

Pharmaceutical Benefits Scheme

Post-market Review of

Chronic Obstructive Pulmonary Disease Medicines

ToR 2

Final Report

August 2017

Contents

Section 2: ToR 2 Review of clinically relevant outcomes	3
2.1 Key findings for ToR 2.....	3
LFA consumer views	3
Stakeholder views (Forum and Public consultation).....	4
2.2 Methodology	4
2.2.1 Identification of relevant evidence.....	4
2.2.2 Data extraction and synthesis.....	5
2.3 Data extraction and synthesis	6
2.3.1 Regulatory agencies.....	6
2.3.2 Peer-reviewed literature.....	14
2.3.3 Health technology assessment and reimbursement agencies	21
2.3.4 Clinical guidelines and clinical studies.....	25
2.3.5 Minimal clinically important difference for COPD outcomes.....	27
2.3.6 Discussion and summary	28

List of Tables

Table 2.1	Literature search criteria for ToR 2.....	5
Table 2.2	COPD outcome information from regulatory agencies	9
Table 2.3	COPD outcomes characterised in peer-reviewed literature from 2010 onwards.....	16
Table 2.4	Outcomes referenced in Public Summary Documents for PBS-listed COPD medications	23
Table 2.5	COPD outcome information from the GOLD Strategy Report	25
Table 2.6	Summary of MCID values for the main outcomes used in COPD	27

Section 2: ToR 2

Review of clinically relevant outcomes

Review the clinical outcomes that are most important or clinically relevant to people with COPD and the extent to which these outcomes are included in the evidence previously provided to PBAC on the cost-effectiveness of these medicines.

2.1 Key findings for ToR 2

- The main outcomes published in the PSDs for COPD submissions since 2002 are FEV₁, St. George's Respiratory Questionnaire (SGRQ), exacerbations, rescue medication and adverse events (AEs).
- The 2014 PSDs for glycopyrronium/indacaterol and umeclidinium/vilanterol (both LAMA/LABA combinations) reflect concern over the translation of FEV₁ into more clinically relevant measures of effect that were not reported in the submissions. Although these products were ultimately listed on the PBS after resubmissions, the PSDs indicate that these issues were not resolved.
- The literature search for published articles discussing outcomes for COPD identified three industry-funded publications that support FEV₁ as a surrogate outcome that is correlated with SGRQ and exacerbations. Authors comment that the correlation at a patient level is weak. In contrast, two industry-funded reviews found a poor correlation between FEV₁ and patient reported outcomes (PROs).
- The GOLD Strategy Report provides evidence of a weak correlation between FEV₁ and SGRQ. The document also presents evidence that there is an increase in risk of exacerbations, hospitalisation and death with worsening of airflow limitation. The document recommends an approach of combining symptomatic assessment with the patient's spirometric classification and/or risk of exacerbations, which is consistent with the above-mentioned PBAC decision-making based on FEV₁, SGRQ and exacerbations.

LFA consumer views

- The key outcome for consumers is to be able to 'breathe' and live as normal a life as possible.
- Some consumers have experienced side effects. There is acceptance that side effects are part of the course, and that the benefits of medications outweigh the potential side effects.
- Consumers reported that it was usually a trusted pharmacist or respiratory nurse who explained the risk of side effects; very rarely their GP or specialist.
- For many, diagnosis was confirmed by spirometry by a specialist in hospital, often following a severe illness.
- For further information, the LFA Consumer Research Report is available at Appendix G.

Stakeholder views (Forum and public consultations)

- Patient symptoms, including breathlessness, are not well correlated with FEV₁.
- FEV₁ has been an accepted consistent objective measure since the first COPD medication was PBS listed, enabling health technology assessment comparisons across time periods.
- The US Food and Drug Administration (FDA) requires a clinical outcome of FEV₁. Recent RCTs also assess other measures of efficacy, often via secondary endpoints, and this data is submitted to the TGA and PBAC for consideration.
- A recently published meta-regression analysis (approximately 120,000 patients) found that for every 100 mL change in pre-dose FEV₁, the hazard ratio (HR) decreased by 21% and the absolute exacerbation rate decreased by 0.06 per patient per year (Zyder et al, 2017).
- The GOLD Strategy Report (2017) ABCD (COPD patient assessment tool) uses respiratory symptoms and exacerbations alone to assign ABCD patient categories.
- COPD Assessment Test (CAT) is a questionnaire for people with COPD and more reflective of PROs. The questionnaire is designed to measure the impact of COPD on a person's life, and how this changes over time.
- Longer term follow up in comparative COPD clinical trials is required to accurately assess the prevention of exacerbations, reduction in symptoms, health-related quality-of-life (HQoL), and safety outcomes.
- For further information, the Stakeholder Forum Summary is available at Appendix F.
- Additional recent published references were provided by stakeholders. A list is available in Appendix U.

2.2 Methodology

The methodology for ToR 2 involved the identification of relevant evidence from published literature and regulatory and reimbursement agencies. Following that, relevant data were extracted from the sources and synthesised to address the issue of clinically relevant outcomes in COPD.

2.2.1 Identification of relevant evidence

The review focused on current views and advice relating to appropriate outcome measures for COPD treatment and includes English language publications from 2010 onwards.

Databases for peer-reviewed literature were searched to identify recent reviews focusing on the clinical outcome measures that should be used to assess treatment effectiveness in patients with COPD. The websites of regulatory and reimbursement agencies were also searched to identify advice relating to appropriate outcome measures for investigation of medicinal products in the management of COPD. The reference lists of relevant papers were scanned for other studies potentially missed in the searches.

Relevant information was also sourced from the literature identified for ToR 1 (clinical practice guidelines relating to pharmacological management of COPD) and ToR 3 (clinical studies reporting the efficacy and safety of COPD medicines), as well as PSDs for COPD medicines, and public submissions on the final ToR for this review.

Table 2.1 summarises the literature search criteria that was used to address ToR 2.

Table 2.1 Literature search criteria for ToR 2

Limit	Eligibility criteria
Database of peer-review literature	PubMed Cochrane Library
Other means to identify relevant information	• Search of websites of regulatory agencies: TGA, EMA, FDA. • Search of websites of HTA and reimbursement agencies: AHRQ, CADTH, NHS HTA/NCCHTA, NHS CRD, NICE, PBAC. • Clinical practice guidelines identified for ToR 1. • Clinical studies identified in literature search for ToR 3. • Scan of public consultation submissions. • Scan of reference lists of relevant reviews, primary articles, clinical and regulatory guidelines.
Publication types	• Full text reviews, trials, reports and guidelines reporting on outcome measures for COPD treatment. • English language only.
Search period	• The review will focus on current views on the most appropriate measures of benefit in the COPD population. Articles published from 2010 onwards until 31st October 2016 (PubMed) and 18th November (Cochrane) were eligible.
Exclusion criteria	• Wrong patient population: does not relate to patients with COPD or mixed airways disease (e.g. ACOS). • Wrong intervention: does not relate to pharmacological management of COPD. • Wrong outcomes: does not report outcome measures.

Source: Research Protocol, approved by RG 2nd August, 2016

Abbreviations: ACOS, asthma- COPD overlap syndrome; AHRQ, Agency for Healthcare Research and Quality; CADTH, Canadian Agency for Drugs and Technologies in Health; COPD, chronic obstructive pulmonary disease; EMA, European Medicines Agency; FDA, Food and Drug Administration; HTA, health technology assessment; NCCHTA, National Coordinating Centre for Health Technology Assessment; NHS CRD, University of York NHS Centre for Reviews and Dissemination; NHS HTA, National Health Service Health Technology Assessment (UK); NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; TGA, Therapeutic Goods Administration; ToR, Term of Reference.

2.2.2 Data extraction and synthesis

Outcomes identified from the literature search of peer-reviewed publications, regulatory agencies, HTA and reimbursement agencies, guidelines and clinical studies were summarised. Those outcomes that are surrogate measures (e.g. change in FEV₁), directly patient relevant (e.g. change in symptoms, frequency and severity of exacerbations), and/or impact on healthcare resource utilisation (e.g. primary care and specialist attendances, emergency room attendance and hospitalisations) were identified. The summary captures information, where reported, on the proposed definitions of effect sizes that are considered to be clinically important.

For each review question, the findings have been synthesised into an overall narrative on what are considered to be important or clinically relevant measures of benefit. This will be discussed in light of the outcomes presented to the PBAC for consideration of the clinical and cost-effectiveness of COPD medicines.

2.3 Data extraction and synthesis

2.3.1 Regulatory agencies

Table 2.2 refers to COPD outcome material from the European Medicines Agency (EMA, 2012) and the United States Food and Drug Administration (FDA, 2016). The document from the FDA is labelled as 'Draft Guidance' that is being distributed for comment purposes only.

The Australian TGA website was searched for analogous information; it refers to the EMA guidance in the Clinical Efficacy and Safety Guidelines section with a disclaimer.¹ These documents provide a general overview of outcomes relevant to COPD.

Lung function

Both the EMA and the FDA documents state that FEV₁ is the most widespread efficacy endpoint used to measure airflow obstruction. It is standardised, easy to perform, reproducible, and is frequently defined as the primary outcome in clinical studies. Both sources state that FEV₁ should be measured both pre- and post-bronchodilator, with the post-bronchodilator measurements to be used for assessment of a new bronchodilator medicine. From these measurements, a time profile curve can be developed to estimate the time-to-effect and duration-of-effect.

The FDA document states that 'For a non-bronchodilator drug, the use of lung function test parameters, such as FEV₁, as a marker of disease status has become validated as a surrogate endpoint through years of clinical and regulatory experience, and is commonly used and accepted as an endpoint to support efficacy.'

Exacerbations

Both the EMA and the FDA documents state that exacerbations are clinically relevant outcomes that are related to morbidity and mortality. The exacerbation measurement should be clinically meaningful in order to determine when worsening of symptoms becomes an exacerbation. The exacerbations should be clearly defined in terms of criteria such as worsening of shortness of breath, increased sputum volume, increased purulence of sputum, changes in treatment, and hospitalisation. Measurements can include duration of exacerbation, severity (mild, moderate, severe), delay in occurrence, time to first exacerbation, or reduction in frequency.

¹ Where European Union (EU) guidelines adopted in Australia include references to EU legislation (including EC Directives and Regulations), the requirements contained in the referenced EU legislation are not applicable to the evaluation of prescription medicines by the TGA. The Australian legislative requirements applying to prescription medicines are contained in the Therapeutic Goods Act 1989 and the Therapeutic Goods Regulations 1990, as well as in various legislative instruments such as Therapeutic Goods Orders, Notices and Determinations, see Legislation.

Patients and investigators' reported outcomes

Health-related quality-of-life (HQoL)

Both the EMA and the FDA refer to the Chronic Respiratory Questionnaire (CRQ) and the St George's Respiratory Questionnaire (SGRQ), which are both multidimensional disease-specific COPD questionnaires that cover different health-related domains.

The FDA document refers to the limitations of these instruments, in that they may be insufficient to determine a treatment effect of a drug narrowly targeted to a specific aspect of COPD. An Appendix to the FDA document discusses the SGRQ in detail, in terms of administration, scoring, components and minimal clinically important difference (MCID), which is determined to be at least four units on the SGRQ scale.

The EMA document states that disease-specific questionnaires are more sensitive to disease changes and better suited to COPD than generic instruments, although general questionnaires such as Short Form-36 (SF-36) and Extended Activity of Daily Living (EADL) are mentioned.

Symptom scales

According to the EMA and quoted guidelines, the three fundamental symptoms of COPD are dyspnoea, sputum production, and cough. The document states that patient diaries should be used to evaluate the symptoms, according to categories such as night time symptoms, night time awakening, daytime symptoms, cough, wheezing, dyspnoea, and sputum production. The FDA document states that patients can evaluate specific symptoms on a categorical, visual, or numerical scale. The FDA warns that, although symptom scores can be valuable for assessing efficacy of a drug aimed at relieving a specific symptom, symptom scores as the sole or primary measure of efficacy in COPD are discouraged because of their subjective nature, precision issues, and lack of standardisation.

Dyspnoea or activity scales

The instruments mentioned by the EMA and the FDA include the Baseline Dyspnoea indices/transition dyspnoea indices (BDI/TDI), the Borg Scale and the Medical Research Council (MRC) Dyspnoea Score. The FDA warns of many reasons why these types of scales make them unsuitable for use as the sole or primary evidence of efficacy, including that they were not specifically developed for drug studies, that the results may be difficult to interpret in terms of clinical significance, and they may rely on patient recall.

Exercise capacity

Both the EMA and the FDA refer to the 6-Minute Walk Test (6-MWT) and the Shuttle Walk as widely used in clinical studies for evaluating exercise capacity. They also refer to treadmill walking or cycle ergometry as endurance tests, combined with lung volume assessment. However, the FDA refers to the limitations of these tests with reference to precision, standardisation, and consistency, which may limit the sensitivity of these measures.

Rescue medication

The EMA refers to the use of rescue medication as a clinical endpoint, in terms of the number of times in a given period that rescue medication is required, and the number of puffs.

Composite scores

The EMA refers to the Body-mass index, airflow Obstruction, Dyspnea, and Exercise – Index (BODE-Index) as a composite outcome used in COPD. It is made up of body mass index (BMI), airflow obstruction measured by FEV₁, dyspnoea measured by MRC Dyspnoea Scale, and exercise capacity measured by the 6-MWT.

Imaging, biomarkers and surrogate endpoints

Both the EMA and the FDA refer to computed tomography (CT) imaging as a possible outcome in COPD in terms of measuring the progression of emphysema and the evaluation of airway wall thickening, although the EMA document states that it is not a fully validated technique in COPD and should not be relied upon as a primary outcome. The FDA document states that these surrogate outcomes can be considered supportive evidence [only], and list biomarkers additional to CT, such as concentration of certain gases in exhaled air or breath condensate, inflammatory mediators or cells in relevant biological fluids, and sensitive measures of airflow based on imaging of radio-labeled gases.

Secondary efficacy outcomes

The EMA states that physical activity and biomarkers of systemic inflammation could be used as secondary outcomes. The FDA discusses that all the above-mentioned outcomes such as exercise capacity, symptom scores, activity scales, and HRQoL instruments could be used as supportive measures.

Table 2.2 COPD outcome information from regulatory agencies

EMA ^{a,b}	FDA ^c
Reference	
Guideline on clinical investigation of medicinal products in the treatment of chronic obstructive pulmonary disease (COPD). Respiratory Drafting Group. European Medicines Agency 2012.	Chronic obstructive pulmonary disease: Developing drugs for treatment. Guidance for industry. DRAFT GUIDANCE – This guidance document is being distributed for comment purposes only. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). May 2016. Clinical/Medical Revision 1.
Lung function	
Changes in spirometric parameters should be measured as a relevant part of the overall effect of any new therapy in the treatment of patients with COPD. Spirometry should be undertaken by trained healthcare professionals according to standardised methods. FEV₁ is the most extensively used parameter for adopting treatment strategies in COPD. FEV₁ is one of the most repeatable lung function parameters and in COPD is a measure of the obstructive element of the disease. If FEV₁ is the primary endpoint, the pre-bronchodilator FEV₁ is the preferred measure in the development of a new product for maintenance treatment, although depending on the mode of action, other lung function parameters (e.g. post-bronchodilator FEV₁) could be the parameter of choice. Whether the preferred parameter is the pre-bronchodilator FEV₁ or the post-bronchodilator FEV₁ or other lung function parameters, this should be justified. It is recommended that FEV₁ is measured both pre- and post-bronchodilator, both at baseline and at repeated visits during each study treatment period. For a bronchodilator serial post-dose FEV₁ measurements should be carried out to characterise the time profile in order to determine time-to-effect and duration-of-effect, particularly in Phase II studies. The maintenance of the effect over time for any drug with an effect on lung function should also be assessed. A central quality assurance system is highly encouraged. The classification of lung function values as 'valid' or 'invalid' should be pre-specified and scientifically justified in the protocol according to acceptable standards. It should be stated and justified how valid or invalid measurements will be used in the study analysis. A description of the quality achieved during spirometric testing should be provided in the study report by means of generally accepted parameters. Other measures of lung function which could also be recorded to characterise the effect of a new active substance include IC, FRC, RV/TLC, FVC and slow VC and DLCO. Some of these measures of lung function may correlate better with improvements in symptoms and exercise tolerance than does FEV₁. They might be considered as appropriate alternative physiological endpoints if validated for use in COPD. Slow VC is preferred to FVC in some cases of severe airflow obstruction, particularly emphysema.	<u>Efficacy assessment</u> Pulmonary function testing by spirometry can be a useful way to assess airflow obstruction and, therefore, can be a useful tool to assess efficacy of a COPD treatment. FEV₁ obtained from typical spirometry is commonly used as an efficacy endpoint because FEV₁ is a reflection of the extent of airway obstruction. Spirometry is also well standardised, easy to perform, and when conducted appropriately gives consistent, reproducible results across different pulmonary function laboratories. Air-trapping and hyperinflation are common features in COPD, particularly in the emphysematous-type, and are reflected in parameters of lung function testing, such as an elevation in the residual volume to total lung capacity ratio. Hyperinflation is believed to be responsible, at least in part, for the sensation of dyspnoea. The division does not have a great deal of regulatory experience in the use of parameters of lung function other than spirometric measures in therapeutic approvals, but is open to considering alternative assessments. These alternatives should be discussed with the division early in drug development. <u>Improving airflow obstruction</u> The primary efficacy endpoint should be change in post-dose FEV₁ for a bronchodilator (e.g. a new beta-adrenergic agent or a new anticholinergic agent) and change in pre-dose FEV₁ for a non-bronchodilator. A bronchodilator drug may improve the FEV₁ from a direct effect on the airway smooth muscle, and a non-bronchodilator drug may improve the FEV₁ by other mechanisms such as reduction of airway inflammation. For a bronchodilator drug, serial post-dose FEV₁ assessments should be performed to characterise a time profile curve that will help in the estimation of time-to-effect and duration-of-effect. Assessments of post-dose FEV₁ for a bronchodilator drug and pre-dose FEV₁ for a non-bronchodilator drug should be performed periodically over the duration of the study to ensure that the beneficial effect is sustained over time. <u>Altering disease progression</u> A preferred primary efficacy endpoint is the serial measurement of FEV₁ over time, with the expectation that the FEV₁ decline slopes will diverge in favour of active treatment (i.e. airflow is preserved relative to the comparator). When the claim is alteration of disease progression, such divergence should exclude the possibility of parallel declines in FEV₁ with the active treatment offset by an initial and sustained bronchodilator effect. This latter circumstance may still be one in which a drug approval is possible (e.g. for a bronchodilation claim), but would not be appropriate for supporting a claim of altering disease progression.
Exacerbations	
Definitions of exacerbation and severity of the exacerbation need to be standardised to allow comparisons between different interventions in different settings. The proposed definition of an exacerbation of COPD is an acute event characterised by a worsening of the patient's respiratory symptoms that is beyond normal day-to-day variations and leads to a change in	<u>Modifying or preventing exacerbations</u> The primary efficacy endpoint should be a clinically meaningful measure of exacerbations. Such measures can include the duration of exacerbations, severity of exacerbations, delay in the occurrence of an exacerbation, or reduction in the frequency of exacerbations. If one of these measures is chosen as the

EMA ^{a,b}	FDA ^c
medication (GOLD 2011). Although criteria for medical interventions might be subject to local differences, the following classification of the severity of exacerbations is recommended for stable COPD patients: • Mild: exacerbations described as an increase in respiratory symptoms that can be controlled by the patient with an increase in usual medication; • Moderate: exacerbations that require treatment with systemic corticosteroids and/or antibiotics; • Severe: exacerbations that require hospitalisation or result in death. The rate of moderate or severe exacerbations is a clinically relevant endpoint related to the associated morbidity and mortality and the usually significantly increased health-care requirement. The frequency and/or severity of exacerbations are important outcome measures that should be considered in clinical studies in COPD. Such measures can include reduction in the number of exacerbations, annual rate and severity of exacerbations. Time to first exacerbation might also be considered. If one of these measures is chosen as the primary efficacy endpoint, the others should be assessed also to ensure that improvement in one endpoint does not result in worsening in another. An evaluation of the frequency of exacerbations should normally be made over a period of at least one year due to seasonal variation in exacerbation rates. The timing of the study treatment may prove important (e.g. capturing the winter cold season in the majority of patients). There should be an established minimum time interval between exacerbations to consider them as different episodes. The end of an exacerbation has to be defined clearly so that a difference between the existing exacerbation and a new exacerbation can be measured. Evaluation by an external adjudication committee is encouraged.	primary efficacy endpoint, the others also should be assessed to ensure that some other measure has not worsened. For instance, a delay in occurrence of a first exacerbation would not be clinically meaningful if the end result were more frequent exacerbations over a longer period of assessment. The protocol should define exacerbations in a way that is clinically meaningful, and specify criteria to determine when worsening of symptoms become an exacerbation. Criteria to consider in defining exacerbation include worsening of shortness of breath, increased sputum volume, increased purulence of sputum, worsening in symptoms requiring changes in treatment, or worsening of symptoms requiring urgent treatment or hospitalisation. Because exacerbations are often associated with precipitous falls in airflow, the rapidity of recovery of a pulmonary function measure, such as FEV₁, following an exacerbation to pre-exacerbation status also can be considered a reasonable primary efficacy endpoint.
Patients' and investigators' reported outcomes	
<u>Health status and HRQoL</u> The impact of disease on a patient's daily life, activity and well-being should be assessed at regular intervals. There is a wide range of questionnaires available. Disease-specific questionnaires (e.g. the CRQ and the SGRQ) cover different health-related domains. Disease-specific instruments tend to be more sensitive to changes and therefore better suited to measure treatment effects in COPD than generic instruments. More recently new tools have been introduced into clinical trials. Among them, the CAT, a patient and clinician rating scale, deserves some consideration for its easy management and its good correlation with the SGRQ. General questionnaires (e.g. SF-36) and questionnaires with a narrower perspective such as the activity of daily living questionnaires (Nottingham EADL or London EADL) or the functional status questionnaires can also provide relevant information, focusing on the number of activities that a patient can perform. Other health-related questionnaires, specific or generic, can be utilised if sufficiently validated and extensively used.	<u>HRQoL instruments</u> HRQoL, such as the SGRQ and the CRQ, are designed to systematically assess many different aspects of the effect of COPD on a patient's life. These instruments can be used to assess efficacy of a drug, but they have some limitations. These instruments are multidimensional and assess various effects of the disease on a patient's life and health status. Therefore, these instruments may be insufficient to determine a treatment effect in cases of a drug narrowly targeted to a specific, but clinically meaningful, aspect of COPD. When they are used to assess efficacy in the setting of multinational trials, the instruments should be validated for all languages and cultures in which the studies are conducted.
<u>COPD symptom scales</u> According to widely accepted COPD treatment guidelines, (ERS, ATS, GOLD), the three cardinal symptoms of COPD are dyspnoea, sputum production and cough. The symptoms can be evaluated over the course of the clinical study by use of patient diaries. Improvements in these symptoms are to be expected with most drugs, but the magnitude of improvement is difficult to estimate and a clinically relevant standard for improvement has not yet been established. This needs to be discussed on a study by study basis. Symptoms to be	<u>Symptom scores</u> Symptom scores determined by asking patients to evaluate specific symptoms on a categorical, visual, or numerical scale can be a simple way to assess efficacy of a drug based on the patient's own assessment of health status. Symptom scores can be valuable for assessing efficacy of a drug specifically aimed at relieving a symptom. In clinical programs aimed at other aspects of COPD, patient-reported symptom scores can be useful in assessing secondary effects of the therapy and may provide important additional evidence

EMA ^{a,b}	FDA ^c
recorded should include – night time symptoms, night time awakening, daytime symptoms, cough, wheezing, dyspnoea, sputum production, etc.	of efficacy. Symptom scores as the sole measure or primary measure of efficacy in COPD are discouraged because of their subjective nature, precision issues, and lack of standardisation. If a symptom score is used, particularly a novel scoring, the issue of validation of the scoring should be addressed. Providing symptom relief The primary efficacy endpoint should reflect the claimed clinical benefit (e.g., a drug intended to reduce cough should show that effect through assessments of coughing, subjectively and/or objectively measured). The selected primary efficacy endpoint should be clinically meaningful, and the magnitude of improvement that is proposed to be shown should be clinically relevant. In addition, if the action of the drug targets the underlying process, but manifests as symptom relief, secondary endpoints should assess other aspects of the drug's effects (e.g., measures of lung function, airflow, sputum production).
<u>Dyspnoea</u> Instruments used to measure dyspnoea should rely on patient-reported outcomes and be multidimensional whenever possible. Dyspnoea can be measured using clinical ratings based on activities of daily living and ratings during an exercise task. The BDI/TDI (clinician-rated scales) and the dyspnoea component of the CRQ (patient-rated scale) are examples of clinical ratings extensively used in randomised controlled trials. The BDI/TDI is a validated instrument developed to measure the impact of dyspnoea on three domains – functional impairment, magnitude of task and magnitude of effort. Alternatively, there are a number of methods for patients with COPD to rate their dyspnoea during an exercise test such as cycle ergometry or treadmill walking. The two more commonly used methods are the Borg CR10, which is preferred, and VAS.	<u>Activity scales</u> Activity scales such as the MRC Dyspnoea Scale, the Borg Scale, and the Mahler BDI/TDI can be used as supportive of efficacy. These scales are relatively simple to administer, but they have limitations that make them unsuitable for use as the sole or primary evidence of efficacy and for supporting specific labelling claims. These scales were not specifically developed for use in clinical studies of drugs and their attributes in longitudinal interventional settings may not be fully elucidated. Also, the results can be difficult to interpret in terms of levels of clinical significance, because for some of these scales the minimal important difference has not been identified and validated. Scales that are third-party rated (e.g., Mahler's dyspnoea indices) may prove less compelling than validated patient-rated instruments, because third-party assessments have been shown in some circumstances to be less reflective of patient status than first-party assessments. In addition, scales that require patients to recall prior symptoms (e.g., how do you feel now compared to baseline?) are problematic, because patients' memories may fade over time, particularly in studies lasting several months.
<u>Patients' questionnaires or diary cards</u> Questionnaires or diary cards should be provided, one for the patient to capture the unreported exacerbations (mild exacerbations) and another for the investigator to collect the reported (moderate-severe) exacerbations. Diary entries may be entered into an electronic diary which, in addition to recording exacerbation data, may also capture symptoms.	N/A
Exercise capacity In patients with COPD exercise testing is useful in the clinical setting to assess the degree of impairment, prognosis and the effects of interventions. Several methods for evaluating exercise capacity have been developed. The severity and cause of exercise intolerance are best assessed by conducting standardised laboratory exercise testing in which detailed physiological respiratory/metabolic measurements are made while patients perform cycle ergometry or walk at a specific speed on the treadmill. Laboratory test protocols can be either constant ('endurance') or incremental work rate tests. Endurance tests rather than incremental testing have been more extensively used in COPD. Cycle and treadmill exercise have been used interchangeably although the former has been used more commonly in clinical studies in COPD, as the work rate for endurance and incremental tests is easier to quantify. Simpler tests can also be used, although the information gathered is more limited. The 6-MWD is a relatively simple test that has been used extensively in studies to evaluate possible benefits of pharmacological intervention; the shuttle walking test, a better standardised and simpler field test, is also widely used.	Reduced capacity for exercise is a typical consequence of airflow obstruction in COPD patients, particularly because of dynamic hyperinflation occurring during exercise. Assessment of exercise capacity by treadmill or cycle ergometry combined with lung volume assessment potentially can be a tool to assess efficacy of a drug. Alternate assessments of exercise capacity, such as the 6-MW or Shuttle Walk, also can be used. However, all these assessments have limitations. For instance, the 6-MWT reflects not only physiological capacity for exercise, but also psychological motivation. Some of these assessments are not rigorously precise and may prove difficult in standardising and garnering consistent results over time. These factors may limit the sensitivity of these measures and, therefore, limit their utility as efficacy endpoints, because true, but small, clinical benefits may be obscured by measurement noise.

EMA ^{a,b}	FDA ^c
Rescue medication	
The use of rescue medication (e.g. β_2 agonist, reliever inhaler) reflects effects on symptoms and therefore can be considered as a clinical endpoint. Both the number of times that rescue medication is required during the day and at night and the number of puffs used on each occasion should be recorded; the number of times that rescue medication is used is the more relevant measure.	NR
Composite scores	
Changes in the BODE-Index are considered of interest. The BODE-Index is a composite index based on BMI, airflow obstruction as measured by FEV ₁ , dyspnoea assessed by the MRC Dyspnoea Scale, and exercise capacity measured by the 6-MWT. Other composite scores might be used if validated and generally accepted.	NR
Imaging	
CT imaging can accurately characterise lung parenchymal changes and facilitate quantitative assessment. Although in clinical practice plain radiography still has an important role in the evaluation of COPD, CT densitometric evaluation might have a role in the assessment of the progression of emphysema and the evaluation of airway wall thickening. As yet the use of CT imaging is not fully validated and therefore is not appropriate for use in clinical studies as a primary or important secondary endpoint. However to explore the possible role that CT imaging might have in clinical studies in COPD, its inclusion as a secondary endpoint should be considered. Other important considerations when using CT imaging concern the total exposure to radiation. If changes in lung structure are to be assessed it should be demonstrated that the observed changes in lung tissue are linked to functional changes which provide clinically meaningful benefit to the patient.	<u>Biomarkers and surrogate endpoints</u> With the exception of lung function tests, there are no well-validated biomarkers or surrogate endpoints that can be used to establish efficacy of a drug for COPD. For a non-bronchodilator drug, the use of lung function test parameters, such as FEV ₁ , as a marker of disease status has become validated as a surrogate endpoint through years of clinical and regulatory experience, and is commonly used and accepted as an endpoint to support efficacy. There are many biomarkers that can be considered for use in clinical studies. Some of these biomarkers include sensitive radiological evaluation of lung tissue structure (such as high-resolution chest CT), concentration of certain gases in exhaled air or breath condensate, inflammatory mediators or cells in relevant biological fluids, and sensitive measures of airflow based on imaging of radiolabeled gases. With the possible exception of the high-resolution CT, none of these biomarkers are sufficiently validated to date for use as the primary evidence of efficacy or for supporting specific labelling claims. Some of the biomarkers may be technically challenging to perform or present important additional considerations (e.g., total X-ray dose exposure in patients subjected to multiple serial CT scans). These biomarkers and surrogates can be considered as supportive of the drug's putative mechanism of action. If proposed as primary assessments of efficacy, discussions with the division early on in development would be useful to allow for earlier phase studies to not only test the drug, but help establish validity of the measure itself. A single study should not be used to establish both the validity of a novel primary endpoint and the efficacy of the drug in question. <u>Modifying lung structure</u> The primary efficacy endpoint can be a sensitive radiological assessment of lung structure with supportive evidence that the regenerated lung tissue is functional and that the treatment provides clinically meaningful benefit to patients.
Secondary efficacy outcomes	
Physical activity should be considered as a potential secondary endpoint. The value of biomarkers of systemic inflammation in COPD is not yet established. To explore the possible role that they might have in clinical studies in COPD, their inclusion as secondary endpoints should be considered.	Secondary efficacy endpoints can provide useful information on the effect of the treatment and should be selected to provide support to the primary efficacy endpoint. Secondary efficacy endpoints also can explore other effects of the drug on the disease. Commonly used secondary efficacy endpoints include various measures of lung function, exercise capacity, symptom scores, activity scales, and HRQoL instruments. Biomarkers can, in some cases, also provide support of efficacy. For some efficacy measures, such as symptom scores, activity scales, and disease-specific, HRQoL instruments, the threshold that defines a

EMA ^{a,b}	FDA ^c
	clinically meaningful improvement may not be well defined for use in clinical studies that test new drugs. Having such a benchmark of effect would be important in interpreting the meaning of differences shown in the clinical trials. Therefore, the protocol should define minimal clinically important difference with appropriate reasoning and justification. Consideration also should be given to the added complexity of the use of these measures in clinical studies for drugs, such as comparisons to baseline, comparisons to placebo, multiplicity, missing data, and the effect of study duration (e.g., recall of baseline status over time). In studies where an objective measure is used as an endpoint, such as FEV₁, use of subjective measures as important secondary assessments may be particularly useful in judging the value of mean changes in the primary endpoint. Similarly, in treatments intended to affect subjective perceptions of the disease through an effect on the underlying pathophysiology of COPD, secondary objective measures also can provide useful additional assessments to support the efficacy of the drug.

Abbreviations: ATS, American Thoracic Society; BDI, Baseline Dyspnoea Indices; BMI, body mass index; BODE, Body-mass index, airflow Obstruction, Dyspnea, and Exercise; CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; CR10, Category Rating Dyspnea Score; CRQ, Chronic Respiratory Questionnaire; CT, computed tomography; DLCO, diffusing capacity of the lung for carbon monoxide; EADL, Extended Activity of Daily Living; EMA, European Medicines Agency; ERS, European Respiratory Society; FDA, Food and Drug Administration; FRC, functional residual capacity; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; HRQoL, Health-Related Quality of Life; IC, inspiratory capacity; MCID, minimal clinically important difference; MRC, Medical Research Council; 6-MWD or 6-MWT, six-minute walking distance or test; N/A, not applicable; NR, not reported; RV/TLC, residual volume/total lung capacity; SF-36, Short Form – 36; SGRQ, St George's Respiratory Questionnaire; TDI, Transition Dyspnoea Indices; TGA, Therapeutic Goods Administration; VAS, visual analogue scales; VC, vital capacity.

^a From [EMA Guideline on clinical investigation of medicinal products in the treatment of COPD](#)

^b The TGA website refers to the EMA guidance in the [Clinical Efficacy and Safety Guidelines](#) section.

^c From US FDA, [COPD :Developing Drugs for Treatment Guidance for Industry](#).

2.3.2 *Peer-reviewed literature*

The literature search in Medline (PubMed) and the Cochrane Library yielded 66 published reviews that were considered for inclusion (62 from PubMed and four from the Cochrane Library). Of these, 18 references were identified for inclusion. Relevant articles from the literature search performed for ToR 3 and from hand searching of reference lists were also included. Table 2.3 shows a summary of COPD outcomes characterised in peer-reviewed literature from 2010 onwards.

The literature summary shown in Table 2.3 indicates that the most commonly reported outcomes in COPD are the following:

- Lung function: FEV₁
- COPD exacerbations
- HRQoL: SGRQ and CRQ
- Symptoms (dyspnoea): BDI/TDI.

Martin et al (2016) reports that there is a significant association between improvements in either FEV₁ and SGRQ score, and lower risk for COPD exacerbations. The authors performed a systematic literature review and regression analysis involving 67 RCTs that reported the relevant COPD endpoints. Of note, the authors found the relationship between FEV₁ and SGRQ score and hospitalisations was less clear, and required further research. The authors argue that their study suggests changes in FEV₁ and SGRQ might serve as reliable surrogate markers of patients' likelihood of experiencing an exacerbation, in the context of few trials of COPD drugs being powered to identify a significant difference in the reduced risk of exacerbations. Since the analysis used aggregated patient data from published trials, the authors point out that any statistical association at the population level cannot be translated into patient-level associations.

The authors of an earlier article (Jones et al, 2011) correlated changes in lung function with patient outcomes (TDI, SGRQ and exacerbations) using a pooled analysis of three indacaterol studies. The analysis demonstrated that in individual subjects, change in FEV₁ is a significant, albeit relatively weak predictor of improvement in PROs. However, the analysis also showed that, at a population level, improvements in FEV₁ with long-acting bronchodilator therapy are strongly correlated with improvements in dyspnoea, health status and exacerbations. The authors concluded that interventions that significantly improve FEV₁ are also likely to be associated with improved clinical outcomes and PROs.

Westwood et al (2011) demonstrated a relationship between improved lung function (FEV₁) and improvements in health status (SGRQ) in patients who are treated with long-acting inhaled bronchodilators. Results of regression modelling indicated that a 100 mL increase in FEV₁ was associated with a reduction in SGRQ total score of 2.5 units, equating to a clinically meaningful reduction of 4 units in SGRQ being associated with an estimated improvement in FEV₁ of 160.6 mL. The authors state that these results were supported by correlation analyses

that demonstrated a moderate negative correlation between change in total SGRQ score and change in trough FEV₁, when all treatment arms were considered.

The three above-mentioned studies (Martin et al, 2016; Jones et al, 2011; Westwood et al, 2011) were all funded by the Novartis Pharmaceutical Company. Novartis is the sponsor for the PBS-listed medicines glycopyrronium (Seebri Breezhaler), indacaterol (Onbrez), and indacaterol/glycopyrronium FDC (Ultibro Breezhaler). The PSDs for all three products (see Table 0.4 in Section 2.3.3) indicate that they were recommended for PBS listing based on FEV₁ alone.

As pointed out by the public submission from the LFA and TSANZ for this COPD review, two articles that acknowledge the poor correlation of FEV₁ with PROs are authored by Jones et al (2012) and Vestbo et al (2014). The former publication (also funded by Novartis) states that 'FEV₁ is a relatively poor correlate of symptoms such as breathlessness and the impact of COPD on daily life'. This publication provides a comprehensive review of PROs that are relevant to COPD, with a focus on the SGRQ and the CRQ. The latter publication (associated with GlaxoSmithKline) describes the ECLIPSE study (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints). This study measured clinical parameters, lung function, exercise tolerance, biomarkers, and amount of emphysema by computed tomography. The authors report that they found a striking heterogeneity among patients with COPD, with poor correlations between FEV₁, symptoms, quality of life (QoL), functional outcomes, and biomarkers.

A summary of the recent literature published around outcomes for COPD is as follows:

- PROs are required to supplement FEV₁, because of its surrogate nature (Jones et al, 2011; Cazzola et al, 2015; van der Molen et al, 2012).
- There is no gold standard for measuring COPD symptoms, as none of the available tools are individually ideal (Glaab et al, 2010).
- PRO measurements, such as for dyspnoea or functional status, may be particularly useful to assess a collective improvement in COPD patients who may be heterogeneous in terms of presentation, severity and progression (Cazzola et al, 2015).
- There should be a consensus in guidelines as to which PROs should be used routinely (Jones et al, 2012; Singh et al, 2014).
- Although there are some non-disease-specific HRQoL outcomes that are used for COPD, there are many disease-specific COPD tools that are also available and validated (Singer et al, 2012: SGRQ, CRQ, University of California San Diego Shortness of Breath Questionnaire [UCSD-SOBQ], modified MRC dyspnoea, BDI/TDI).
- Some PRO outcomes that have recently been validated include the Night time Symptoms of COPD instrument (NiSCI; Hareendran et al, 2013), Capacity of Daily Living during the Morning questionnaire (CDLM) and Global Chest Symptoms Questionnaire (GCSQ) (Partridge et al, 2010), and Shortness of Breath with Daily Activities (SOBDA; Tabberer et al, 2015).
- A composite outcome that has recently been studied is called 'Clinical Deterioration in COPD' (CID; Singh et al, 2016); it is a composite of FEV₁, SGRQ and exacerbation.

Table 2.3 COPD outcomes characterised in peer-reviewed literature from 2010 onwards

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
Cazzola 2015		
Frequently used outcomes Health status and HRQoL: - St George's Respiratory Questionnaire (SGRQ), St George's Respiratory Questionnaire for COPD (SGRQ-C) - Chronic Respiratory Disease Questionnaire (CRDQ), Short-Form Chronic Respiratory Disease Questionnaire (SF-CRDQ) - Clinical COPD Questionnaire (CCQ) - COPD Assessment Test (CAT). Symptom diary measures: - Exacerbations of Chronic Pulmonary Disease tool (EXACT) – PRO/EXACT – Respiratory Symptoms - Breathlessness, Cough and Sputum Scale. Breathlessness/dyspnoea: - Baseline Dyspnoea Index/Transition Dyspnoea Index (BDI/TDI) - Medical Research Council (MRC) Dyspnoea Scale, modified – Medical Research Council (mMRC) Dyspnoea Scale. Emerging measures - Mgill COPD QoL questionnaire - Visual Simplified Respiratory Questionnaire (VSRQ) - Dyspnoea-12 - Dyspnoea Management Questionnaire Computer Adaptive Test (DMQ-CAT) - Shortness of Breath with Daily Activities (SOBDA) Questionnaire - Global Chest Symptoms Questionnaire (GCSQ) - Capacity of Daily Living during the Morning questionnaire (CDLM) - Living with COPD (LCOPD).	1. Review. 2. Comprehensive evaluation of psychometric properties of available PRO instruments and the ability of each of them to characterise pharmaceutical treatment effects from published clinical trial evidence.	PRO measurements of dyspnoea or functional status provide insights into the effects of treatment on everyday life by reflecting whether or not patients perceive improvement in their symptoms or their abilities to perform daily activities, regardless of whether FEV ₁ has improved or not. This feature may be particularly useful when a treatment has multiple beneficial effects, which individually may be too small to register as a change on an assessment of an individual parameter but collectively may produce improvement. The CCQ, EXACT-PRO/E-RS, and the CAT appear to be the most promising instruments.
Ekstrom 2015		
Breathlessness Unidimensional: - Exacerbations of Chronic Pulmonary Disease tool (EXACT) – Respiratory Symptoms (E-RS) 100-mm Visual Analogue Scale (VAS) - Numeric rating scale (NRS) - Shortness of Breath with Daily Activities (SOBDA) Questionnaire. Multidimensional: - Dyspnoea-12 - Multidimensional Dyspnoea Profile (MDP).	1. Review. 2. Tools and MCIDs for breathlessness.	There have been substantial developments in instruments able to provide reliable and valid unidimensional and multidimensional measurement of self-reported breathlessness and in the understanding of the MCID for chronic breathlessness. Routine use of agreed outcome measures in clinical practice and research are crucial steps to improve our understanding of the science of breathlessness and its impact on patients' outcomes. Validated tools are now available for: - A symptom diary of daily life (E-RS) and the SOBDA in stable COPD, - Breathlessness not related to activity as a multidimensional summary score (Dyspnoea-12), or - Measuring dimensions separately (MDP). MCIDs are available for the intensity of chronic refractory breathlessness on a VAS or NRS, E-RS, and SOBDA. The MCID needs to be established for the multidimensional Dyspnea-12 and MDP instruments.
Garrow 2015		
Sleep disorders Validated in COPD: - COPD and Asthma Sleep Impact Scale (CASIS). Not validated in COPD: - Epworth Sleepiness Scale (ESS) - Pittsburgh Sleep Quality Index (PSQI).	1. Systematic review. 2. Review of the literature to identify disease specific and non-disease-specific sleep PROs that have been validated for use in COPD patients. Also	The results highlight a need for existing non-disease-specific instruments to be validated in COPD populations and also a need for new disease-specific measures to assess the impact of sleep problems in COPD.

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
	examined the psychometric properties of identified sleep outcome measures and extracted point and variability estimates of sleep instruments used in COPD studies.	
Glaab 2010		
All outcomes • Lung function (FEV₁) • Lung volumes (total lung capacity, functional residual capacity, residual volume, inspiratory capacity) • Exercise capacity and physical activity (6-Minute Walk Test (6-MWT), Shuttle Walk Test (SWT), bicycle ergometer, treadmill, accelerometer) • Dyspnoea (Baseline Dyspnoea Index/Transition Dyspnoea Index (BDI/TDI), Medical Research Council (MRC) Scale, Borg Scale) • Health status (St George's Respiratory Questionnaire (SGRQ), Chronic Respiratory Disease Questionnaire (CRDQ), SF-36) • Exacerbations (frequency of exacerbations, time to first exacerbation, severity and duration of exacerbations) • Multidimensional scoring systems: BMI, airflow Obstruction, Dyspnoea, and Exercise (BODE) Index • Mortality.	1. Review. 2. Strengths and limitations of outcome measures in COPD.	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. Disclosure: the first author (Glaab) was an employee of Boehringer Ingelheim Pharmaceutical Company at the time of submission of the manuscript.
Hareendran 2013		
New PRO instrument • Night time symptoms of COPD Instrument (NiSCI).	1. Article. 2. Development and validation of a PRO for night time symptoms of COPD.	There is currently no standardised way to measure night time symptoms in clinical studies to examine potential treatment benefit. This study documents evidence of content validity for the NiSCI, a new PRO instrument to evaluate COPD night time symptoms. Future research will involve item reduction, evaluation of the instrument's psychometric properties, and exploration of additional impacts of night time symptoms of COPD.
Jones 2011		
• FEV₁ • Transition Dyspnoea Index (TDI) • St George's Respiratory Questionnaire (SGRQ) • Exacerbation frequency.	1. Article. 2. Correlating changes in lung function with patient outcomes.	It is commonly stated that spirometry does not fully capture the impact of COPD on a patient's health. Our analysis of a large cohort of patients has demonstrated that in individual subjects, change in FEV₁ is a significant, albeit relatively weak predictor of improvement in PROs. However, the current analysis also shows that, at a population level, improvements in FEV₁ with long-acting bronchodilator therapy are strongly correlated with improvements in dyspnoea, health status and exacerbations. This suggests that interventions which significantly improve FEV₁ are also likely to be associated with improved clinical and patient-reported outcomes. Disclosure: funded by Novartis Pharmaceutical Company.
Jones 2012		
Breathlessness Most commonly used: • Medical Research Council (MRC) Dyspnoea Scale • Modified Borg Scale • Transition Dyspnoea Index (TDI) • Self-Administered Computerised Transition Dyspnoea Index (SAC TDI)	1. Review. 2. Overview of PRO concepts in COPD to accompany spirometry, such as breathlessness, physical functioning, and health status, and	Ideally, there should be a set of standardised comprehensive PRO instruments that are approved by regulatory authorities and used consistently during drug development and research. Health status questionnaires, such as the SGRQ and the CRQ, provide a comprehensive

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
<u>Less frequently used:</u> - Chronic Respiratory Disease Questionnaire (CRDQ) – dyspnoea component - University of California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) <u>More recent scales:</u> - Shortness of Breath with Daily Activities Questionnaire (SOBDA) - Global Chest Symptoms Questionnaire (GCSQ) - Dyspnoea-12 - Dyspnea Management Questionnaire Computer Adaptive Test (DMQ-CAT) Physical functioning <u>Subjective, not useful:</u> - Follick's Diary - Minnesota Leisure Time Physical Activity Questionnaire. <u>Newer PRO tools to evaluate physical functioning:</u> - PROactive tools - Capacity of Daily Living during the Morning questionnaire (CDLM) - London Chest Activities of Daily Living (LCADL) questionnaire Exacerbations - Exacerbations of Chronic Pulmonary Disease tool (EXACT) Health status and QoL outcomes <u>Most commonly used:</u> - St George's Respiratory Questionnaire (SRGQ) - Chronic Respiratory Disease Questionnaire (CRDQ) <u>Newer tools:</u> - COPD Assessment Test (CAT) - Clinical COPD Questionnaire - Living with COPD (LCOPD).	evaluate the tools used for measuring these concepts.	assessment of the overall effect of the disease and have been well tested in a variety of clinical settings and populations. They are known to be responsive to a wide range of therapeutic interventions, and can provide an overall measure of the response to treatment. Total scores, such as those obtained with the SGRQ, are "black box" measurements and provide little or no information of the specific nature of the benefit or any insight into mechanisms of benefit. A number of more specific tools have been developed to evaluate various aspects of the disease, although many of these were developed prior to new regulatory guidelines and may only be valid as secondary or supportive outcome measures in clinical trials. There is a very large body of published evidence concerning the SGRQ in research studies of all kinds. It has been accepted by the EMA as a symptomatic outcome measure in COPD trials and it is becoming accepted as an outcome measure for COPD studies by the FDA. A white paper to support that purpose is being put together by a consortium working with the COPD Foundation in the US. For a shorter measure, both the CCQ and CAT have demonstrated validity and responsiveness. Disclosure: funded by Novartis Pharmaceutical Company.
Martin 2016		
Various outcomes - FEV₁ - St George's Respiratory Questionnaire (SGRQ) - Exacerbations.	1. Systematic review. 2. The association of FEV₁ and SGRQ with exacerbations.	The regression analysis demonstrated a significant association between improvements in FEV₁ and SGRQ score and lower risk for COPD exacerbations. Even in cases of non-significant relationships, results were in the expected direction with few exceptions. The results of this analysis offer health care providers and payers a broader picture of the relationship between exacerbations and mean change in FEV₁ as well as SGRQ score, and will help inform clinical and formulary-making decisions while stimulating new research questions for future prospective studies. Disclosure: funded by Novartis Pharmaceutical Company.
Partridge 2010		
PRO questionnaires - Capacity of Daily Living during the Morning questionnaire (CDLM) - Global Chest Symptoms Questionnaire (GCSQ).	1. Article. 2. Development and validation of PROs, CDLM and GCSQ.	Both the CDLM questionnaire and the GCSQ are easy-to-use, reliable, responsive, self-administered questionnaires that report on patients' symptoms and ability to perform morning activities. The CDLM questionnaire and GCSQ could be incorporated into multinational clinical trials to assess the impact of COPD on morning symptoms and the patient's ability to perform morning activities. Further evaluation of the CDLM questionnaire and GCSQ will determine the utility of these tools in general clinical practice.
Santus 2014		
Lung function (sRAW).	1. Article.	The present results support the notion that the

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
	2. Assessment of bronchodilator effects with respect to sRAW.	effects of bronchodilators should be evaluated from the changes of airway resistance in the resting tidal volume range rather than FEV ₁ and/or FVC, especially in connection with the dependent improvement of dyspnoea. In contrast, both FEV ₁ and sRAW are equally good evaluators of the severity of bronchoconstriction at baseline. It is also shown that the decrease of RV and ITGV is mainly due to changes in the mechanical properties of the peripheral airways leading to decreased resistance and closing pressure. Finally, the ability to lower dyspnoea at rest appears mainly related to the concomitant reduction of airway resistance rather than lung deflation.
Singer 2012		
PROs for HRQoL <u>Generic</u> • 36-Item Short Form Survey (SF-36) • Nottingham Health Profile • Sickness Impact Profile <u>Disease-specific</u> • St George's Respiratory Questionnaire (SGRQ) • Chronic Respiratory Disease Questionnaire (CRQ) • University of California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) • modified Medical Research Council (mMRC) - dyspnoea • Baseline Dyspnoea Index/Transition Dyspnoea Index (BDI/TDI) <u>Multi-attribute or preference-based utility</u> Indirect: • EuroQoL Five Dimensions Questionnaire (EQ-5D) • Quality of Well Being (QWB) • HUI (Health Utilities Index) Mark 3 Direct: • Standard Gamble Other: • 100-mm Visual Analogue Scale (VAS).	1. Review 2. Defining PROs in COPD.	Patients with COPD have reduced HRQoL, and HRQoL worsens as COPD progresses. Although long-term interventions such as smoking cessation, supplemental oxygen therapy for hypoxemia, LVRS, and lung transplantation may save lives, these and other therapies have variable but important effects on PROs. Measures of HRQoL can serve as measures of disease severity and predict outcome. Moreover, these measures are sensitive to change following interventions, and can thus be used as measures of intervention effect. Future research will provide a better understanding of the effects of COPD on HRQoL and the impact of various interventions.
Singh 2014		
<u>Exercise capacity</u> • 6-Minute Walk Test (6-MWT) • Body-mass index, airflow Obstruction, Dyspnea, and Exercise (BODE) index • Incremental Shuttle Walking Test (ISWT) • Endurance Shuttle Walking Test (ESWT) • Cycle endurance test <u>HRQoL</u> • Chronic Respiratory Disease Questionnaire (CRQ) • St George's Respiratory Questionnaire (SGRQ) • COPD Assessment Test (CAT) <u>Functional performance</u> • Pulmonary Functional Status and Dyspnoea Questionnaire (PFSDQ) • Pulmonary Function Status Scale (PFSS) • London Chest Activities of Daily Living (LCADL) scale • Canadian Occupational Performance Measure <u>Anxiety and depression</u> • Hospital Anxiety and Depression Scale (HADS) <u>Symptoms (breathlessness)</u> • 100-mm Visual Analogue Scale (VAS) • Borg breathlessness scale	1. Review. 2. Outcome assessment in pulmonary rehabilitation.	It is vital that a comprehensive assessment be conducted to support the delivery of a successful pulmonary rehabilitation program. The development of a wide range of outcome measures should not detract from the main components of an assessment. It should be evident that it is not possible for a particular pulmonary rehabilitation program to incorporate all these outcome measures into its routine assessments. Rather, a limited number of assessments are made across outcome areas. Over time, there may be many developments that enhance the assessment of individuals with chronic respiratory disease before commencing a rehabilitation program. A sophisticated suite of outcomes might allow rehabilitation to be personalised.

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
- Medical Research Council (MRC) Dyspnoea Scale Knowledge and Self-Efficacy - Lung Information Needs Questionnaire (LINQ) - Bristol COPD Knowledge Questionnaire (BCKQ) - COPD self-efficacy scale - Pulmonary Rehabilitation Adapted Index of Self-Efficacy (PRAISE).		
Singh 2016		
Exploratory composite endpoint: Clinically important deterioration (CID – a decrease of ≥ 100 mL in trough FEV₁ or ≥ 4-unit increase in SGRQ total score or an on-treatment moderate-to-severe COPD exacerbation).	1. Clinical study. 2. Prevention of CIDs in COPD with umeclidinium/vilanterol.	This exploratory analysis, using a new assessment of clinical deterioration in COPD, revealed that a majority of symptomatic patients with low exacerbation risk experienced a deterioration during the 24-week study periods. UMEC/VI reduces the risk of a first CID versus placebo or bronchodilator monotherapy.
Tabberer 2015		
Shortness of Breath with Daily Activities (SOBDA) Questionnaire	1. Meta-analysis. 2. Confirmation of the reliability and responsiveness of the SOBDA questionnaire.	The reliability, validity, and responsiveness of the SOBDA questionnaire as a PRO measure to quantify dyspnoea was supported in a large clinical trial population of patients with moderate–very severe COPD.
van der Molen 2012		
All outcomes - FEV₁ - Inspiratory capacity (IC) - Dyspnoea (Transition Dyspnoea Index (TDI)) - Respiratory symptoms (shortness of breath, chest tightness, night time awakenings, and total respiratory symptom scores, cough scores) - Rescue medication use - Exacerbations - Health status (St George's Respiratory Questionnaire (SGRQ)) - Safety.	1. Review. 2. LABA/LAMA combination therapies demonstrate greater improvements in patient-centred outcomes than in FEV ₁ .	The value of FEV ₁ alone as a surrogate marker of COPD is limited, and patient-centred outcomes are important for both adequate recognition of the disease and effective treatment of patients.
Vestbo 2014		
Predictive surrogate endpoints - FEV₁ - Modified Medical Research Council (mMRC) score 6-Minute Walk Test (6-MWT) - St George's Respiratory Questionnaire for COPD (SGRQ-C) - Number of exacerbations.	1. Clinical study 2. Longitudinal characterisation of the heterogeneity and variability of COPD.	The authors report that they found a striking heterogeneity among patients with COPD, with poor correlations between FEV ₁ , symptoms, QoL, functional outcomes, and biomarkers.
Westwood 2011		
Relationship between two outcomes - FEV₁ - St George's Respiratory Questionnaire (SGRQ) - Exacerbations - Dyspnoea.	1. Systematic review. 2. Relationship between changes in FEV ₁ and changes in health status with bronchodilator therapy.	Improvement in mean trough FEV ₁ is associated with proportional improvements in health status (SGRQ). Disclosure: funded by Novartis Pharmaceutical Company.
Wilt 2012		
Various outcomes - Exacerbations - FEV₁ - HRQoL - Harms.	1. Review. 2. Type of evidence and outcomes to include in COPD guidelines.	Outcomes should address both benefits and downsides, with consideration of the definitions, severity, and time course of the outcomes.
Yohannes 2011		
Various outcomes - HRQoL (St George's Respiratory Questionnaire (SGRQ)) - Dyspnoea (Transition Dyspnoea Index (TDI))	1. Systematic review and meta-analysis. 2. Clinically relevant	Tiotropium showed superior efficacy for QoL and dyspnoea compared with other agents.

Outcome/s	1. Type of publication 2. Subject of publication	Authors comments/conclusion
- Exacerbations - Hospitalisations - Harms.	outcomes with respect to tiotropium therapy.	

Abbreviations: BCKQ, Bristol COPD Knowledge Questionnaire; BCSS, Breathlessness, Cough and Sputum Scale; BDI, Baseline Dyspnoea Index; BODE, Body-mass index, airflow Obstruction, Dyspnea, and Exercise; CASIS, COPD and Asthma Sleep Impact Scale; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; CDLM, Capacity of Daily Living during the Morning questionnaire; CID, clinically important deterioration; COPD, chronic obstructive pulmonary disease; CRQ, Chronic Respiratory Disease Questionnaire; DMQ-CAT, Dyspnoea Management Questionnaire Computer Adaptive Test; EMA, European Medicines Agency; EQ-5D, EuroQoL Five Dimensions Questionnaire; E-RS, EXACT – Respiratory Symptoms; ESS, Epworth Sleepiness Scale; ESWT, Endurance Shuttle Walking Test; EXACT, Exacerbations of Chronic Pulmonary Disease tool; FDA, Food and Drug Administration; FEV₁, forced expiratory volume; GCSQ, Global Chest Symptoms Questionnaire; HADS, Hospital Anxiety and Depression Scale; HRQoL, health-related quality of life; HUI, Health Utilities Index; IC, inspiratory capacity; ISWT, Incremental Shuttle Walking Test; ITGV, intrathoracic gas volume; LABA, long-acting beta-2 agonist; LAMA, long-acting muscarinic antagonist; LCADL, London Chest Activities of Daily Living; LCOPD, Living with COPD; LINQ, Lung Information Needs Questionnaire; MCID, minimal clinically important difference; MDP, Multidimensional Dyspnoea Profile; mMRC, modified Medical Research Council; MRC, Medical Research Council; NiSCI, Night time symptoms of COPD Instrument; NRS, Numeric Rating Scale; PFSQ, Pulmonary Functional Status and Dyspnoea Questionnaire; PFSS, Pulmonary Function Status Scale; PRAISE, Pulmonary Rehabilitation Adapted Index of Self-Efficacy; PRO, patient-reported outcomes; PSQI, Pittsburgh Sleep Quality Index; QoL, quality of life; QWB, Quality of Well Being; SF-36, 36-Item Short Form Survey; SF-CRQ, Short-Form Chronic Respiratory Disease Questionnaire; SAC, Self-administered computerised; SGRQ, St George's Respiratory Questionnaire; SGRQ-C, St George's Respiratory Questionnaire for COPD; SOBDA, Shortness of Breath with Daily Activities Questionnaire; sRaw, specific airway resistance; SWT, Shuttle Walk Test; TDI, Transition Dyspnoea Index; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire; VAS, 100-mm Visual Analogue Scale; VSRQ, Visual Simplified Respiratory Questionnaire.

The authors of a review of strengths and limitations of outcomes in COPD (Glaab et al, 2010) have described the difficulties of using exacerbations as an outcome, as follows:

- There is no standardised definition of an exacerbation, making comparative evaluations of clinical study results difficult.*
- 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.*
- 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, ischaemic 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.*
- No MCID has been established yet.*

Note that the first author (Glaab) was an employee of Boehringer Ingelheim at the time of submission of the manuscript. Boehringer Ingelheim is the sponsor for the PBS-listed medicines ipratropium (various tradenames), tiotropium (Spiriva and Spiriva Respimat), and tiotropium/olodaterol FDC (Spiolto Respimat).

The PSD for tiotropium/olodaterol (see Table 0.4 in Section 2.3.3) indicates that the FDC was recommended for PBS listing based on the outcomes FEV₁ and exacerbations. The PSDs for the other two products are not available.

2.3.3 Health technology assessment and reimbursement agencies

No information specifically related to COPD outcomes judged as relevant to this review was found on the following sources: the AHRQ, CADTH, NHS HTA/NCCHTA, NHS CRD and NICE.

Pertinent details from the PSDs for PBS-listed COPD medications are shown in Table 0.4. The outcomes reported are dominated by FEV₁, which is captured for every submission except one (fluticasone propionate/salmeterol, recommended in 2007). The SGRQ is the most commonly utilised HRQoL measure for submissions, which is expected given that it is a COPD/asthma disease-specific measure. Also reported often are exacerbations of COPD. Outcomes that are referred to rarely or not at all in the PSDs are BDI/TDI, rescue medications, hospitalisations, and mortality. These outcomes may have been reported in the included studies in the submissions, but are not highlighted in the PSDs. Also, hospitalisation may be part of a definition of an exacerbation, especially for a severe exacerbation.

The PSDs for the glycopyrronium/indacaterol and the umeclidinium/vilanterol LAMA/LABA FDCs raise the issue of relevance of trough FEV₁ to more direct, patient-relevant measures of effect. The March 2014 PSDs for both products state that:

The PBAC noted that it had previously accepted trough FEV₁ as a surrogate measure of effect. However, the PBAC considered that additional clinical outcomes such as frequency of exacerbations and hospitalisations would be informative as more direct, patient-relevant measures of effect.

As the incremental gain in FEV₁ of [glycopyrronium/indacaterol / umeclidinium/vilanterol] FDC was not able to be translated into more clinically relevant measures of effect (e.g. frequency of exacerbations, hospitalisations), the PBAC considered it was unable to determine and value the incremental benefit associated with use the FDC compared with use of components given concurrently. Therefore, the Committee was unable to determine an appropriate price for the FDC.

Both FDC products were rejected at the March 2014 meeting, and both products were subsequently recommended for listing at the July 2015 meeting with the following note in their respective PSDs:

The PBAC noted that the primary concerns raised in the March 2014 submission were not adequately addressed in the resubmission as the incremental benefit of the combination product could not be translated into clinically relevant measures of effect. However, the Committee accepted that there are both benefits and cost savings for patients who are already using individual LAMA and LABA in separate devices.

According to the PSDs for the LAMA products, there were generally a range of outcomes considered for these agents, including FEV₁ and various PRO. These agents were recommended for listing between 2013 and 2015. Similarly, the PSDs for the ICS/LABA FDCs refer to FEV₁ and a range of other outcomes. These products were recommended for listing between 2007 and 2014. In contrast, the LAMA/LABA FDCs recently recommended for listing in 2014-15 focus heavily on FEV₁ without associated PRO.

Table 0.4 Outcomes referenced in Public Summary Documents (PSDs) for PBS-listed COPD medications

Name of active ingredient (brand name and strength)	PSD date (outcome)	Source	Outcomes accepted by PBAC as clinically relevant
SAMA			
Ipratropium	N/A	N/A	N/A
LAMA			
Acclidinium (Bretaris Genuair 400 µg actuations)	Mar 2014 (recommended)	Acclidinium Mar 2014 PSD	• Change from baseline in trough FEV₁ at 6-24 weeks, 52 weeks • Change from baseline in SGRQ • COPD exacerbations • AEs
Glycopyrronium (Seebri Breezhaler 50 µg inhalation capsules)	Nov 2013(recommended, managed entry ^a) Nov 2015 (update of RCT data ^b)	Glycopyrronium Nov 2013 PSD Glycopyrronium Nov 2015 PSD	• Change from baseline in trough FEV₁ at 12, 26, 52 64 weeks • AEs
Tiotropium (Spiriva 18 µg inhalation capsules; Spiriva Respimat 2.5 µg actuations)	N/A	N/A	• Outcomes from the tiotropium submission:^c • FEV₁ • FVC • PEFR • BDI/TDI • SGRQ • SF-36 • COPD exacerbations • COPD hospitalisations • all-cause hospitalisations • AEs • deaths
Umeclidinium (Incruse Ellipta 62.5 µg actuations)	Jul 2014 (recommended)	Umeclidinium July 2014 PSD	• Change from baseline in trough FEV₁ at 12, 24 weeks • TDI focal score • Change from baseline in SGRQ • Rescue medication at 12 weeks • COPD exacerbations at 24 weeks • AEs
LABA			
Indacaterol (Onbrez 150 µg, 300 µg inhalation capsules)	Nov 2010 (rejected) Jul 2011 (recommended)	Indacaterol Nov 2010 PSD Indacaterol July 2011 PSD	• Change from baseline in trough FEV₁ at 12 weeks • AEs
LAMA/LABA			
Acclidinium/eformoterol ^a (Brimica Genuair 340/12 µg actuations)	Jul 2015 (recommended)	Acclidinium/eformoterol July 2015 PSD	• Pre-dose FEV₁ • Change from baseline in trough FEV₁ at 24 weeks • 1-hour post-dose FEV₁ • AEs

Post-market Review of COPD Medicines

Name of active ingredient (brand name and strength)	PSD date (outcome)	Source	Outcomes accepted by PBAC as clinically relevant
Glycopyrronium/indacaterol (Ultibro Breezhaler 50/110 µg inhalation capsules)	Mar 2014 (rejected) Jul 2014 (recommended)	Glycopyrronium/indacaterol Mar 2014 PSD Glycopyrronium/indacaterol July 2014 PSD	- Change from baseline in trough FEV₁ at 4, 12 and 26 weeks - AEs
Tiotropium/olodaterol (Spiolto Respimat 2.5/2.5 µg actuations)	Jul 2015 (recommended)	Tiotropium/olodaterol July 2015 PSD	- Change from baseline in trough FEV₁ at 24 weeks - Change from baseline in trough FEV₁ exacerbations - AEs
Umeclidinium/vilanterol (Anoro Ellipta 62.5/25 µg actuations)	Mar 2014 (rejected) Jul 2014 (recommended) Nov 2014 (rejected)	Umeclidinium/vilanterol March 2014 PSD Umeclidinium/vilanterol July 2014 PSD Umeclidinium/vilanterol Nov 2014 PSD	- Change from baseline in trough FEV₁ at 12, 24/26 weeks - AEs
OCS			
Prednisolone (Solone 5 mg, 25 mg tablets)	N/A	N/A	N/A
ICS/LABA			
Budesonide/eformoterol (Symbicort Turbuhaler DPI 400/12 µg actuations)	Nov 2010 (recommended)	Budesonide/eformoterol Nov 2010 PSD	- Rate of COPD exacerbations - Mean change from baseline as measured by SGRQ - Mean change from baseline for pre- and post-dose FEV₁ - AEs
Budesonide/eformoterol (Symbicort Rapihaler MDI 200/6 µg actuations)	Jul 2013 (recommended)	Budesonide/eformoterol July 2013 PSD	As above.
Fluticasone propionate/salmeterol (Seretide Accuhaler 500/50 µg actuations; Seretide MDI 250/25 µg actuations)	March 2007 (recommended)	Fluticasone propionate/salmeterol Mar 2007 PSD	- Rate of health care utilisation exacerbations - Mortality
Fluticasone furoate/vilanterol (Breo Ellipta 100/25 µg actuations)	Mar 2014 (rejected) July 2014 (recommended)	Fluticasone furoate/vilanterol Mar 2014 PSD Fluticasone furoate/vilanterol July 2014 PSD	- Change from baseline in trough 24-hour weighted mean serial FEV₁ at 12 weeks - QoL determined using change in baseline SGRQ-C total score at 12 weeks - Health status using change in baseline EQ-5D values at 12 weeks - AEs

Abbreviations: AE, adverse event; COPD, chronic obstructive pulmonary disease; DPI, dry powder inhaler; EQ-5D, EuroQol five dimensions questionnaire; FEV, forced expiratory volume in 1 second; ICS, inhaled corticosteroid; LABA, long-acting beta-2 agonist; LAMA, long-acting muscarinic antagonist; MDI, metered dose inhaler; OCS, oral corticosteroid; PBAC, Pharmaceutical Benefits Advisory Committee; PEFR, peak expiratory flow rate; PSD, Public Summary Document; RCT, randomised controlled trial; SF-36, Short Form-36; SGRQ, St George's Respiratory Questionnaire; SGRQ-C, St George's Respiratory Questionnaire – COPD; TDI, Transition Dyspnoea Index.

a The PBAC requested that the sponsor provide data from the GLISTEN trial when it was completed, to confirm the assessment of comparative effectiveness and safety of glycopyrronium in combination with ICS/LABA.

b The PBAC noted that the publication supplied by the sponsor reaffirmed the view of the Committee in November 2013 that glycopyrronium was non-inferior in regards to efficacy and safety with tiotropium.

c There is no PSD available for tiotropium; this list of outcomes was obtained from the tiotropium submission.

2.3.4 Clinical guidelines and clinical studies

As discussed in Section **Error! Reference source not found.**, the two key COPD clinical practice guidelines of relevance to Australian practice are the Australian-based COPD-X and the US-based GOLD Strategy Report. The GOLD Strategy Report contains information relevant to outcomes in COPD, which is summarised in Table 0.5. The COPD-X guidelines do not contain outcome information relevant to this review.

Table 0.5 COPD outcome information from the GOLD Strategy Report (2016)

GOLD Strategy Report ^a
Lung function In patients with FEV₁/FVC < 0.70: • GOLD 1: Mild FEV₁ ≥ 80% predicted • GOLD 2: Moderate 50% ≤ FEV₁ < 80% predicted • GOLD 3: Severe 30% ≤ FEV₁ < 50% predicted • GOLD 4: Very Severe FEV₁ < 30% predicted. There is only a weak correlation between FEV₁, symptoms and impairment of a patient's HRQoL. Within any given category, patients may have anything between relatively well preserved to very poor health status. For this reason, formal symptomatic assessment is also required.
Symptoms • mMRC questionnaire – previously a simple measure of breathlessness used in COPD, but now replaced by more comprehensive symptom assessment. • CRQ and SGRQ – too complex to use in routine practice. • CAT and CCQ – shorter and suitable comprehensive measures. • CAT – The CAT is an 8-item unidimensional measure of health status impairment in COPD. It was developed to be applicable worldwide and validated translations are available in a wide range of languages. The score ranges from 0-40, correlates very closely with the SGRQ, and has been extensively documented in numerous publications (COPD Assessment Test Website). • CCQ – The CCQ is a 10 item self-administered questionnaire developed to measure clinical control in patients with COPD. Although the concept of 'control' in COPD remains controversial, the CCQ is short and easy to administer. It is reliable and responsive, is available in a range of languages, and has been validated (Clinical COPD Questionnaire Website). A MCID during rehabilitation of -0.4 for the CCQ has been identified.
Exacerbations An exacerbation of COPD is defined as an acute event characterised by a worsening of the patient's respiratory symptoms that is beyond normal day-to-day variations and leads to a change in medication. The rate at which exacerbations occur varies greatly between patients. The best predictor of having frequent exacerbations (2 or more exacerbations per year) is a history of previous treated events. In addition, worsening airflow limitation is associated with an increasing prevalence of exacerbations and risk of death. Hospitalisation for a COPD exacerbation is associated with a poor prognosis with increased risk of death. A large body of data has been accumulated in patients classified using GOLD spirometric grading systems. These show an increase in risk of exacerbations, hospitalisation and death with worsening of airflow limitation, demonstrated by data derived from prospective large medium-term clinical trials (TORCH, UPLIFT and ECLIPSE). The data illustrate clearly the increased risk of exacerbations, hospitalisation and death between spirometric levels. Up to 20% of GOLD 2 (Moderate airflow limitation) patients may experience frequent exacerbations requiring treatment with antibiotics and/or systemic corticosteroids. The risk of exacerbations significantly increases in GOLD 3 (Severe) and GOLD 4 (Very Severe). Since exacerbations increase the decline in lung function, deterioration in health status and risk of death, the assessment of exacerbation risk can also be seen as an assessment of the risk of poor outcomes in general.
Exercise testing Walking tests can be useful for assessing disability and are used to assess the effectiveness of pulmonary rehabilitation. Both the paced SWT and the unpaced 6-MWT can be used. As the course length has a substantial impact on the distance walked, existing reference equations established for a 30 m course cannot be applied to predict the distance achieved on shorter courses. Laboratory testing using cycle or treadmill ergometry can identify co-existing or alternative conditions e.g. cardiac diagnoses.
Composite scores Several variables including FEV₁, exercise tolerance assessed by walking distance or peak oxygen consumption, weight loss, and reduction in arterial oxygen tension identify patients at increased risk for mortality. A relatively simple approach to identifying disease severity using a combination of most of the above variables has been proposed. The BODE method gives a composite score that is a better predictor of subsequent survival than any component singly, and its properties as a measurement tool are under investigation. Simpler alternatives not including an exercise test have been suggested but all these approaches need validation across a wide range of disease severities and in different clinical settings to confirm that they are suitable for routine clinical use.

Abbreviations: BMI, body mass index; BODE, BMI, Obstruction, Dyspnoea and Exercise; CAT, COPD Assessment Test; CCQ, COPD Control Questionnaire; CRQ, Chronic Respiratory Questionnaire; FEV₁, forced expiratory volume; MCID, minimal clinically important difference; mMRC, modified Medical Research Council; SGRQ, St George's Respiratory Questionnaire; SWT, Shuttle Walk Test.

^a From [Global Strategy for Diagnosis, Management, and Prevention of COPD 2016](#)

The GOLD Strategy Report contains a section on 'Assessment of disease'. Although the information presented appears to be intended for patient assessment in clinical practice, as opposed to clinical trials, it is still useful in the context of a discussion around clinically relevant outcomes.

The GOLD Strategy Report states that there is only a weak correlation between FEV₁ and HRQoL (Jones et al, 2009). The guidelines show that if SGRQ is plotted against post-bronchodilator FEV₁, patients may have anything between relatively well preserved to very poor health status (see Figure 2.2, p14 of the guidelines, which shows a figure adapted from Jones et al, 2009). Jones et al (2011) showed that although change in FEV₁ is a significant but relatively weak predictor of improvement in PROs at the patient level, improvements in FEV₁ with long-acting bronchodilator therapy are strongly correlated with SGRQ at the population level (see discussion in Section 2.3.2).

The GOLD Strategy Report presents evidence that there is an increase in risk of exacerbations, hospitalisation and death with worsening of airflow limitation. This is demonstrated by data derived from prospective large medium-term clinical trials (TORCH, UPLIFT and ECLIPSE), showing that as the GOLD spirometric level increases from mild to very severe COPD, so too do the rates of exacerbations, hospitalisations and mortality.

The GOLD Strategy Report states that an understanding of the impact of COPD on an individual patient combines symptomatic assessment with the patient's spirometric classification and/or risk of exacerbations. The GOLD Strategy Report provides a worked example of combined COPD assessment. The guidelines recommend combining a symptomatic assessment using CAT, spirometric classification using FEV₁, and/or risk of exacerbations using either the GOLD spirometric classification or the individual patient's history of exacerbations. The guidelines state that this approach, combined with an assessment of potential comorbidities, reflects the complexity of COPD better than the unidimensional analysis of airflow limitation previously used for staging the disease.

The outcomes reported in the clinical trials (Section 2) are consistent with those that are listed in Section 2.3.2 of this report as being commonly reported and of current interest to researchers. It is apparent from the clinical trials that lung function can be tested using several different methods, with spirometry being the most common. Various spirometric outcomes are reported in the clinical trials, most commonly:

- trough FEV₁ or forced volume vital capacity (FVC), generally measured between 23 and 24 hours post-dose
- peak FEV₁ or FVC
- area under the curve (AUC) for FEV₁ over different periods of time (e.g. 0-6 hours; 0-12 hours; 12-24 hours; 0-24 hours).

2.3.5 Minimal clinically important difference for COPD outcomes

A discussion around MCID is important when considering PRO. The MCID reflects the minimum difference in a particular outcome that is clinically significant to a patient. Singh et al (2016) have commented that studies comparing outcomes for a FDC with each of their mono-components do not always satisfy the MCID, because MCIDs have traditionally been used for placebo-controlled studies and the magnitude between two active treatments may not be as great. Table 0.6 shows values for MCID that have been reported for COPD outcomes. In the clinical section of this review, it is evident that some interventions show a statistically significant benefit to patients with respect to FEV₁, but do not meet the threshold for MCID.

Table 0.6 Summary of MCID values for the main outcomes used in COPD

Outcome	Surrogate outcomes	Patient-relevant outcomes	MCID	Reference for MCID
Mortality				
Death			N/A	N/A
Lung function				
Change in FEV ₁	✓		Not established yet. The authors suggested that an appropriate range of values for the MCID for FEV ₁ might be 100-140 mL but the MCID for FEV ₁ remains poorly defined for COPD.	Glaab 2010
Lung volume				
RV, TLC, FRC			Neither a standardised classification for the assessment of severity of hyperinflation nor a MCID 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.	Glaab 2010
Exacerbations				
Duration		✓	Not established yet.	Glaab 2010
Severity		✓		
Delay in occurrence		✓		
Time to first exacerbation		✓		
Reduction in frequency		✓		
Health status and HRQoL				
SGRQ		✓	At least 4 unit reduction 2 to 8 points, most often a value of 4 ^b	FDA (Table 2.2) ^a Glaab 2010 ^b
CRQ		✓	0.5 unit increase per question.	Wilt 2012
SF-36			Not established yet.	Glaab 2010
CDLM		✓	0.20 ^c	Partridge 2010
CCQ		✓	-0.4	GOLD Strategy Report 2016
Symptoms				
GCSQ		✓	0.15 ^c	Partridge 2010
Dyspnoea				
BDI/TDI		✓	At least 1 unit in TDI 1 unit	van der Molen 2012 Glaab 2010
SOBDA			0.13-0.25 points	Ekström 2015
Dyspnoea-12			No MCIDs reported.	Ekström 2015

Outcome	Surrogate outcomes	Patient-relevant outcomes	MCID	Reference for MCID
MDP			No MCIDs reported.	Ekström 2015
VAS			9 mm (95% CI 2.1 to 15.8)	Ekström 2015
E-RS			1 point	Ekström 2015
CR-10 or Borg Scale			1 unit	Glaab 2010
Exercise capacity				
6-MWT	✓		53 m	Wilt 2012
			54-80 m	Glaab 2010
			Approximately 50 m	Singh 2014 and references therein
SWT	✓		47.5 m	Glaab 2010
Treadmill	✓		Not established yet.	Glaab 2010
Cycle ergometry	✓		Not established yet.	Glaab 2010
Composite scores				
BODE-Index			Not established yet.	Glaab 2010
CID (FEV ₁ , SGRQ, exacerbations)			Trough FEV ₁ : decrease of ≥100 mL SGRQ: ≥4-unit increase On-treatment moderate-severe exacerbation	Singh 2016 Singh 2016 Singh 2016

Abbreviations: BDI, Baseline Dyspnoea Index; BODE, Body-mass index, airflow Obstruction, Dyspnea, and Exercise; CAT, COPD Assessment Test; CCQ, COPD Control Questionnaire; CI, confidence interval; CID, clinically important deteriorations; COPD, chronic obstructive pulmonary disease; CRQ, Chronic Respiratory Questionnaire; CT, computed tomography; FRC, functional residual capacity; FVC, forced vital capacity; GCSQ, Global Chest Symptoms Questionnaire; HRQoL, Health-Related Quality of Life; IC, inspiratory capacity; MCID, minimal clinically important difference; MDP, Multidimensional Dyspnoea Profile; MRC, Medical Research Council; 6-MWT, Six-Minute Walk Test; N/A, not applicable; NR, not reported; RV, residual volume; SGRQ, St George's Respiratory Questionnaire; SOBDA, Shortness of Breath with Daily Activities; SWT, Shuttle Walk Test; TDI, Transition Dyspnoea Index; TLC, total lung capacity.

^a From US FDA, [Chronic Obstructive Pulmonary Disease: Developing Drugs for Treatment \(Draft Guidance\)](#).

^b The authors state that there is little published empiric evidence supporting the MCID of four points.

^c This value corresponds to an SGRQ-C MCID of 4.

2.3.6 Discussion and summary

In the original 2002 tiotropium submission for COPD, the PBAC was first presented with the relationship between FEV₁ and disability. This submission reported outcome results for lung function (FEV₁, FVC, PEFR), dyspnoea (BDI/TDI), HRQoL (SGRQ, SF-36), COPD exacerbations, COPD hospitalisations, all-cause hospitalisations, AEs, and deaths. The submission did not refer to the surrogate nature of the FEV₁ outcome in COPD, and FEV₁ was called the gold standard measure for airflow obstruction. As expected, the references used to justify the selection of these outcomes in the 2002 submission are now out of date.

With the exception of PEFR, data for all the outcomes utilised for the tiotropium submission are still collected in recent studies. Furthermore, the main outcomes published in the PSDs for COPD submissions since 2002 (FEV₁, SGRQ, exacerbations, rescue medication and AEs; see Table 0.4) are the same as those presented in the tiotropium submission. As discussed in Section 2.3.3, the commentary in the 2014 PSDs for glycopyrronium/indacaterol and umeclidinium/vilanterol (both LAMA/LABA combinations) reflects concern over the translation of FEV₁ into more clinically relevant measures of effect that were not reported in the submissions. Although these products were ultimately listed on the PBS after resubmissions, the PSDs indicate that these issues were not resolved.

The literature search for published articles discussing outcomes for COPD showed that there is recent interest in demonstrating that FEV₁ is correlated with PROs. The publication by

Martin et al (2016), which shows a significant association between improvements in either FEV₁ or SGRQ score and lower risk for COPD exacerbations, supports the emphasis that the PBAC have placed on FEV₁ in their decision-making. Two other publications also support FEV₁ as a surrogate endpoint that is correlated with SGRQ and exacerbations. Authors comment that the correlation at a patient level is weak. All three studies were funded by Novartis, the manufacturer of three currently PBS-listed COPD medications. In contrast, two industry-funded reviews found a poor correlation between FEV₁ and PROs. The literature search also showed that there is a general consensus among authors that PROs are required to supplement FEV₁ due to its surrogate nature, that there is no single PRO that is an ideal tool, and that clinical guidelines could identify which PROs should be used routinely.

The regulatory agencies that provide information on clinically relevant outcomes in COPD are the FDA and the EMA. Both agencies emphasise that lung function (in the form of FEV₁) is the most widespread efficacy endpoint used in the measurement of airflow obstruction. The agencies also emphasise exacerbations, HRQoL, symptom scales, and exercise capacity as outcomes relevant to COPD.

The GOLD Strategy Report provides evidence of a weak correlation between FEV₁ and SGRQ. The guidelines also present evidence that there is an increase in risk of exacerbations, hospitalisation and death with worsening of airflow limitation. The guidelines recommend an approach of combining symptomatic assessment with the patient's spirometric classification and/or risk of exacerbations, which is consistent with the above-mentioned PBAC decision-making based on FEV₁, SGRQ and exacerbations.