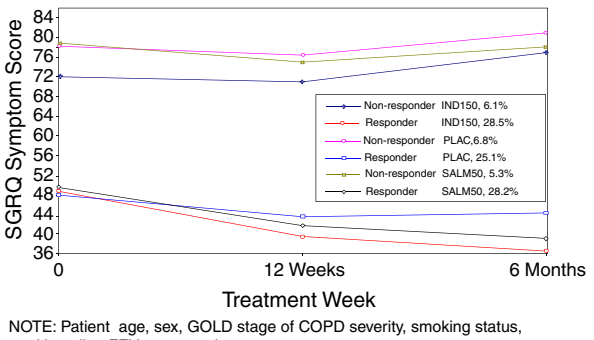


Fig. 5. Growth mixture models of SGRQ symptoms with mMRC dyspnea as a time-varying covariate — INLIGHT-2.

The 6-month analysis of the INVOLVE trial showed that a subset of non-responders also emerged. Unlike the other two studies, in INVOLVE three latent classes emerged: non-responders, responders, and partial-responders (patients who showed some treatment response but not as great as those in the responder class). Although scores for responders were consistently between those of the other two classes in INVOLVE, they were almost identical to those of the responders in INHANCE and INLIGHT-2; the non-responder scores were also consistent between the three studies. The partial-responders began the trial with quite low SGRQ Symptom scores and dropped, on average, about 6–8 points, compared with 14–21 points for the responders. They were also the class with the best overall baseline characteristics, but were significantly older. So, for the patients in INVOLVE it appears that those in better health and with better baseline SGRQ Symptom scores are less likely to show the dramatic improvement seen in responders.

A key treatment goal in COPD is symptomatic improvement and accordingly the present *post hoc* evaluation analyzed the Symptom domain of the SGRQ as a way of more closely assessing the symptomatic effects of COPD treatment. In the absence of pre-defined responders/non-responders (as in the case with the SGRQ Symptoms domain) GMMs can identify such subgroups efficiently and *post hoc* comparisons can help in understanding the differences between subgroups.

The analyses included several rigorous decision points about the significance of the latent subgroups and the credibility of the number and constitution of the subgroups. It is clear in the present analyses that the combined use of empirical and qualitative (i.e., visual) results aids in decisions on the optimal number of latent classes to extract. Once the latent class solution was selected for each trial, the sizes and the consistency of sizes of the latent subclasses, and similarity of the baseline scores and slopes of change within each of trials added confidence that the solutions selected were the most likely. Even in INVOLVE, the non-responders and responders were very much like those in the other two trials. So it appears that common processes and experiences are working in all three trials, but something unique – the partial-responders – occurred in the INVOLVE trial.

The GMM analyses used in the present study demonstrate that the use of this relatively recent analytic approach to examining change and variability in responses can aid in

Table 4
Mean SGRQ Symptom Scores by Visit for Responders, Non-Responders and Partial-Responders: Based on latent class assignment from Growth Mixture Models.

Time point	INHANCE								INLIGHT-2						INVOLVE									
	IND150		IND300		TIOT		PLAC		IND150		SALM50		PLAC		IND300			IND600			FORM			PLAC	
	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR	R	PR	NR	R	PR	NR	R	PR	NR	R	PR	NR
Baseline	51	82	50	75	51	81	51	76	49	72	50	79	48	78	62	28	67	64	28	66	62	29	69	52	26	75
Week 12	44	76	43	71	45	79	45	75	39	72	42	75	44	76	52	22	65	54	22	65	53	23	68	46	22	71
6 months	40	78	38	72	43	80	43	77	36	76	39	78	45	81	45	21	69	50	20	67	46	22	72	47	19	73
Difference from baseline to 6 months	−11	−4	−12	−3	−8	−1	−8	1	−13	4	−11	−1	−3	3	−17	−7	2	−14	−8	1	−16	−7	3	−5	−7	−2

Note: IND = indacaterol; TIOT = tiotropium; SALM = salmeterol; PLAC = placebo. R = responders; NR = non-responders; PR = partial-responders.

detecting subgroups exhibiting differential change efficiently with statistical tests that can increase confidence in the results. The GMM analyses made it possible to determine whether certain subgroups were more or less responsive to treatment, helping to explain the overall variability in treatment response.

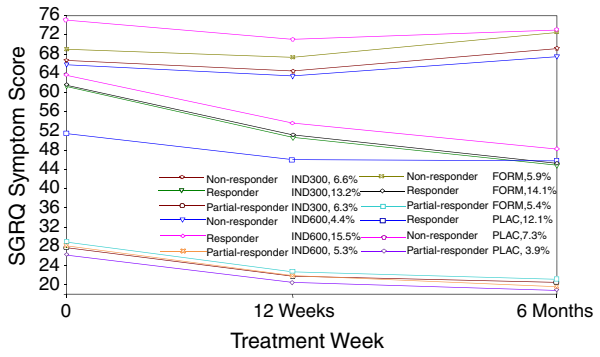
Predictably, a *post hoc* analysis of this type has limitations. First, there were no pre-defined subgroups of patients, or pre-determined analyses. It may be argued that this type of subgroup analysis is exploratory and represents a starting point for subsequent confirmatory analyses or clinical trials. That is, the results of analyses of only one study would only constitute exploratory findings until confirmed with analyses of an additional study. However, the opportunity to replicate the analyses in this study in similar trials, and the consistency of many of the results, increased our confidence in the results.

Second, notwithstanding the similarities in design of the two 6-month clinical trials, the treatment options were not identical necessitating the analyses of each study individually thus limiting the numbers of patients in each non-responder subgroup.

Finally, the SGRQ Symptom scale is not validated as a standalone subscale; only the SGRQ Total score is. Nonetheless, it focuses more closely on the symptom experience of patients with COPD and showed differences in change between active treatment groups and placebo, giving us greater opportunity to detect patient sub-groups.

Randomized control trials are designed to ensure that each treatment arm has a similar amount of heterogeneity due to patient randomization. This does not preclude consideration of significance tests on baseline characteristics [41]. However, in the absence of baseline differences, heterogeneity in treatment response can still exist and may result from influences outside the design of the trial, including differences in genetic makeup of patients. This heterogeneity in treatment response is often reflected in large variations around group means. Examining heterogeneity in treatment response can yield new insights into treatment efficacy [42]. The present analyses showed that when subgroups of non-responders are identified and modeled clearly, there are subsets of patients with COPD that exhibit an even more pronounced symptom relief from treatment than was apparent with a focus on the main trial treatment groups. This was evident from the progressive improvement in SGRQ Symptoms during the 6-month study periods. The characteristics of these subgroups produce insights into differences between responders and non-responders, and from a clinical practice perspective, could provide useful information regarding identification of patients likely to benefit the most from the treatment.

GMMs provide a way to analyze responses at an individual level and thus investigate whether there are subsets of differential responders that are masked when whole group means are analyzed. If included early in a clinical trial program, the results can provide insights into subgroups most/least likely to respond to treatment. This information could inform the design of later trials thereby reducing heterogeneity and providing a more accurate assessment of treatment efficacy. However, the extent to which subgroups of differential responders can be adequately characterized depends largely on the extent of data collected as part of the study. Pharmaceutical companies intending to apply this



NOTE: Patient age, sex, GOLD stage of COPD severity, smoking status, and baseline FEV_1 are covariates

Fig. 6. Growth mixture models of SGRQ symptoms with mMRC dyspnea as a time-varying covariate — INVOLVE.

methodology in their trials should anticipate factors that result in heterogeneity of response and design studies to collect data that could be used in *post hoc* analyses to characterize the subgroups of differential responders. This could reduce the chances of finding subgroups who are more responsive to treatment but who cannot be clearly defined on observed characteristics.

6. Conclusion

Patients with COPD represent a heterogeneous population in terms of their reporting of symptoms and response to treatment. The GMM analyses provided greater insight into treatment response than was evident with the LGMs. We found that the GMM methods are able to identify sub-groups of responders and non-responders to the SGRQ symptoms component score and that when the responders are analyzed separately from the non-responders there are increased treatment effects and, in some cases, increased differences between treatments. Application of this novel methodology could be of value in examining other endpoints in COPD and other disease areas. Future work could involve studying the accuracy of classifying patients to their respective responder groups based on the baseline predictor variables.

Disclosure statement

Dr Stull and Ms Houghton are employed by RTI Health Solutions (RTI-HS), while Dr Wiklund is employed by United BioSource Corporation (UBC). Both companies provide consulting and other research services to pharmaceutical, device, government, and non-government organizations. In their salaried positions, they work with a variety of companies and organizations. They receive no payment or honoraria directly from these organizations for services rendered. Mr Gale is an employee of Novartis. Professor Jones has received remuneration for providing consulting services to Novartis.

The authors would like to acknowledge the contribution of David Young who provided editorial review in the preparation of this manuscript.

At the time of submission of the manuscript, Dr Stull and Ms Houghton were employed by UBC. RTI-HS was not affiliated in the development of this work.

Table 5
Baseline differences between responders and non-responders: based on latent class assignment for the SGRQ symptoms domain from growth mixture models.

Variable	INHANCE		INLIGHT-2		INVOLVE		
	Non-responders (n = 252)	Responders (n = 729)	Non-responders (n = 158)	Responders (n = 671)	Non-responders (n = 300)	Partial-responders (n = 284)	Responders (n = 797)
SGRQ symptoms domain: mean	73.7	50.5***	79.3	48.3***	73.3	24.6[d]	60.4[e, f]
Smoking history (% smoker)	60%	40%***	70%	40%***	50%	30%[d]	40%[f]
Baseline FEV_1	1.3	1.3	1.3	1.3	1.3	1.4[d]	1.3[f]
COPD severity (GOLD) n (%)[a]							
At risk	0 (0%)	1 (0.1%)	0 (0%)	3 (0.5%)	10 (3.3%)	2 (0.7%)	12 (1.5%)
Mild	13 (5.2%)	26 (3.6%)	0 (0%)	18 (2.7%)	3 (1.0%)	7 (2.5%)	13 (1.6%)
Moderate	130 (51.6%)	424 (58.2%)	76 (48.1%)	377 (56.2%)	146 (48.7%)	172 (60.6%)	420 (52.7%)
Severe	109 (43.3%)	277 (38.0%)	82 (51.9%)	272 (40.5%)	136 (45.3%)	100 (35.2%)	342 (42.9%)
Very severe	0 (0%)	1 (0.1%)	0 (0%)	1 (0.2%)	5 (1.7%)	3 (1.1%)	10 (1.3%)
Patient age	61.8	64.0***	61.8	63.7*	62.5	65.2[d]	63.6[f]
MMRC baseline dyspnea[a]	1.9	1.6***	1.7	1.5*	1.9	1.4[d]	1.6[e, f]
Reversibility (SABA)[b]	1.5	1.5	1.4	1.5	1.5	1.6[d]	1.5
Reversibility (anti-cholinergic)[c]	1.6	1.5	1.5	1.5	1.5	1.6[d]	1.5[f]
Race							
White	231 (95.6)	581 (84.9)	151 (95.6)	482 (71.8)	280 (92.3)	257 (90.5)	736 (92.3)
Black	10 (4.0)	14 (2.1)	–	1 (0.2)	1 (0.3)	1 (0.4)	1 (0.1)
Asian	9 (3.6)	80 (11.7)	–	126 (18.8)	6 (2.0)	2 (0.7)	16 (2.0)
Other	2 (0.8)	9 (1.3)	7 (4.4)	62 (9.2)	13 (4.3)	24 (8.5)	44 (5.5)

[a] 1 = not troubled by breathlessness except in strenuous exercise, 5 = too breathless to leave house, or breathless when dressing or undressing.
[b] = demonstrated increase in FEV_1 compared to the pre-bronchodilator value, within 30 min after inhalation of 4 × 100 μg puffs of salbutamol/albuterol.
[c] = demonstrated increase in FEV_1 compared to the pre-bronchodilator value, 1 h after inhalation of 2 × 21 μg ipratropium bromide.
[d] = difference ($p<0.05$) between non-responders and partial-responders.
[e] = difference ($p<0.05$) between non-responders and responders.
[f] = difference ($p<0.05$) between partial-responders and responders.
* $p<0.05$.
** $p<0.01$.
*** $p<0.001$.

Role of the funding source

Novartis, Inc. provided funding for this study. Mr. Gale, as an employee of Novartis, Inc., made available the data for the analyses conducted in this study. Mr. Gale collaborated with Drs. Stull and Wiklund and Ms. Houghton on the development of the research questions that were examined and interpretation of the overall results in relation to the drug under study. Mr. Gale collaborated with writing and revision of the Introduction and Discussion sections to ensure the clinical accuracy of the treatments being evaluated.

Acknowledgments

Funding for this study was provided by Novartis, Inc.

References

[1] Kaplan SH, Billimek J, Sorkin DH, Ngo-Metzger Q, Greenfield S. Who can respond to treatment? Identifying patient characteristics related to heterogeneity of treatment effects. Med Care Jun 2010;48(6 Suppl): S9–S16.
[2] Leoutsakos JM, Muthen BO, Breitner JC, Lyketsos CG. For the ADAPT Research Team. Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. Int J Geriatr Psychiatry May 10 2011, doi:10.1002/gps.2723.
[3] Muthen B, Brown HC. Estimating drug effects in the presence of placebo response: causal inference using growth mixture modeling. Stat Med Nov 30 2009;28(27):3363–85.
[4] Muthén B, Brown CH, Hunter A, Cook IA, Leuchter AF. General approaches to analysis of course: applying growth mixture modeling to randomized trials of depression medication. In: Shrout PE, editor. Causality and psychopathology: finding the determinants of disorders and their cures. New York: Oxford University Press; 2011. p. 159–78.
[5] Stull DE. Analyzing growth and change: latent variable growth curve modeling with an application to clinical trials. Qual Life Res Feb 2008;17(1): 47–59.
[6] Muthen B, Brown CH, Masyn K, Jo B, Khoo ST, Yang CC, et al. General growth mixture modeling for randomized preventive interventions. Biostatistics Dec 2002;3(4):459–75.
[7] Kerstjens HA, Brand PL, Quanjer PH, van der Bruggen-Bogaarts BA, Koeter GH, Postma DS. Variability of bronchodilator response and effects of inhaled corticosteroid treatment in obstructive airways disease. Dutch CNSLD study group. Thorax Jul 1993;48(7):722–9.
[8] Taskin D, Kesten S. Long-term benefits with tiotropium in COPD patients with and without short = term bronchodilator responses. Chest 2003;123(5):1441–9.
[9] Calverley PM, Rennard SI. What have we learned from large drug treatment trials in COPD? Lancet Sep 1 2007;370(9589):774–85.
[10] Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med Sep 15 2007;176(6):532–55.
[11] WHO. World Health Organization. Chronic respiratory diseases. Chronic obstructive pulmonary disease (COPD). 2010 [updated 2010; cited 2 June 2010]; Available from http://www.who.int/respiratory/copd/en/.
[12] Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J Jun 2004;23(6):932–46.
[13] GOLD. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for diagnosis, management, and prevention of COPD. 2009 [updated 2009; cited 2 June 2010]; Available from http://www.goldcopd.org/Guidelineitem.asp?l1=2&l2=1&intId=2003.
[14] Jones PW, Agusti AG. Outcomes and markers in the assessment of chronic obstructive pulmonary disease. Eur Respir J Apr 2006;27(4): 822–32.
[15] Donohue JF. Minimal clinically important differences in COPD lung function. COPD Mar 2005;2(1):111–24.
[16] Miravitlles M, Anzueto A, Legnani D, Forstmeier L, Fargel M. Patient's perception of exacerbations of COPD—the PERCEIVE study. Respir Med Mar 2007;101(3):453–60.
[17] Cazzola M, MacNee W, Martinez FJ, Rabe KF, Franciosi LG, Barnes PJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. Eur Respir J Feb 2008;31(2):416–69.
[18] Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest Mar 1988;93(3):580–6.
[19] Canadian Thoracic Society. Canadian Thoracic Society recommendations for management of chronic obstructive pulmonary disease — 2003. Can Respir J 2003;10(Suppl A):11A–65A.
[20] Gold PM. The 2007 GOLD Guidelines: a comprehensive care framework. Respir Care Aug 2009;54(8):1040–9.
[21] EMEA Committee for Proprietary Medicinal Products (CPMP). Points to consider on clinical investigation of medicinal products in the chronic treatment of patients with chronic obstructive pulmonary disease (COPD); 1999.
[22] Food and Drug Administration (FDA. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist 2009;74(235):65132–3.
[23] Beier J, Chanez P, Martinot JB, Schreurs AJ, Tkacova R, Bao W, et al. Safety, tolerability and efficacy of indacaterol, a novel once-daily beta (2)-agonist, in patients with COPD: a 28-day randomised, placebo controlled clinical trial. Pulm Pharmacol Ther 2007;20(6):740–9.
[24] Rennard S, Bantje T, Centanni S, Chanez P, Chuchalin A, D'Urzo A, et al. A dose-ranging study of indacaterol in obstructive airways disease, with a tiotropium comparison. Respir Med Jul 2008;102(7): 1033–44.
[25] Sustained 24-h bronchodilation with indacaterol once-daily in COPD: a 26-week efficacy and safety study. In: Fogarty C, Worth H, Hébert J, Iqbal A, Owen R, Higgins M, et al, editors. American Thoracic Society conference; 2009. [May 15–20 2009; San Diego, CA, USA].
[26] Fogarty C, Hébert J, Iqbal A, Owen R, Higgins M, Kramer B, editors. Indacaterol once-daily provides effective 24-h bronchodilation in COPD patients: a 26-week evaluation versus placebo and tiotropium. Vienna, Austria: European Respiratory Society; 2009. [12–16 September 2009].
[27] Indacaterol once-daily improves day and night-time symptom control in COPD patients: a 26-week study versus placebo and tiotropium. In: Lötvall J, Cosio BG, Iqbal A, Swales J, Owen R, Kramer B, et al, editors. European Respiratory Society conference; 2009. [12–16 September 2009; Vienna, Austria].
[28] Indacaterol once-daily improves health-related quality of life in COPD patients: a 26-week comparison with placebo and tiotropium. In: Yorgancıolu A, Mahler DA, Iqbal A, Owen R, Higgins M, Kramer B, editors. European Respiratory Society conference; 2009. [12–16 September 2009; Vienna, Austria].
[29] Kornmann O, Dahl R, Centanni S, Dogra A, Owen R, Lassen C, et al. Once-daily indacaterol vs twice-daily salmeterol for COPD: a placebo-controlled comparison. Eur Respir J Aug 6 2010.
[30] Dahl R, Chung KF, Buhl R, Magnussen H, Nonikov V, Jack D, et al. INVOLVE (INdacaterol: Value in COPD: Longer Term Validation of Efficacy and Safety) Study Investigators. Efficacy of a new once-daily long-acting inhaled beta2-agonist indacaterol versus twice-daily formoterol in COPD. Thorax Jun 2010;65(6):473–9.
[31] Donohue JF, Fogarty C, Lotvall J, Mahler DA, Worth H, Yorgancioglu A, et al. Once-daily bronchodilators for chronic obstructive pulmonary disease: indacaterol versus tiotropium. Am J Respir Crit Care Med Jun 3 2010.
[32] Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. Chest Jun 1984;85(6): 751–8.
[33] Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis Jun 1992;145(6):1321–7.
[34] Curtis JR, Patrick DL. The assessment of health status among patients with COPD. Eur Respir J Suppl Jun 2003;41:36s–45s.
[35] Indacaterol once-daily improves health-related quality of life (HRQOL) in COPD patients: a 52-week study. In: Magnussen H, Paggiaro P, Jack D, Owen R, Higgins M, Kramer B, editors. European Respiratory Society conference; 2009. [12–16 September 2009; Vienna, Austria].
[36] Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. J Clin Epidemiol Dec 2003;56(12): 1170–6.
[37] Lubke G, Muthen BO. Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. Structural Equation Modeling: A Multidisciplinary Journal 2007;14(1): 26–47.

[38] Muthén LK, Muthén BO. Mplus. Version 4.21. Los Angeles: Muthén & Muthén; 2007.
[39] Leite WL, Cooper LA. Detecting social desirability bias using factor mixture models. Multivar Behav Res 2010;45(2):271–93.
[40] Tofighi D, Enders CK. Identifying the correct number of classes in growth mixture models. In: Hancock GR, Samuelsen KM, editors. Advances in latent variable mixture models 2008(Information Age Publishing, Inc):317–41.
[41] Berger VW. Do not test for baseline imbalances unless they are known to be present? Qual Life Res May 2009;18(4):399.
[42] Stull DE. Uncovering heterogeneity in clinical studies: factor mixture models growth mixture models. ISPOR 14th Annual International Meeting; 2009; Paris, France; 2009.