Outcome measures in chronic obstructive pulmonary disease (COPD): strengths and limitations

Thomas Glaab¹, Claus Vogelmeier² and Roland Buhl*¹

Abstract
Current methods for assessing clinical outcomes in COPD mainly rely on physiological tests combined with the use of questionnaires. The present review considers commonly used outcome measures such as lung function, health status, exercise capacity and physical activity, dyspnoea, exacerbations, the multi-dimensional BODE score, and mortality. Based on current published data, we provide a concise overview of the principles, strengths and weaknesses, and discuss open questions related to each methodology. Reviewed is the current set of markers for measuring clinically relevant outcomes with particular emphasis on their limitations and opportunities that should be recognized when assessing and interpreting their use in clinical trials of COPD.

Introduction
Chronic obstructive pulmonary disease (COPD) is a heterogeneous, multi-component disease associated with significant clinical burden. Though the presence of airflow limitation is well recognised as the pathophysiological basis, COPD as a complex disorder requires a multifaceted approach with regard to clinical assessment and response to therapy. This has prompted an intense search for clinical trial endpoints that may adequately reflect the success or failure of treatment. Current methods for assessing COPD progression mainly rely on lung function tests with a particular focus on forced expiratory volume in 1 second (FEV₁). However, clinical and patient-reported outcome measures such as dyspnoea, exercise capacity, physical activity, exacerbations, health status and mortality have been recognized and applied as an essential part of the clinical assessment of COPD beyond FEV₁ measurements [1,2] (figure 1).

In recent years, a profound analysis of available outcomes and markers has been provided by the scientific community [3,4]. The objective of this review is to provide a concise overview of the feasibility, strengths and limitations of major outcome measures commonly applied in current COPD trials.

Lung function: forced expiratory volume in 1 second (FEV₁)
Relevance
It is well established that patients with COPD lose lung function at a steeper rate than subjects without COPD. Post-bronchodilator forced expiratory volume (FEV₁) is the single most important marker to determine severity and treatment algorithms in COPD. The decline of FEV₁ over time has been traditionally used to indicate disease progression.

Measures
The diagnosis, staging and treatment of COPD in current guidelines is based on the fixed ratio of FEV₁/FVC (forced vital capacity) and the percentage predicted FEV₁ value.

The methodology for measuring forced expiratory maneuvers by spirometry has been standardized by ATS/ERS [5]. Specific training to yield reproducible and reliable results is mandatory.

Strengths
• FEV₁ and FVC measurements are highly reproducible if performed adequately.
• Reduction in lung function is a risk factor for all cause and cardiovascular mortality [6-8], and impaired health status [9].

* Correspondence: r.buhl@3-med.klinik.uni-mainz.de
¹ Pulmonary Department, University Hospital, Johannes Gutenberg-University, Mainz, Germany
Full list of author information is available at the end of the article

© 2010 Glaab et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
absent, thus creating the opportunity of early intervention [10].

**Limitations**
- FEV1 measurements are based on an artificial manoeuvre and do not always correlate with clinically relevant outcomes such as dyspnoea, health status, exercise capacity, or exacerbations [4,11].
- Patients with similar FEV1 may represent different underlying phenotypes.
- Reference equations for lung function by European Community for Coal and Steel are disputed and limited in predicting lung function in the general population [12].
- Changes in lung volumes can occur without concomitant changes in FEV1 and are more closely related than FEV1 changes to exercise performance [13].
- No minimal important difference (MID) has been established yet. It was suggested that an appropriate range of values for the MID for FEV1 might be 100-140 mL [4] but the MID for FEV1 remains poorly defined for COPD [14].

**Open Questions and Outlook**
FEV1, while a crucial marker, is far from being the only measure to comprehensively characterize patients with COPD. Additional outcome measures are usually needed to assess the clinical benefit of therapeutic agents. The relationships between changes in airway structure and measures of lung function require further investigation.

**Lung volumes**

**Relevance**
Changes in absolute lung volumes can occur in COPD patients even in the absence of FEV1 changes. Progressive hyperinflation due to airflow limitation and loss of lung elastic recoil not only increases the work required during inspiration but also profoundly decreases the ventilatory reserve and increases the sense of effort and dyspnoea [15].

The assessment of absolute lung volumes has been standardized but is technically more demanding than simple spirometry. Specific training to yield reproducible and reliable results is essential.

**Measures**
Static lung hyperinflation and its increase during exercise (dynamic hyperinflation) are measured as elevations of total lung capacity (TLC), functional residual capacity (FRC), residual volume (RV) and as a decrease in inspiratory capacity (IC). The variability of lung volume measurements has been reviewed elsewhere [16].

**Strengths**
- Indices of dynamic hyperinflation correlate better than FEV1 with activity limitation and exertional dyspnoea [13,15] and pharmacological and surgical lung volume reduction have been associated with improvements in exercise performance and dyspnoea [17,18].
• A severely reduced IC/TLC ratio with a threshold value of 25% has been shown to predict mortality in COPD patients [19].

**Limitations**

• Body plethysmography remains the gold standard for the measurement of lung volumes such as TLC, FRC and RV. Spirometrically derived assessments of lung hyperinflation are more difficult to interpret in the absence of simultaneous bodyplethysmographic volume measurements to rule out a concomitant restrictive ventilatory disorder [15].

• The reproducibility of FRC, IC and RV in absolute values has yet to be demonstrated. Measurement of IC alone is not a reliable marker of lung hyperinflation and does not consistently reflect changes in FRC or TLC [15].

• Neither a standardized classification for the assessment of severity of hyperinflation nor a MID have been established yet. In practice, values of RV, TLC and FRC exceeding 120-130% of the predicted value are regarded to be clinically relevant, but these cut-offs are not validated.

• The natural course of dynamic hyperinflation in COPD is unknown and seems likely to be highly variable among COPD patients [15].

**Open Questions and Outlook**

In the absence of any consensus on the definition and/or severity of hyperinflation, it has been proposed that hyperinflation - preferentially expressed as % predicted - should be specified in terms of the volume compartment referred to and the measuring method used [12]. So far, there have been no studies aimed at exploring the longitudinal course of dynamic hyperinflation and its impact on the course of the disease in COPD patients.

**Exercise Capacity and Physical Activity**

**Relevance**

Reduced exercise capacity is considered to be a consequence of airflow obstruction, primarily because of dynamic hyperinflation occurring during exercise. Reduced physical activity of patients is a result of COPD, but at the same time promotes worsening and progression of the disease [20].

**Measures**

There are different approaches to determine the exercise capacity or activity levels of COPD patients (table 1): Higher exercise tolerance measured via laboratory or field tests can be translated to higher levels of activity. In addition, physical activity during daily life can be assessed directly by measuring energy expenditure or by mechanical assessment of movement.

**Exercise Capacity**

**6-Minute Walk Test (6MWT)**

Measurement of the distance walked during a 6-minute period on a level surface [21]. The principal outcome of this self-paced test is the distance covered. The MID is estimated to be 54-80 meters [27].

**Strengths**

• 6MWT is relatively simple to perform and well tolerated.

• 6MWT reflects everyday life-like activity.

• 6MWT is validated and standardized [28].

• The test results correlate with lung function, health status, and maximal VO₂ [29], and have shown to be predictive for mortality [30].

**Limitations**

• There are many sources of variability, e.g. patient’s motivation, weight, height, age, sex, co-morbidities, and day-to-day variability [28].

• 6MWT was significantly reduced only in COPD patients with GOLD stages III and IV [26,31].

• Assessment of the 6MWT is associated with spatial requirements and is personnel- and time-consuming.

• Standards of 6MWT are not always realisable. This might influence the results, e.g. shorter corridors reduce the distance covered because of time-consuming change in direction.

• Learning effect: Walking distance is up to 17% higher for a second test performed a day later [28].

**Table 1: Different methods to determine exercise capacity or activity levels in COPD**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>field tests to determine exercise capacity</td>
<td>6-Minute Walk Test [21] Shuttle Walk Test [22,23]</td>
</tr>
<tr>
<td>assessment of activity levels</td>
<td>accelerometer [26]</td>
</tr>
</tbody>
</table>
Shuttle Walk Test (SWT)
There are two forms of assessment: In the Incremental Shuttle Walk Test, walking speed is set by the frequency of an acoustic signal. The frequency increases progressively until patients can no longer pick up the pace. The principal outcome is the distance covered. The MID is estimated to be 47.5 meters [32].

The Endurance Shuttle Walk Test has been developed to determine sub-maximal exercise capacity with the acoustic signal frequency being constant throughout the walk [23]. The principal outcome is the duration of exercise. No MID has yet been described.

Strengths
• SWT is relatively simple to perform and well-tolerated.
• Learning effects are minimal.
• Walking pace is externally controlled.

Limitations
• Instructions for SWT are time consuming.
• The test is less extensively validated than the 6MWT. Solid evidence for validity still has to be provided [27,33].
• SWT does not reflect common daily activities that require endurance and pacing.

Ergometry
To evaluate the exercise response, bicycle-ergometer or treadmill are commonly used in two different test modes. In incremental-workload tests, work-rate is increased progressively as a mild continuous ramp under computer control with the principal outcome being the distance covered. Alternatively, constant-workload tests have been performed at sub-maximal levels of exercise intensity which is typically set between 75% and 85% of the maximum workload during incremental tests [20]. The principal outcome is the duration of workload.

Reasons for break-off, e.g. leg discomfort vs. breathlessness, provide additional insights [34].

Strengths
• Standardized protocols are available [35].
• Treadmill walking reflects an activity of daily living.
• Cycle ergometer is less prone to introduce movement or noise artefacts into measurements than treadmill, and electrocardiogram and blood pressure are generally easier to measure [35].
• Additional physiological and clinical variables, such as peak O₂ uptake, CO₂ output, minute ventilation, heart rate, dyspnoea, and leg discomfort can be determined in parallel.

Limitations
• The workload not only depends on speed and inclination of the treadmill but also on the weight of the subject and pacing strategy. Body weight has much less effect on bicycle ergometry performance [30].
• Cycling is less closely related to the patient’s activities of daily living.

• Resources: Ergometers are relatively expensive, treadmills require much space.
• No MID has been established yet.

Physical Activity

Sensors for physical activity
The methods that are available to quantify physical activity in daily life include direct observation, assessment of energy expenditure, and the use of physical activity questionnaires or motion sensors. In particular, motion sensors are practical tools for clinical trials or practice. Accelerometers are electronic devices that record energy expenditure or mechanically assess movement. The devices are usually worn on patients’ arm or waist. Accelerometers read out stored data as movement intensity and as quantity and can also provide data on body posture.

Strengths
• Accelerometers generate objective data by determination of quantity and intensity of body movements.
• Significant limitations of physical activity can already be detected in patients with moderate COPD (GOLD stage II) [31].

Limitations
• Solid evidence for reliability, validity and responsiveness for different types of accelerometers still has to be provided [26,31].
• Some activity sensors are poorly accepted by patients [36].
• Variability in sensitivity among accelerometers of a given model has been detected [37].
• Accelerometers may be sensitive to artefacts like car vibrations [26].
• Activity sensors may actually fail to accurately capture the inactive life style of patients with COPD [38].
• Physical activity patterns vary from day to day and between week-days and weekend due to the patient’s health, or external factors [31]. In long-term studies, another source of variability may be seasonal climate changes, hours of daylight and weather [38].
• Observation bias: a greater level of activity may be induced during the measurement period that results in overestimation of the activity [39]. On the other hand, underreporting bias may evolve from poor compliance [26,39].
• No MID has been established yet.

Open questions and outlook
Exercise capacity is an important clinical outcome in interventional trials of COPD, but it is still debatable what is the most valid, reliable, and responsive measurement of changes within subjects.

Physical activity may become a key outcome measure not only in clinical trials of COPD, but also in rehabilitation programs and for patients’ self-management. Even
though the technical assessment of physical activity is improving rapidly, not all new techniques have been developed to the point where their clinical utility has been validated.

Little is known about the agreement of exercise capacity as measured using different methods. Therefore, indirect comparisons of treatment effects on exercise capacity are obscured by different methods of assessment applied in various trials.

**Dyspnoea Relevance**
For patients with COPD, dyspnoea is the most frequent complaint for which they seek medical attention. However, dyspnoea is a subjective measure that poorly correlates with objective assessments of lung function, exercise capacity, and other outcomes [1].

**Measures**
Different approaches have been used to measure dyspnoea in clinical trials, amongst which the BDI/TDI, Borg-Scale, and MRC are applied most often (table 2).

**Baseline Dyspnoea Index/Transition Dyspnoea Index (BDI/TDI)**
The BDI and TDI represent one of the most commonly applied instruments for dyspnoea rating in clinical trials, describing symptoms at a single point in time (e.g., baseline (BDI)), and measuring changes in breathlessness from this baseline state over time (TDI) [40].

BDI and TDI ratings are obtained in the course of an interview conducted by an experienced observer, who asks open-ended questions about the patient’s experience of breathlessness during everyday activities, which are then translated into numerical values.

**Strengths**
• BDI and TDI ratings provide multi-dimensional measurements of breathlessness (functional impairment, magnitude of task, and magnitude of effort) related to activities of daily living.
• A MID with a difference of 1 unit for the mean total score being considered clinically meaningful is available though it is mainly based on retrospective data analysis [41,42].

**Limitations**
• Interviewer bias: Neither interviewer questions nor the translation of patients’ answers to ratings are standardized, enforcing thorough interviewer training.
• Recall bias: The patient has to recall baseline state (BDI) in order to answer questions regarding the TDI.
• Assessment bias: Interviewer blinding to patients’ clinical status is necessary to prevent assessment bias.

**Medical Research Council (MRC) Scale**
The MRC dyspnoea scale was developed as a simple and standardised method of categorising disability in COPD [4].

The patient selects a grade on the self-applied 5-point instrument that describes everyday situations or activity levels provoking breathlessness and impairment. A MID has not been established.

**Strengths**
• The method has been widely used in the past [43-46].

**Limitations**
• A possible underestimation bias due to avoidance of exertion has to be taken into account [47].
• The MRC is relatively insensitive to change, e.g. due to therapeutic intervention [48,49].
• There are relatively scarce clinical data on validation, responsiveness, and sensitivity [43].

**Borg-Scale (CR-10)**
The CR-10 or Borg-Scale has been developed primarily as an objective tool to measure exertional dyspnoea in COPD patients [50,51]. Although the 10-point category ratio scale is easy to use, concise and detailed instructions for patients are indispensable for appropriate application [52]. Based on retrospective analysis, a MID for the Borg-Scale in the range of 1 unit has been discussed [4].

**Open Questions and Outlook**
More research is needed to optimize and validate questionnaire items including direct patient involvement in instrument generation to improve their utility in clinical trials. Little is known about the impact of concomitant disorders on outcomes, e.g. if disorders such as anxiety or depression influence perceived dyspnoea and - if so - to which extent those instruments applied today reflect that influence. Furthermore, studies are needed to show

---

**Table 2: Dyspnoea measurement scales**

<table>
<thead>
<tr>
<th>Type of scale</th>
<th>Type of stimulus</th>
<th>Items</th>
<th>Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI/TDI</td>
<td>multi-dimensional</td>
<td>everyday activities</td>
<td>8/9 interview</td>
</tr>
<tr>
<td>MRC-Scale</td>
<td>uni-dimensional</td>
<td>everyday activities</td>
<td>1 self-administered by patient</td>
</tr>
<tr>
<td>Borg-Scale</td>
<td>uni-dimensional</td>
<td>under exertion</td>
<td>1 self-administered by patient</td>
</tr>
</tbody>
</table>

BDI: Baseline Dyspnoea Index; TDI: Transition Dyspnoea Index; MRC: Medical Research Council.
which of the existing methodologies, e.g. questions or word lists, should be preferred in the context of COPD.

Health Status

Relevance
Health-status is considered one of the main patient-related outcomes in clinical trials. It is important to make a distinction between quality of life (QoL), which is unique to the individual, and health status measurement, which is a standardized quantification of the impact of disease [53].

Measures
Health-status as a concept of high complexity is assessed indirectly and requires the application of specially designed questionnaires (table 3).

St. George’s Respiratory Questionnaire (SGRQ)
The SGRQ was originally developed to measure health status in patients with respiratory disease, e.g. COPD or asthma [54]. A COPD-specific version is available [55].

The SGRQ covers domains of symptoms (frequency and severity of respiratory symptoms), activity (effects on and adjustment of everyday activities), and psychosocial impact, from which a total score with a possible maximum of 100 points is calculated.

The MID was assessed by various methods. Changes of 2 to 8 points were considered clinically meaningful, with a value of 4 applied most often [56].

Strengths
• The SGRQ has been widely used in clinical trials as a secondary endpoint to assess the effects of treatment and management interventions on health status in COPD.
• It may be considered a quasi standard in clinical trials.

Limitations
• The instrument is time-consuming to implement and is therefore of limited applicability in day-to-day clinical practice.
• There is a trend bias due to non-poled questions (first possible answer is usually “yes” and indicates worse health-status) [57].
• The processing of missing answers is unsatisfactory. A missing answer is considered as if the patient had answered “no” (indicating better health-status) [57].

• SGRQ scores were shown to be influenced by subjects’ sex, age, education, and by comorbidities [58].
• Suitability of MID for individual patients as opposed to patient group comparisons has yet to be shown.
• Linearity of differences between SGRQ values has not been shown, especially not in different stages of severity. Thus, it is unknown, whether a reduction in SGRQ total score by 4 points (e.g. from 44 to 40) represents a subjective improvement in health status equivalent to a reduction from 64 to 60.
• There is little published empiric evidence supporting the MID of four points [59].

Chronic Respiratory Disease Questionnaire (CRQ)
The CRQ measures physical-functional and emotional limitations due to chronic lung diseases including COPD [60]. It refers to activity-related dyspnoea with results covering dyspnoea, fatigue, emotion, and mastery. The questionnaire has primarily been applied in rehabilitation trials of COPD patients [61].

The patient is asked to recall the five most important activities that caused breathlessness over the past two weeks. A total score as well as individual subscale scores can be calculated. A difference of 0.5 for the mean domain scores is considered clinically meaningful [62].

Strengths and limitations
A distinctive property of this instrument is the patient-specific selection of five activities, which cause dyspnoea for the individual patient. This way the instrument adapts to the specific conditions of the patient and is sensitive to treatment. On the other hand, the instrument is less suitable for inter-individual comparisons, as it mirrors individual physical limitations. The questionnaire is not interchangeable with other disease-specific instruments and has not yet been shown to be responsive to long-term disease progression.

Medical Outcomes Study Short Form-36 (SF-36)
The SF-36 is a generic health survey [63]. The patient is asked to complete 36 items of the questionnaire. The instrument allows the patient to self-assess psychic, physical, and social aspects of his or her quality of life.

Table 3: Health-status measurement instruments

<table>
<thead>
<tr>
<th>Instrument Type</th>
<th>Domains</th>
<th>Items</th>
<th>Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGRQ</td>
<td>disease-specific symptoms, activities, psychosocial impact</td>
<td>76</td>
<td>self-administered by the patient</td>
</tr>
<tr>
<td>CRQ</td>
<td>dyspnea, emotional function, fatigue, mastery</td>
<td>20</td>
<td>interview</td>
</tr>
<tr>
<td>SF-36</td>
<td>physical and social function, mental health, energy/vitality, health perception, physical and mental role limitation, pain</td>
<td>36</td>
<td>self-administered by the patient</td>
</tr>
</tbody>
</table>

SGRQ: St. George’s Respiratory Questionnaire; CRQ: Chronic Respiratory Disease Questionnaire; SF-36: Short Form 36.
**Strengths and limitations**

SF-36 is the best-known questionnaire to measure health status. The instrument has shown to be discriminative, responsive to long-term disease progression, easy to use, and has been validated in several languages. However, as a generic measure, it is considered less responsive than disease-specific instruments in COPD and is not consistently responsive to therapeutic effects. No MID has been established yet.

**Open questions and outlook**

Further development of user-friendly, inexpensive instruments to enable fast and easy health status assessment in clinical trials as well as in daily practice is clearly required. Ways to involve patients in questionnaire generation should be further explored. More information is needed on the time course of health-status alterations (e.g., induced by therapeutic intervention or secondary to COPD exacerbations) and on the utility and efficacy of health status instruments in less severe COPD.

**Exacerbations**

**Relevance**

Exacerbations of COPD indicate clinical instability and progression of the disease and are associated with increased morbidity, deterioration of comorbidities, reduced health status, physical and physiologic deterioration and an increased risk of mortality [64,65]. The prevention or reduction of exacerbations thus constitutes a major treatment goal [1].

**Measures**

Verification by patient interview, healthcare databases or prospectively from diary cards. Endpoints: frequency of exacerbations, time to first exacerbation, severity and duration of exacerbations.

**Strengths**

• The event-based approach considers the need for systemic corticosteroids and/or antibiotics or hospitalisation due to an exacerbation. This definition may be more robust and is relatively easy to record.

• The symptom-based definition of exacerbations considers individual patient's perception of clinical status.

**Limitations**

• There is no standardized definition of an exacerbation, making comparative evaluations of clinical study results difficult [1,66].

• The symptom- and event-based approach involves subjective and recall bias, particularly because patients often have a poor understanding of exacerbation symptoms, resulting in substantial underreporting of exacerbations [67].

• The definition by use of health care resources is health system specific and affected by many other factors (social support, comorbidities, baseline health status, clinical expert behaviour).

• Differential diagnoses to exacerbations such as pneumonia, heart failure, ischemic heart disease, pulmonary embolism have to be taken into account.

• Seasonal variations in exacerbation frequency usually require long-term studies of at least one year duration [4,68].

• No MID has been established yet [4].

**Open questions and outlook**

There is a clear need to standardize the evaluation of the onset, frequency, severity and duration of COPD exacerbations as well as to assess therapeutic effects on exacerbations in COPD. Given the potential clinical relevance of even single exacerbations it appears quite difficult to determine exactly what cut-off levels should be used in terms of MIDs. In addition, more work is needed to develop simple feasible criteria for defining exacerbations in clinical practice and to analyse the multiple factors that contribute to decisions to assess the severity stage of exacerbations. In that context, the EXACT-PRO initiative began to develop and evaluate a novel patient-reported outcome tool to measure the rate, duration and severity of exacerbations of COPD [69].

**Multidimensional scoring systems - BODE**

**Relevance**

So far the only multidimensional scoring system that has gained broader acceptance is the BODE index which has been developed as a prognostic marker for COPD patients in an attempt to integrate not only the respiratory but also the systemic expressions of COPD in a single grading system [70].

**Measures**

It comprises the four components nutritional state (BMI), airflow limitation (Obstruction; FEV1), breathlessness (MRC Dyspnoea scale), and Exercise capacity (6MWD, distance walked in 6 min). Replacing the 6MWD with a component for exacerbation frequency (BODEx index) resulted in fully preserved power to predict the mortality risk in a prospective observational study, while expanding the BODE index with exacerbation frequency as a fifth component (e-BODE index) did not further improve its predictive power [71]. A truncated version of the BODE index has been presented in which the exercise component is omitted (BOD index) [72].

The validity of the BODE index as a prognostic marker to predict mortality in COPD patients has recently been challenged by a study demonstrating that the risk of all-cause mortality over 3 years was considerably underestimated by the BODE index in a population of severe COPD patients, while on the contrary it was overestimated in another population with milder disease, indicating that important predictors may still be missing in this index [73]. Nevertheless, the BODE index has been used.
to assess therapeutic efficacy in interventional studies investigating effects of lung volume reduction surgery [74-76], pulmonary rehabilitation [77,78], and physical training [79], but so far not in pharmacological intervention trials.

**Strengths**
- The BODE index integrates different facets of COPD and the risks associated with significant comorbidities.
- It provides better power than that of its individual components (e.g., FEV₁) to predict mortality and future exacerbations in patient populations with severe-to-very severe COPD [70,80].
- Its assessment is straightforward.

**Limitations**
- The BODE index has not primarily been developed to assess effects of therapeutic interventions and a MID has not yet been defined.
- The BODE index has been optimized to predict one-year mortality. The factors most critically affecting short-term survival might differ from those determining survival over a longer term. Thus, its suitability for assessment of patients with mild-to-moderate COPD is as yet less validated.
- The FEV₁ categories in the airway obstruction component are not consistent with the current GOLD staging system.
- No published experience with BODE index as a clinical outcome parameter in pharmacological intervention studies is currently available.

**Open questions and outlook**
More widespread application of the BODE index as an outcome parameter in clinical trials is currently hampered by the lack of experience in pharmacological intervention studies. Furthermore, its validity as a prognostic marker in a population of patients affected by mild-to-moderate COPD and its power to predict survival over longer periods of time as yet have to be proven.

**Mortality**

**Relevance**
Long-term observations of large patient populations have shown an increased risk for all-cause mortality in COPD patients that rises proportionally to severity classes [6,8,81,82]. Mortality can be recorded as all-cause mortality and cause-specific mortality.

**Strengths**
- All-cause mortality is the most robust and reliable outcome of clinical trials in COPD and is relatively easy to follow-up [4,83,84].

**Limitations**
- Standardized methods to accurately define the cause of death (e.g. respiratory versus cardiovascular mortality) have not been established yet. Moreover, the careful analysis of the cause of death requires substantial effort.
- Retrospective mortality data may be confounded by inherent statistical bias [85,86], and even prospective studies are susceptible to bias due to missing follow-up of withdrawals [86,87].
- Adequately powered mortality trials require high patient numbers and extended study duration [3,84,88].
- It is as yet unclear, whether COPD-specific mortality is increased in patients with milder forms of COPD (GOLD stages I and II) [2,89].
- Mortality tends to be lower in participants of clinical trials than is found in routine clinical care [90].

**Open Questions and Outlook**
One important issue is the statistical approach to analyse the events of death. Intent-to-treat (ITT)-analyses, aiming for complete follow-up of deaths are recommended for unbiased comparison between treatment groups and should be used preferentially as shown in major trials [83,84,87].

For a confident, robust assessment, mortality should be the primary outcome of a prospective trial. Clinical trials evaluating death as a primary or secondary endpoint should have a data safety monitoring board and an independent adjudicating committee [3,4,91].

**Conclusion**
The understanding of the merits and limitations of current methods for assessing physiological and clinical outcomes of COPD is crucial for the interpretation and design of clinical trials. Unfortunately, in contrast to monitoring lung function, there is no gold standard for measuring symptoms such as dyspnoea, health status, exercise capacity, physical activity, or exacerbations, since none of the available methods is optimal in all regards. Accordingly, no single outcome measure can be recommended for the assessment of treatment response in COPD. More research is needed to improve and simplify questionnaire-based markers or technologies to assess outcomes such as physical activity or health status in order to enable wider use in clinical trials as well as in primary care. A further step in that direction may be the recent development of a COPD assessment test [92].

Implementation of MIDs may also help to assess which changes of outcome markers can be considered clinically relevant. However, MIDs hardly reflect the heterogeneity, variability, and severity of COPD, as well as the numerous confounding factors contributing to the clinical presentation of the disease.

Further, no biomarkers have been established yet to reflect the inflammatory and destructive process in the lung or to indicate responsiveness to treatment. However, further research in this area is important as pulmonary biomarkers - whether physiological or biochemical - are urgently needed if clinical trials are to be shorter and more discriminating than at present.
Finally, comorbid conditions such as cardiovascular disease, anxiety and depressive disorders, lung cancer and osteoporosis are often observed in COPD patients and are likely to affect COPD outcomes. The impact of these conditions together with the influences of concomitant medication on COPD are variable and for many of them still uncertain; nevertheless, they may alter COPD phenotype, disease progression and survival, and responses to treatment. A systematic evaluation of comorbidities and co-medication should be considered as part of COPD management as they may influence the results of clinical outcome measures.

Competing interests
TG was employee of Boehringer-Ingelheim at the time of manuscript submission. CV has given presentations at industry symposia sponsored by Altana, AstraZeneca, Aventis, Bayer, Boehringer-Ingelheim, Pfizer, GlaxoSmithKline, Merck Darmstadt, Talecris. He has also received consulting fees from Altana, AstraZeneca, Bayer, Boehringer-Ingelheim, Novartis, Pfizer, GlaxoSmithKline, Talecris. RB has received reimbursement for attending scientific conferences, and/or fees for speaking and/or consulting from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmith Kline, Janssen-Cilag, Novartis, Nycomed, and Pfizer. The Pulmonary Department at Mainz University Hospital received financial compensation for services performed during participation in single- and multicenter clinical phase-IIV trials organized by various pharmaceutical companies.

Authors’ contributions
TG conceived of the review, drafted and coordinated the manuscript. CV and TA conceived of the review, critically discussed and helped to draft the manuscript. All authors read and approved the final manuscript.

Author Details
1Pulmonary Department, University Hospital, Johannes Gutenberg-University, Mainz, Germany and 2Department of Pneumology, University Hospital Giessen and Marburg, Marburg, Germany

Received: 23 December 2009 Accepted: 17 June 2010 Published: 17 June 2010

References

77. Cote CG, Celli BR. Pulmonary rehabilitation and the BODE index in COPD. *Eur Respir J* 2005, 6:630-636.


Cite this article as: Glaab et al., Outcome measures in chronic obstructive pulmonary disease (COPD): strengths and limitations Respiratory Research 2010, 11:79