National Institute for Health and Care Excellence

Final

Chronic obstructive pulmonary disease in over 16s: diagnosis and management

[D] Diagnosing COPD and predicting outcomes

NICE guideline NG115

Evidence reviews

December 2018

Final

These evidence reviews were developed by the NICE Guideline Updates Team

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3175-0

Contents

Confirming COPD diagnosis

Review question

In people with suspected COPD, what is the most accurate and appropriate test (for example imaging or biomarkers) to confirm the diagnosis?

Introduction

Clinical diagnosis of COPD is based on the results of spirometry. This test is carried out on people presenting with symptoms that are associated with COPD including breathlessness, chronic cough or sputum production and /or a history of risk factors such as current or previous tobacco smoking and/or other smoked drugs and occupational exposures. However, imaging tests carried out to investigate other issues may identify people with signs of airway disease who are otherwise asymptomatic. In addition to spirometry, other tests, such as chest X-rays, can used to investigate alternate diagnoses that may explain symptoms, for example lung cancer. Such tests may also detect concomitant abnormalities at the time of the initial diagnostic evaluation.

This review aims to determine the diagnostic accuracy of tests for the diagnosis of COPD in people with a diagnostic or non-diagnostic spirometry result or without spirometry results. For this guideline update, the term COPD covers people with chronic bronchitis, emphysema, and chronic airflow limitation or obstruction. The population of interest are people with COPD, COPD with asthma, COPD with bronchopulmonary dysplasia, or COPD with bronchiectasis.

This review identified studies that fulfilled the conditions specified in Table 1. For full details of the review protocol, see appendix A.

Table 1 PICO: confirming diagnosis of COPD

Population	People with suspected COPD
Assessment tools	Imaging, including: • Chest CT • Chest X-ray • 18F-FDG-PET • Lung MRI with or without O2, 3He or 129Xe • Pulmonary Scintigraphy (Ventilation-Perfusion ($\dot{V}/\dot{Q}$) SPECT) Other tests: • Full Reversibility of airways obstruction in response to bronchodilator on spirometry (adjusted for BMI) • Sputum culture • Serial peak flow measurements (peak expiratory flow rate (PEFR), or peak expiratory flow (PEF). Both L/min). • Serum Alpha-1 antitrypsin and other tests for alpha-1 antitrypsin deficiency • Transfer factor for carbon monoxide (TLCO) • ECG • Echocardiogram • Pulse oximetry (peripheral oxygen saturation, SpO_2) • Arterial blood gas analysis

	• Sputum myeloperoxidase and serum interleukin-6 • Systemic inflammatory markers including eosinophil count • Full blood count
Reference standard	• Clinical diagnosis of COPD by any means including Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, guideline; American Thoracic Society criteria for COPD; European Respiratory Society criteria) • Post-bronchodilator spirometry in a stable patient • CT demonstration of emphysema • Histopathology grading of emphysema
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A, and the methods section in appendix B.

Subgroup analyses were not carried out for this review because the included studies did not report data for the categories of interest in an accessible format.

The search strategies used in this review are detailed in appendix C.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

Included studies

A single systematic search was carried out for the 3 review questions in this evidence review to identify observational studies and systematic reviews of observational studies, which found 15,231 references (see appendix C for literature search strategy). Evidence included in the original guideline, evidence identified from the surveillance review, studies referenced in identified systematic reviews, and references from included studies were also reviewed, which added a total of 15 references. An additional reference (Smith 2017) which was published after the date of the systematic search was identified by a member of the guideline committee. In total, 15,247 references were identified for screening at title and abstract level using priority screening. From the first 7,658 references screened, 7,506 were excluded based on their titles and abstracts and 152 references were ordered for screening based on their full texts. Based on the rules for using priority screening software (see appendix B), the screening was terminated at this point, and the remaining 7,589 were not screened on title and abstract.

Of the 152 references screened as full texts, 49 references were included for the 3 review questions based on their meeting the inclusion criteria specified in the review protocol (appendix A). The clinical evidence study selection is presented as a diagram in appendix C. Of the 49 included references, 5 presented data on

diagnostic accuracy of tests for the diagnosis of COPD and met the inclusion criteria for this review.

A second set of searches was conducted at the end of the guideline development process for all updated review questions using the original search strategies, to capture papers published whilst the guideline was being developed. These searches, which included articles up to February 2018, returned 3,100 references in total for all the questions included in the update, and these were screened on title and abstract. No additional relevant references were found for this review question.

The process of study identification is summarised in the diagram in appendix D.

For the full evidence tables and full GRADE profiles for included studies, please see appendix E and appendix G. The references of individual included studies are given in appendix K.

Excluded studies

Excluded studies are listed in appendix I, with reasons for their exclusion, and in appendix K as full references.

Summary of clinical studies included in the evidence review

The included studies are summarised in Table 2 and Table 3 below. See appendix E for full evidence tables.

Table 2 Summary table of included studies – systematic review

Author (year)	Study details	Index test (s)	Reference standard (s)	Outcomes
Li (2012) Countries of included studies were not reported	• Dates searched All of the databases were searched from their inception to October 2011. • Databases searched PUBMED, EMBASE, CNKI, VIP, CBM, WANFANG, The Cochrane Library. • Sources of funding Not stated.	• Chest CT	Pulmonary function tests	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio

Table 3 Summary table of included prospective cohort studies

Author (year)	Sample characteristics	Index test (s)	Reference standard(s)	Outcomes
Garcia-Pachon (2004) Spain	• Sample size: 210 • % female: 27% • Mean age (SD): 62 years (11) • Smoking status and history History of smoking of more than 20 pack–years in 110 participants • FEV1, % predicted (mean, SD) 103 participants had FEV1 value <80%	• Pulse oximetry (peripheral oxygen saturation, SpO2) % of arterial oxygen saturation: <96 <97 <98	• Post-bronchodilator spirometry in a stable patient COPD was defined as FEV1/FVC <0.70.	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio
Kurashima (2005) Japan	• Sample size: 516 • % female: 10.5% • Mean age (SD)	• Chest CT High resolution	• Clinical diagnosis of COPD	• Sensitivity • Specificity • Positive likelihood ratio

Author (year)	Sample characteristics	Index test (s)	Reference standard(s)	Outcomes
	69.0 years (0.1) • Smoking status and history Never smoked 10.9% Ex-smoker 79.3% Current smoker 9.8% • FEV1, % predicted (mean, SD) 58.6 (1.0)	thoracic CT	GOLD	• Negative likelihood ratio
Miniati (2011) Italy	• Sample size: 225 • % female: Derivation sample= 19% Validation sample= 44% • Median age (interquartile range [IQR]) Derivation sample= 65 years (46 to 70) Validation sample= 66 years (57 to 73)	• Chest X-ray Computer-aided procedure to recognise emphysema on digital chest X-ray	• CT demonstration of emphysema	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio
Tilemann (2011) Germany	• Sample size: 210 • %female Asthma 64% COPD 52.8% Partial reversibility 46.2% No obstructive airways disease (OAD) 58.7% • Mean age (SD) Asthma 38.0 years (14.6) COPD 56.8 years (11.7) Partial reversibility 57.9 years (11.2) No OAD 42.3 years (14.4) • Smoking status and history Asthma Current smokers 19.8% Past smokers 12.8%	• Systemic inflammatory markers including eosinophil count and high-sensitivity C-reactive protein concentrations (hs-CRP).	• Post-bronchodilator spirometry in a stable patient	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio

Author (year)	Sample characteristics	Index test (s)	Reference standard(s)	Outcomes
	Never smokers 67.4% COPD Current smokers 47.2% Past smokers 36.1% Never smokers 16.7% Partial reversibility Current smokers 61.5% Past smokers 23.1% Never smokers 15.4% No OAD Current smokers 28.0% Past smokers 12.0% Never smokers 60.0% • FEV1, % predicted (mean, SD) Asthma 99.7 (12.0) COPD 69.1 (17.1) Partial reversibility 67.6 (17.2) No OAD 106.3 (12.8)			

Quality assessment of clinical studies included in the evidence review

The systematic review and observational studies were assessed for risk of bias and applicability and this information is presented in the evidence tables in appendix E. See appendix G for full GRADE tables.

Economic evidence

Included studies

A single search was conducted to cover all review question topics in this guideline update. This search returned 16,299 records, all of which were excluded on title and abstract for this review question.

Evidence statements

The evidence statements based on likelihood ratios were written with reference to the size of the likelihood ratios in the GRADE tables in appendix G, using the interpretation detailed in the methods section on diagnostic test accuracy (Table 8) for both point estimates and confidence intervals. For example, positive likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate an increase in the probability of death and negative likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate a decrease in the probability of death in people with COPD. Evidence statements were grouped according to the size of the increase or decrease.

Clinical evidence statements

Confirming COPD diagnosis with computed tomography

Results that increase the probability of having COPD (based on positive likelihood ratios)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **very large**:

- Low-dose computed tomography with reference standard: emphysema index in expiration (very low quality, 95% CI goes from slight to very large)
- 16 Multi-slice computed tomography with reference standard: pixel index in maximum expiratory (low quality, 95% CI goes from moderate to very large)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **large**:

- 16 Multi-slice computed tomography with reference standard: full expiration average lung density, (low quality, 95% CI goes from moderate to very large)
- Computed tomography with reference standard: pulmonary function tests (very low quality, 95% CI goes from moderate to very large)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **moderate**:

- 16 Multi-slice computed tomography with reference standard: blood flow (very low quality, 95% CI goes from slight to large)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **slight**:

- High resolution computed tomography with reference standard: GOLD (very low quality, 95% CI goes from slight to moderate)

Results that decrease the probability of having COPD (based on negative likelihood ratios)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **very large**:

- 16 Multi-slice computed tomography with reference standard: pixel index in maximum expiratory (low quality, 95% CI goes from large to very large)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **large**:

- Low-dose computed tomography with reference standard: emphysema index in expiration (low quality, 95% CI goes from moderate to very large)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **moderate**:

- 16 Multi-slice computed tomography with reference standard: full expiration average lung density, (low quality, 95% CI goes from moderate to large)
- 16 Multi-slice computed tomography with reference standard: blood flow (very low quality, 95% CI goes from slight to large)
- Computed tomography with reference standard: pulmonary function tests (very low quality, 95% CI goes from moderate to large)
- High resolution computed tomography with reference standard: GOLD (low quality, 95% CI goes from moderate to moderate)

Confirming COPD diagnosis with chest X-ray (reference standard: computed tomography)

Results that increase the probability of having COPD (based on positive likelihood ratios)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **very large**:

- Computer-aided procedure to recognise emphysema on digital chest X-ray (low quality, 95% CI goes from large to very large)

Results that decrease the probability of having COPD (based on negative likelihood ratios)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **large**:

- Computer-aided procedure to recognise emphysema on digital chest X-ray (low quality, 95% CI goes from moderate to very large)

Confirming COPD diagnosis with pulse oximetry (reference standard: post-bronchodilator spirometry FEV1/FVC <0. 70)

Results that increase the probability of having COPD (based on positive likelihood ratios)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **moderate**:

- Pulse oximetry at cut-off arterial oxygen saturation <96% (very low quality, 95% CI goes from slight to moderate)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **slight**:

- Pulse oximetry at cut-off arterial oxygen saturation <97% (low quality, 95% CI goes from slight to slight)
- Pulse oximetry at cut-off arterial oxygen saturation <98% (low quality, 95% CI goes from slight to slight)

Results that decrease the probability of having COPD (based on negative likelihood ratios)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **slight**:

- Pulse oximetry at cut-off arterial oxygen saturation >96% (very low quality, 95% CI goes from slight to moderate)
- Pulse oximetry at cut-off arterial oxygen saturation >97% (very low quality, 95% CI goes from slight to moderate)
- Pulse oximetry at cut-off arterial oxygen saturation >98% (very low quality, 95% CI goes from slight to moderate)

Confirming COPD diagnosis with hs-CRP (reference standard: pulmonary function tests)

Results that increase the probability of having COPD (based on positive likelihood ratios)

The following positive test results **increase** the probability a person has COPD to a degree that is likely to be **moderate**:

- hs-CRP at 2.39mg/L (very low quality, 95% CI goes from slight to moderate)
- hs-CRP at 3.5mg/L (very low quality, 95% CI goes from slight to moderate)

Results that decrease the probability of having COPD (based on negative likelihood ratios)

The following negative test results **decrease** the probability a person has COPD to a degree that is likely to be **slight**:

- hs-CRP at 2.39mg/L (very low quality, 95% CI goes from slight to moderate)
- hs-CRP at 3.5mg/L (very low quality, 95% CI goes from slight to slight)

Economic evidence statements

No relevant economic evidence was identified for this review question.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that the critical outcome was whether the likelihood of a diagnosis of COPD was increased using a particular index test. For all index tests,

sensitivity, specificity, positive and negative likelihood ratios were used to identify the most accurate tests.

The quality of the evidence

The quality of the evidence ranged from very low to moderate. The reasons for downgrading the evidence were risk of bias (for example, due to uncertainty about whether the tests were interpreted independently and a lack of pre-specified thresholds); heterogeneity (inconsistency across studies); indirectness (lack of reporting on inclusion/exclusion criteria of participants); imprecision (95% CI crossed a defined MID).

The committee thought that all of the reference standards were acceptable, but discussed the validity of the index tests in detail. The committee highlighted that some of the techniques included in the CT studies were no longer up to date, such as the 16 multi-slice CT. Regarding computer-aided chest X-ray, the committee highlighted that this technique is not used in clinical practice and that the computer-aided procedure to recognise emphysema on a digital chest X-ray was developed and tested for research and it has not been tested in routine clinical practice.

The committee discussed the use of pulse oximetry and hs-CRP as diagnostic tests for COPD. They noted that pulse oximetry is normally used as a measure of severity of COPD, rather than during diagnosis. In addition, there are many other causes of low oxygen saturation that are not related to COPD or that occur in patients with COPD independently of this disease. The committee noted that raised hs-CRP could have many causes and that the absence of a specific link to COPD meant that it was not suitable for use as a diagnostic test for COPD. The committee accepted that studies examining pulse oximetry and hs-CRP were included in this evidence review because they met the review protocol criteria and there were clinical trials testing them in the context of COPD diagnosis. However, based on the issues discussed above and the evidence for low diagnostic utility of these tests, the committee decided not to recommend pulse oximetry or hs-CRP for the diagnosis of COPD.

Benefits and harms

The committee agreed that that CT scans and chest X-rays are accurate tests for identifying people who would test positive for COPD using spirometry, including people without symptoms. The committee noted that it is possible for asymptomatic patients to have a CT scan or chest X-ray as part of another investigation that shows emphysema or signs or symptoms of chronic airways disease. Therefore, the committee agreed to recommend considering spirometry and GP respiratory review for patients with emphysema or signs of chronic airways disease detected by incidental CT scan or chest X-ray.

Although there was no evidence on what to do for people who have emphysema or signs of chronic airways disease on a CT scan or chest X-ray, but who have no symptoms, the committee thought it was important to provide some guidance on this matter based on their experience and on current practice in the NHS. They made separate recommendations for current smokers and current non-smokers to reflect their different levels of risk of developing lung disease and to ensure that other relevant factors were taken into account, such as a personal or family history of lung or liver disease, which could explain the test results, and the increase in the risk of lung cancer for people with emphysema. The committee made a research recommendation to provide information about the characteristics of people diagnosed with COPD based on a CT scan to try to determine whether people identified in this manner are sufficiently similar to those diagnosed in the usual way to ensure that they will benefit from the same treatment pathways. If they are found to differ in

important characteristics then different treatments or treatment pathways may be required for this group of people with COPD.

The committee also reviewed the list of additional investigations and changed the order of these investigations to better reflect current practice. The committee moved transfer factor for carbon monoxide (TLCO) to the end because this test is not available in primary care, unlike the other investigations. The committee decided to add 'serum' to alpha-1 antitrypsin to clarify the type of test required. They thought it was important to add another role for a CT scan of the thorax relating to the investigation of signs that may suggest another lung diagnosis. The committee decided to replace the word 'surgery' with 'lung volume reduction' because bronchoscopic techniques for lung volume reduction are now available. They agreed that ECG and serum natriuretic peptides were also used to assess the clinical suspicion of cardiac disease and pulmonary hypertension, so this was added to the role of these investigations. The committee also decided to add that TLCO could be used as another test of suitability for lung volume reduction. Pulse oximetry was removed from this list because its main role is to assess and monitor exacerbations (see recommendations 1.3.2 and 1.3.42 in the short guideline), COPD disease progression and to assess the need for oxygen therapy (see recommendations 1.2.61, 1.2.63, and 1.3.32 in the short guideline).

The committee highlighted that the new recommendation should not trigger substantial additional tests because suspicion of COPD is usually based on signs and symptoms. CT scan and chest X-ray showing emphysema or signs of chronic airways disease can be incidental findings in asymptomatic patients having investigations for other conditions.

Cost effectiveness and resource use

The committee considered the cost effectiveness of carrying out spirometry in patients with signs of COPD on CT scan or chest X-ray. It was concluded that spirometry in this scenario is likely to be cost effective, considering its low cost compared with the downstream benefits of correctly diagnosing COPD. The committee was confident in recommending smoking cessation advice, treatment and services for current smokers, as the cost effectiveness of such interventions has been demonstrated in previous guidance (Stop smoking interventions and services [NG92]).

The committee determined that, in the appropriate circumstances, serial domiciliary peak flow measurements, CT scan of the thorax, ECG or assessing serum natriuretic peptides, echocardiogram, sputum culture, and transfer factor for carbon monoxide and lung volumes, are likely to be cost-effective investigations, given the balance of benefits provided from a correct diagnosis compared with the cost of investigation.

The committee concluded that the recommendations are in-line with current practice, and are therefore unlikely to have a significant impact on resource use.

Other factors the committee took into account

The committee discussed that it would be useful to have guidance for radiologists on how to report emphysema and that this should include a recommendation to refer for spirometry and GP respiratory review if emphysema or signs of chronic airways disease are found. Whilst this was outside the scope of the guideline, they agreed it would be an appropriate route for the recommendations to be implemented.

Predicting outcomes for people with a new diagnosis of COPD

Review question

In people with suspected COPD, which tests (for example imaging or biomarkers) are the most accurate to identify whether they are at risk of poor outcomes and whether they will develop mild, moderate or severe COPD?

Introduction

This review aims to determine the prognostic accuracy of tests to predict outcomes in people with COPD at the point of diagnosis. At this stage, accurate disease prognosis could help physicians tailor the appropriate level of monitoring and treatment for a person with COPD, with the goal of achieving improved outcomes. Among the tests listed in Table 4, multidimensional indices have been shown to be important predictors of outcome for people with an existing diagnosis of COPD and may also be of value at the point of diagnosis.

This review identified studies that fulfilled the conditions specified in Table 4. (For full details of the review protocol, see appendix A.)

Table 4 PICO: prognosis for people with newly diagnosed COPD

Population	People with a new diagnosis of COPD (by any means including Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, guideline; American Thoracic Society criteria for COPD; European Respiratory Society criteria)

Assessment tools	• Imaging including CT, 18F-FDG-PET
	• BMI
	• Biomarkers
	• MRC dyspnoea (breathlessness) tool/ Borg dyspnoea (breathlessness) score
	• Multidimensional assessment indices including:
	  ○ BODE
	  ○ CAT (self-administered COPD assessment test)
	  ○ GOLD
	  ○ DECAF (hospital based for acute exacerbations and pneumonia in COPD)
	  ○ DOSE
	  ○ COPD Diagnostic Questionnaire
	• Polycythaemia (full blood count, FBC)
	• Oxygen saturation (SaO2)
	• 6 minute walk distance (6MWD)
	• Tests for anxiety (e.g. General anxiety disorder 7, GAD7; Hospital Anxiety and Depression Scale, HADS)
	• Tests for depression (e.g. patient health questionnaire 9, PHQ9; Hospital Anxiety and Depression Scale, HADS)

Outcomes	• Mortality • Hospitalisations (no hospitalisation versus hospitalisation) • Exacerbations (exacerbations versus no exacerbations)

	• Severity of COPD (as defined by Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, 2017 and NICE clinical guideline 101 (2010), based on predicted airflow limitation (FEV1 %) in patients with FEV1/FVC <0.70. This will be scored as mild versus not mild (moderate and severe), moderate versus severe and severe versus not severe (mild and moderate) as the data permits.
Measures	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A, and the methods section in appendix B.

Subgroup analyses were not carried out for this review because the included studies did not report data for the categories of interest in an accessible format.

The search strategies used in this review are detailed in appendix C.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

Included studies

A single systematic search was carried out for the 3 review questions in this evidence review to identify observational studies and systematic reviews of observational studies, which found 15,231 references (see appendix C for literature search strategy). Evidence included in the original guideline, evidence identified from the surveillance review, studies referenced in identified systematic reviews, and references from included studies were also reviewed, which added a total of 15 references. An additional reference (Smith 2017) which was published after the date of the systematic search was identified by a member of the guideline committee. In total, 15,247 references were identified for screening at title and abstract level. From the first 7,658 references screened, 7,506 were excluded based on their titles and abstracts and 152 references were ordered for screening based on their full texts. Based on the rules for using priority screening software (see appendix B), the screening was terminated at this point, and the remaining 7,589 were not screened on title and abstract.

Of the 152 references screened as full texts, 49 references were included for the 3 review questions based on their meeting the inclusion criteria specified in the review protocol (appendix A). The clinical evidence study selection is presented as a diagram in appendix C. Of the 49 included references, none was deemed relevant for this review question.

A second set of searches was conducted at the end of the guideline development process for all updated review questions using the original search strategies, to capture papers published whilst the guideline was being developed. These searches returned 3,100 references in total for all the questions included in the update, and

these were screened on title and abstract. No additional relevant references were found for this review question.

The process of study identification is summarised in the diagram in appendix D.

Excluded studies

Excluded studies are listed in appendix I, with reasons for their exclusion, and in appendix K as full references.

Economic evidence

Included studies

A single search was conducted to cover all review question topics in this guideline update. This search returned 16,299 records, all of which were excluded on title and abstract for this review question.

Evidence statements

Clinical evidence statements

No relevant evidence was identified for this review question.

Economic evidence statements

No relevant economic evidence was identified for this review question.

The committee's discussion of the evidence

No evidence was found specifically addressing prognosis at the point of diagnosis. The committee therefore made recommendations for both review questions in the predicting outcomes using multidimensional indices in people with stable COPD section. Discussions on predicting outcomes at the time of diagnosis are contained within the section on prognosis in people with stable COPD.

Predicting outcomes using multidimensional severity assessment indices

Review question

In people with stable COPD, does routine assessment using a multidimensional severity assessment index (such as BODE [BMI, airflow obstruction, dyspnoea/ breathlessness and exercise capacity]) better predict outcomes than forced expiratory volume in 1 second (FEV1) alone?

Introduction

Accurate disease prognosis could help clinicians tailor the appropriate level of monitoring and treatment for a person with COPD, with the goal of achieving improved outcomes. FEV1 status has been shown to be predictive of outcome in COPD, but the inclusion of additional or alternative measures may improve its prognostic ability. Multidimensional indices, which by definition measure multiple domains, have been shown to be important predictors of outcome for people with COPD. Depending on the index these may include: breathlessness; exercise capacity; airway obstruction; depression; body mass index (BMI); sleep and smoking status. This review aims to determine the prognostic ability of these indices in comparison to FEV1 (including the GOLD classification schemes) in people with an existing diagnosis of COPD.

This review identified studies that fulfilled the conditions specified in Table 5. (For full details of the review protocol, see appendix A.) In particular, studies recruiting people with COPD from hospital during or immediately following an exacerbation were excluded as their condition was considered to be unstable. For the purpose of this review, people with stable COPD were defined as being free from an exacerbation for at least one month. However, studies that recruited people with COPD from the community, hospital out-patient clinics or from primary care were included even if they failed to specify that participants were exacerbation free for this amount of time.

Prognostic indices that covered multiple measures (multivariable indices) were not included unless the measures also covered a number of separate domains. Finally, indices involving biomarkers were excluded as these would require the physician to request additional tests over and above routine information that would be available regarding a person with COPD.

Table 5 PICO: prognosis for people with an existing diagnosis of COPD

Population	People diagnosed with COPD

Assessment tools	• FEV1 alone
	• Multidimensional assessment indices including:
	  ○ BODE
	  ○ CAT (self-administered COPD assessment test)
	  ○ GOLD
	  ○ DECAF (hospital based for acute exacerbations and pneumonia in COPD)
	  ○ DOSE
	  ○ COPD Diagnostic Questionnaire

Outcomes	• Mortality • Hospitalisations • Exacerbations • Change in FEV1
Measures	• Sensitivity/specificity (preferred outcomes) • c-statistic, • Hazard ratios • Model fit (e.g. r-squared)

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A, and the methods section in appendix B. The categories used to assess c-statistic test classification accuracy range from poor to outstanding and are detailed in the methods section under other prognostic evidence.

Meta-analysis of the c-statistic data was not carried out for this review due to the absence of 95% CI for the majority of studies. Hazard ratio data was also not meta-analysed as the models were not adjusted for the same potential confounding factors and data fell into several formats (per point increase, compared to a low severity reference category or compared to a high severity reference category). Instead, a number of decision rules were used to analyse data for c-statistics and hazard ratios across multiple studies. These are detailed in the methods section in appendix B.

The majority of the proposed subgroup analyses were not carried out for this review because the included studies did not report data for the categories of interest in an accessible format. However, data was available for severe exacerbations (including moderate to severe and severe exacerbations as one group) and this is presented as a separate analysis.

The search strategies used in this review are detailed in appendix C.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Deviation from review protocol

In a deviation from the review protocol, studies that presented multivariate prognostic models were included in this review even if they failed to adjust for age, smoking status and comorbidities or the adjusted confounders were not specified. This was due to the small number of fully adjusted studies that were identified. The limitations of these studies were discussed with the Committee.

Clinical evidence

Included studies

A single systematic search was carried out for the 3 review question in this evidence review to identify observational studies and systematic reviews of observational studies, which found 15,231 references (see appendix C for literature search strategy). Evidence included in the original guideline, evidence identified from the surveillance review, studies referenced in identified systematic reviews, and references from included studies were also reviewed, which added a total of 15

references. An additional reference (Smith 2017) which was published after the date of the systematic search was identified by a member of the guideline committee. In total, 15,247 references were identified for screening at title and abstract level. From the first 7,658 references screened, 7,506 were excluded based on their titles and abstracts and 152 references were ordered for screening based on their full texts. Based on the rules for using priority screening software (see appendix B), the screening was terminated at this point, and the remaining 7,589 were not screened on title and abstract.

Of the 152 references screened as full texts, 49 references were included for the 3 review questions based on their meeting the inclusion criteria specified in the review protocol (appendix A). The clinical evidence study selection is presented as a diagram in appendix C. Of the 49 included references, 44 were deemed relevant for this review question. These included papers presenting data on multiple prospective cohorts (e.g. Marin 2013).

A second set of searches was conducted at the end of the guideline development process for all updated review questions using the original search strategies, to capture papers published whilst the guideline was being developed. These searches returned 3,100 references in total for all the questions included in the update, and these were screened on title and abstract. Two papers were identified as being potentially relevant for this review question, but they were excluded after full text screening.

The process of study identification is summarised in the diagram in appendix D.

For the full evidence tables and full GRADE profiles for included studies, please see appendix E and appendix G. The references of individual included studies are given in appendix K.

Excluded studies

The excluded studies are listed with reasons for their exclusion in appendix I, and as full references in appendix K.

Summary of clinical studies included in the evidence review

The included studies are summarised in Table 6 below. (Please refer to appendix E for full evidence tables.) The included prospective cohort studies covered several prognostic indices including BODE, DOSE, HADO and SAFE; and variations on these indices such as i-BODE, BODEx, HADO-AH. Other multidimensional measures that could be used for prognosis such as CCQ, CAT, COPD severity score, GOLD 2011, and 2013 were also reported. GOLD prior to 2011 and FEV1 were reported as comparators.

Table 6 Summary table of included studies. The table only includes outcomes, measures and indices that were analysed in the GRADE tables. Refer to the evidence tables for details of all of the other measures, outcomes and indices included in the studies.

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
Andrianopoulos (2015)	• BODE index	• c-statistic • Sensitivity and specificity • Hazard ratios	• Mortality • Hospitalisations	• Age • Smoking status • Gender • Body Mass Index (BMI) • FEV1 %, predicted • SGRQ (St George's Respiratory Questionnaire total score) • FEV1/FVC ratio • Emphysema
Ansari (2016)	• BODS index • BODAS index • BOD index	• c-statistic • Sensitivity and specificity	• Mortality	N/A
Casanova (2005)	• BODE index	• c-statistic	• Mortality	• N/A
Casanova (2015)	• CCQ • CAT	• c-statistic	• Mortality	• N/A
Celli (2004)	• BODE index • FEV1	• c-statistic • Hazard ratios	• Mortality	• Comorbidities

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
Chan (2016)	• BOD index • GOLD 2011 • GOLD 2007	• c-statistic • Hazard ratios	• Mortality • Exacerbations	• Unspecified
Chan (2017)	• BOSA	• c-statistic • Hazard ratios	• Mortality	• Gender • Race
Chen (2015a and b)	• BODEx index • GOLD 2013 • GOLD 2007	• c-statistic	• Mortality • Exacerbations	• N/A
Cote (2008)	• BODE index • mBODE%	• c-statistic	• Mortality	• N/A
de Torres (2008)	• BODE index	• Hazard ratios	• Mortality	• Age • Smoking (pack years) • Presence of cardiovascular risk factors or disease • Treatment with inhaled corticosteroids
de Torres (2014)	• BODE index • GOLD 2011 • BODE and COTE combined	• c-statistic	• Mortality	• N/A
Divo (2012)	• BODE and COTE combined	• c-statistic	• Mortality	• N/A
Eisner (2010)	COPD severity score	• Hazard ratios	• Hospitalisations	• Age • Gender • Race • Smoking history • Educational attainment
Esteban (2006)	• HADO score	• c-statistic • Hazard ratios	• Mortality	• Age • Smoking status

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
Esteban (2010)	• BODE index • HADO score	• c-statistic • Odds ratios	• Mortality	• Age • Smoking (pack years) • Comorbidities • Number of hospitalisations in the previous year
Esteban (2011)	• BODS index • ADO index • HADO score • HADO-AH index	• c-statistic	• Mortality	• N/A
Faganello (2010)	• BODE index	• c-statistic • Sensitivity and specificity	• Exacerbations.	• Age • Smoking status • Smoking (pack years) • GOLD stage • 6 MWD (6 minute walk distance) • mMRC dyspnoea/breathlessness • SGRQ (St George's Respiratory Questionnaire total score) • SpO2 (Peripheral oxygen saturation)
Goossens (2014)	• GOLD A-D (probably GOLD 2011)	• c-statistic	• Mortality	• N/A
Imfeld (2006)	• BODE index • FEV1	• c-statistic	• Mortality	• N/A
Johannessen (2013)	• GOLD 2011 • GOLD 2007	• c-statistic • Hazard ratios	• Mortality • Hospitalisations	• Age • Smoking status • Comorbidities

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
				• Gender • Body Mass Index (BMI)
Lee (2014)	• GOLD (not specified) *Stages 1-4* • CAT (COPD Assessment Test) *Categories: 0-9, 10-19, 20-29, 30-40.*	• c-statistic • Sensitivity and specificity • Hazard ratios	• Exacerbations	• Age • Smoking status • Number of comorbidities • Country • Number of exacerbations in the previous year • Gender • Body Mass Index (BMI) • Influenza vaccination • Duration of COPD • GOLD stage
Leivseth (2013)	• GOLD (not specified) *1-4 severity grouping* • GOLD 2011	• Hazard ratios	• Mortality	• Age • Smoking status • Educational attainment
Marin (2009)	• BODE index	• c-statistic • Sensitivity and specificity	• Exacerbations	• N/A
Marin (2011)	• BODE index	• c-statistic	• Mortality	• N/A
Marin (2013)	• BODE index • e-BODE • DOSE index • ADO index • HADO score	• c-statistic	• Mortality	• N/A

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
	• SAFE index cardiovascular			
Mattila (2015)	• GOLD (not specified).	• Hazard ratios	• Mortality	• Age • Smoking status • Gender
Moberg (2014)	• i-BODE	• Hazard ratios	• Mortality • Hospitalisations	• Age • Smoking status • Smoking (pack years) • Gender • Oxygen saturation at rest • Desaturation >4% during shuttle walking test (SWT) • Maintenance prednisolone • LTOT
Motegi (2013)	• BODE index • DOSE index • ADO index • GOLD 2007 (stage 1-4)	• c-statistic	• Exacerbations	• N/A
Moy (2014)	• BODE index	• c-statistic	• Hospitalisations • Exacerbations	• N/A
Neo (2017)	• BODE index • ADO index	• c-statistic	• Mortality	• N/A
Omachi (2008)	• COPD severity score	• c-statistic	• Hospitalisations	• Age • Smoking history • Comorbidities • Race • Educational attainment

Author (year)	Relevant prognostic factor(s)	Measures	Outcome(s)	Multivariate regression model adjusted covariates
Ou (2014)	• BODEx index •CPI	• c-statistic	• Mortality	• N/A
Pedone (2014)	• BODE index	• c-statistic	• Mortality	• N/A
Pothirat (2015)	• CAT	• c-statistic	• Exacerbations	• N/A
Puhan (2009)	• BODE index	• c-statistic	• Mortality	• N/A
Soler-Cataluna (2009)	• BODE index • e-BODE	• Hazard ratios	• Mortality	• Comorbidities • Exacerbation frequency • Blood gases *PaO2, PaCO2*
Stolz (2014a)	• BODE index	• c-statistic	• Mortality	•N/A
Stolz (2014b)	• BOD index	• c-statistic	• Mortality	•N/A
Suetomo (2014)	• CAT • GOLD 2009	• Sensitivity and specificity	• Hospitalisations • Exacerbations	•N/A
Sundh (2012a and 2012b)	• CCQ • DOSE index	• Hazard ratios	• Mortality	• Age • Comorbidities • Gender
Thabut (2014)	• BODE index	• c-statistic • Hazard ratios	• Mortality	• Augmentation therapy and centre
Varol (2014)	• CAT	• c-statistic	• Exacerbations	•N/A

Quality assessment of clinical studies included in the evidence review

The included prognostic cohort studies were assessed for risk of bias and applicability and this information is presented in the evidence tables in appendix E. Please refer to appendix G for GRADE tables.

Economic evidence

Included studies

A single search was conducted to cover all review question topics in this guideline update. This search returned 16,299 records, all of which were excluded on title and abstract for this review question.

Evidence statements

The evidence statements based on likelihood ratios were written with reference to the size of the likelihood ratios in the GRADE tables in appendix G, using the interpretation detailed in the methods section on diagnostic test accuracy (Table 8) for both point estimates and confidence intervals. For example, positive likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate an increase in the probability of death and negative likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate a decrease in the probability of death in people with COPD. Evidence statements were grouped according to the size of the increase or decrease.

The format of the evidence statements for c-statistic and HR data are explained in the methods in appendix B.

Clinical evidence statements

All-cause mortality

Sensitivity, specificity and likelihood ratios

Results that increase the probability of death at an average follow-up of 5 years (based on positive likelihood ratios)

The following test results **increase** the probability of death to a degree that is likely to be **moderate**:

- BODE >4 (moderate quality, 95% CI goes from slight to moderate)

The following test results **increase** the probability of death to a degree that is likely to be **slight**:

- BODE ≥4 (moderate quality, 95% CI goes from slight to slight)
- BODAS >5 (very low quality, 95% CI goes from slight to moderate)
- BOD >2 (very low quality, 95% CI goes from slight to moderate)
- BODS >4 (low quality, 95% CI goes from slight to slight)
- GOLD >1 (matrix [new classification A to D]) (low quality, 95% CI goes from slight to slight)
- GOLD >2 (old GOLD stages 1 to 4) (low quality, 95% CI goes from slight to slight)
- ADO >3 (very low quality, 95% CI goes from slight to moderate)

Results that decrease the probability of death (based on negative likelihood ratios)

The following results **decrease** the probability of death to a degree that is likely to be **moderate**:

- BODAS ≤5 (very low quality, 95% CI goes from slight to moderate)
- GOLD ≤1 (matrix [new classification A to D]) (very low quality, 95% CI goes from slight to large)

The following results **decrease** the probability of death to a degree that is likely to be **slight**:

- BODE <4 (moderate quality, 95% CI goes from slight to slight)
- BODE ≤4 (moderate quality, 95% CI goes from slight to moderate)
- BOD ≤2 (very low quality, 95% CI goes from slight to slight)
- BODS ≤4 (very low quality, 95% CI goes from slight to moderate)
- GOLD ≤2 (old GOLD stages 1 to 4) (very low quality, 95% CI goes from slight to slight)
- ADO ≤3 (very low quality, 95% CI goes from slight to moderate)

c-statistics

Test with good median classification accuracy

- BODAS index (very low quality, range from adequate to good)
- BODE and COTE (high quality)
- Clinical basic model (age-adjusted Charlson comorbidity score, sex, FEV1 % predicted and smoking status, low quality)
- CPI (high quality)
- GOLD 2011 (very low quality, range from poor to excellent)
- HADO index (moderate quality, range from adequate to good)
- HADO-AH index (moderate quality, range from good to excellent)

Tests with adequate median classification accuracy

- ADO index (very low quality, range from poor to excellent)
- BOD index (low quality, range from adequate to good)
- BODE index (very low quality, range from poor to excellent)
- BODE index ≥4 (high quality)
- BODEx index (very low quality, range from poor to good)
- BODS index (very low quality, range from adequate to good)
- e-BODE index (very low quality, range from adequate to good)
- BOSA index (low quality, range from adequate to good)
- DOSE index (very low quality, range from poor to good)
- FEV1 % predicted (very low quality, range from poor to adequate)
- GOLD stages 2-4 (high quality)
- GOLD 2007 (very low, range from poor to excellent)
- GOLD 2013 (moderate quality, range from adequate to good)
- mBODE% (moderate quality, range from adequate to good)

- SAFE index (low quality)

Tests with poor median classification accuracy

- CAT (high quality)
- CCQ (high quality)
- GOLD stages A-D (moderate quality, range from poor to adequate)

Hazard ratios

The following instruments reported data on hazard ratios per unit increase on a scale and are ordered from largest to smallest:

- BODE index (low quality)
- i-BODE index (moderate quality)

The following instruments were reported using comparison of groups to a reference group. They are ranked in order of the largest to smallest (based on the median most extreme category hazard ratio, with data reversed where necessary so the comparisons go in the same direction):

- e-BODE (high quality)
- DOSE (moderate quality)
- BODEx (high quality)
- BOD (low quality)
- BOSA (moderate quality)
- HADO (high quality)
- GOLD before 2011 (low quality)
- CCQ (high quality)
- GOLD 2011 (low quality)
- BODE (moderate quality)

Mortality due to respiratory causes

c-statistics

Tests with an excellent median classification accuracy

- BODE index (low quality, range from good to outstanding).
- GOLD 2011 (high quality)
- HADO index (moderate quality, range from excellent to outstanding)

Test with a good median classification accuracy

- GOLD 2007 (very low quality, range from poor to excellent)
- GOLD 2013 (moderate quality, range from adequate to good)

Tests with an adequate median classification accuracy

- FEV1 (% predicted) (high quality)

Hazard ratios

The following instrument reported data on hazard ratios per unit increase on a scale:

- BODE (high quality)

The following instruments were reported using comparison of groups to a reference group. They are ranked in order of the largest to smallest (based on the median most extreme category hazard ratio):

- GOLD 2007 (low quality)
- GOLD 2011 (high quality)

All-cause hospitalisations

Sensitivity, specificity and likelihood ratios

Results that increase the probability of hospitalisations at an average follow-up of 2 years (based on positive likelihood ratios)

The following test results **increase** the probability of hospitalisations to a degree that is likely to be **moderate**:

- GOLD stages III and IV (moderate quality, 95% CI goes from slight to moderate)

The following test results **increase** the probability of hospitalisations to a degree that is likely to be **slight**:

- BODE ≥3 (moderate quality, 95% CI goes from slight to moderate)
- CAT ≥10 (moderate quality, 95% CI goes from slight to moderate)

Results that decrease the probability of hospitalisations (based on negative likelihood ratios)

The following results **decrease** the probability of hospitalisations to a degree that is likely to be **slight**:

- BODE <3 (moderate quality, 95% CI goes from slight to slight))

The following results were **not significantly different** from random chance:

- CAT <10 (moderate quality, 95% CI goes from large decrease to moderate increase)
- GOLD stages I and II (moderate quality, 95% CI goes from moderate decrease to slight increase)

c-statistics

Tests with a good median classification accuracy

- GOLD 2007 (high quality)
- GOLD 2013 (high quality)

Tests with an adequate median classification accuracy

- BODE ≥3 (high quality)

Hazard ratios

The following instruments reported data on hazard ratios per unit increase on a scale and are ordered from largest to smallest:

- COPD severity score (low quality)
- i-BODE (moderate quality)

The following instruments were reported using comparison of groups to a reference group. They are ranked in order of the largest to smallest (based on the median most extreme category hazard ratio):

- GOLD 2007 (high quality) and GOLD 2011 (high quality) (identical HRs)
- BODE (moderate quality)

Respiratory specific hospitalisations

c-statistics

Tests with an outstanding median classification accuracy

- Model 2 (age, race, educational attainment, tobacco history, and medical comorbidities (heart failure, coronary artery disease, diabetes, and sleep apnoea with COPD severity score) (moderate quality)

Test with a good median classification accuracy

- Model 1 (age, race, educational attainment, tobacco history, and medical comorbidities (heart failure, coronary artery disease, diabetes, and sleep apnoea) (moderate quality)

Tests with an adequate median classification accuracy

- BODEX index (low quality)

Hazard ratios

The following instrument reported data on hazard ratios per unit increase on a scale:

- i-BODE (high quality)

The following instruments were reported using comparison of groups to a reference group. They are ranked in order of the largest to smallest (based on the median most extreme category hazard ratio):

- GOLD 2011 (high quality)
- GOLD 2007 (high quality)

Exacerbations

Sensitivity, specificity and likelihood ratios

Results that increase the probability of exacerbations at an average follow-up of 2 years and 6 months (based on positive likelihood ratios)

The following test results **increase** the probability of exacerbations to a degree that is likely to be **moderate**:

- BODE index >1.9 (moderate quality, 95% CI goes from slight to large)
- GOLD 2003 stage III (moderate quality, 95% CI goes from slight to moderate)
- CAT ≥10 (moderate quality, 95% CI goes from slight to moderate)
- GOLD stages III and IV (moderate quality, 95% CI goes from slight to large)

The following test results **increase** the probability of exacerbations to a degree that is likely to be **slight**:

- CAT (cut-off 17/40) (moderate quality, 95% CI goes from slight to moderate)
- BODE class II (stages 3 to 4) (moderate quality, 95% CI goes from slight to moderate)

Results that decrease the probability of exacerbations (based on negative likelihood ratios)

The following results **decrease** the probability of exacerbations to a degree that is likely to be **moderate**:

- BODE index <1.9 (high quality, 95% CI goes from moderate to moderate)
- CAT <10 (moderate quality, 95% CI goes from slight to moderate)

The following results **decrease** the probability of exacerbations to a degree that is likely to be **slight**:

- CAT <17 (moderate quality, 95% CI goes from slight to slight)
- BODE class I (score 0 to 2) (moderate quality, 95% CI goes from slight to moderate)
- GOLD 2003 stage I and II (moderate quality, 95% CI goes from slight to slight)
- GOLD stages I and II (moderate quality, 95% CI goes from slight to moderate)

c-statistics

Test with a good median classification accuracy

- BODEx (low quality, range adequate to good)
- CAT (low quality, range from good to excellent)
- DOSE index (low quality, range adequate to excellent)
- GOLD 2012 (low quality, range from good to excellent)

Tests with an adequate median classification accuracy

- ADO index (low quality, range poor to good)
- BOD index (moderate quality)
- BODE index (very low quality, range adequate to excellent)
- BODE index Stage 3-4 (low quality)
- GOLD stages 1-4 (low quality, range poor to good)
- GOLD 2003 stage III (low quality)
- GOLD 2007 (low quality, range from poor to good)
- GOLD 2011 (moderate quality)

Hazard ratios

The following instruments were reported using comparison of groups to a reference group. They are ranked in order of the largest to smallest (based on the median most extreme category hazard ratio, with data reversed where necessary so the comparisons go in the same direction):

- GOLD 2007 (moderate quality)
- GOLD 2011 (moderate quality)
- BOD index (moderate quality)
- CAT (moderate quality)

Severe exacerbations

c-statistics

Tests with an excellent median classification accuracy

- BODE index (moderate quality, range from excellent to outstanding)

Test with a good median classification accuracy

- CAT (moderate quality, range from adequate to good)
- GOLD 2013 (high quality, range from good to excellent)

Tests with an adequate median classification accuracy

- GOLD 2007 (low quality, range from adequate to good)

Hazard ratios

The following instrument was reported using comparison of groups to a reference group:

- CAT (high quality)

Economic evidence statements

No relevant economic evidence was identified for this review question.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that the most important outcome from the perspective of someone with COPD was survival time (mortality), followed by hospitalisations and severe exacerbations, which negatively affect quality of life and reducing the impact of breathlessness. The distinction between mortality and hospitalisations due to all-causes or respiratory events was not considered helpful as this distinction did not matter to people with COPD.

This review used 3 groups of measures to assess prognostic accuracy of tests: sensitivity, specificity and likelihood ratios; c-statistics and hazard ratios. Sensitivity, specificity and likelihood ratios were considered to be the highest standard of evidence available as they involve evaluating the performance of a measure using a specified threshold, and link directly to decision making. However, the limited amount of this data available meant that c-statistics and hazard ratios were the primary outcomes used to assess the performance of multidimensional indices in predicting outcomes for people with COPD. They provide an indication of classification accuracy and the risk of an event associated with the classification.

The quality of the evidence

The committee discussed the multiple tests and indices that were included in the review. They agreed that for an index to be useful in a primary care context it needed to be easily administered and consist of components that were easy and quick to assess during a consultation or were readily available in the medical records of people with COPD. The committee agreed with the inclusion of the majority of tests reported in this review, however, mBODE% was excluded from consideration as one of the components (oxygen uptake measured at peak exercise, VO2) is not routinely assessed. They noted that the COPD Assessment test (CAT) and Clinical COPD Questionnaire (CCQ) were not multi-dimensional as they examined the single domain of health status. As such, they agreed to exclude the data on these tests as they did not fit the review protocol closely enough. The committee also agreed that DECAF was not a suitable multidimensional index to include in this review for

prognosis in a person with stable COPD, since it is designed for use only in a hospital.

The committee discussed the role of c-statistics in determining prognostic test classification accuracy and tried to establish what value could be considered clinically useful in the context of prognostic tests in general and specifically for COPD. The committee agreed that values closest to 1 are most useful, with values of around 0.5 having little prognostic value. They considered other prognostic tests in clinical use, such as the QRISK2 test which is used for cardiovascular disease. Based on these discussions, the committee agreed that a c-statistic of > 0.75 could indicate a useful test for prognosis in COPD.

Classification accuracy alone is unhelpful in predicting outcomes such as mortality. The committee agreed that for a test to be clinically useful it needed to have a good classification accuracy coupled with a larger risk of mortality, shown by the hazard ratio, in more severe groups compared to less severe groups, or with increasing points (severity) on the test. This was decided as in practice, disease prognosis impacts and discussions are focused on people with severe and very severe COPD.

The data on c-statistics could not be meta-analysed due to the lack of 95% CI in multiple studies. The committee agreed with the approach used to synthesis the data using medians as a measure of central tendency and noted the uncertainty surrounding the original point estimates that lacked 95% CI and the resulting pooled estimates. The committee also noted the difficulties in comparing hazard ratio data across studies where the multivariate regression models were adjusted for different covariates (such as smoking status, age and comorbidities). They confirmed that the aforementioned factors were especially important prognostic factors for COPD and could confound the results if not taken into account. However, they decided to look at the evidence on indices from models that were partially adjusted, as there were few papers with fully adjusted models, and then look in detail at which factors were included in the adjustments for any potentially useful indices.

The committee commented that the size of many of the trials reported was very small in comparison to prognostic cohorts for other diseases, which may have many thousands of participants. This will be associated with reduced certainty in the results from the COPD trials, especially when sample sizes were < 500. The committee agreed with the use of a threshold of 500 participants to downgrade a trial for imprecision.

The committee agreed that data on FEV1 data should not be merged with GOLD 2007, 1-4 or GOLD before 2011 data as they were not completely equivalent, but did provide useful comparators to the multidimensional indices. In the context of this review, GOLD 2011 was considered to be a multidimensional index as it included exacerbations, hospitalisations and CAT or breathlessness scores. The committee agreed that GOLD 2011 A-D categorisation was not as useful as the older GOLD 1-4 system in predicting outcomes as the FEV1 component from GOLD 1-4 was omitted and A-D categories did not directly correspond to increasing severity of disease. GOLD 2017 reintroduced consideration of FEV, but the prognostic usefulness of this categorisation was not examined here as the only relevant study identified was retrospective, while this review was limited to including prospective cohort studies.

The committee noted that the c-statistics for respiratory mortality were larger than for all-cause mortality and that this was not surprising as the indices were using components known to predict or linked to respiratory events. They also noted that it was not surprising that indices that included exacerbations or hospitalisations were

better at predicting these events as, for example, having frequent exacerbations is known to be linked with an increased risk of more exacerbations in the future.

Benefits and harms

The committee noted that the clinical usefulness of c-statistic data varied across tests and diseases. In other situations, such as predicting people at risk of stroke, a positive test result is linked to a change in treatment. However, in the case of COPD, being classified accurately into a group of people who are at higher risk of death is less helpful if there is no change in treatment to follow after classification, although it may prove important in enabling advanced care planning conversations to be offered. They noted that the 2010 recommendation for the use of BODE is not linked to a course of action based on the information obtained.

The committee discussed what actions could be taken to reduce risk identified by a multidimensional index such as DOSE or HADO-AH that contain potentially modifiable components. The DOSE index takes into account the number of recent exacerbations and these could be targeted with improvements in disease management and self-management, including smoking cessation, to improve prognosis. They noted that FEV1 status could not be modified in the same manner.

The committee agreed to not recommend the use of any of the multidimensional indices examined as they either had c-statistics that were no better than those for FEV1 alone (or indices such as GOLD 2007 based solely on FEV1), consisted of components that would not be routinely available in primary care, and/or were associated with low hazard ratios. They noted that of the indices that were suitable for primary care, none were better at classifying people reliably into high- and low-risk groups than FEV1, or they were no better at predicting outcomes than FEV1 alone. As a result, the committee agreed that there was no additional value in using a complex index instead of FEV1 for disease prognosis. The committee included BODE as an example in the do not use recommendation because it was recommended previously and they thought that non-specialists might not know what a multi-dimensional index was in this context, but were likely to recognise this index. The committee stressed that BODE was not mentioned here as an example of a particularly poor index based on classification accuracy or prognostic potential.

Several indices in particular were considered as potential replacements for BODE. DOSE was not recommended, despite a large hazard ratio (8.00) for mortality compared to the least severe category, because the c-statistics from multiple cohorts were low, ranging from 0.5 to 0.75, with a median value of 0.62. In addition, the hazard ratio data was only available from a single cohort. The committee also noted that DOSE does not include other factors with known prognostic value such as hospitalisations. HADO-AH had a c-statistic > 0.75, but the data was based on one study and the activity component (8km walk) was not considered a relevant measure for most people living with COPD.

The committee agreed to remove the recommendation to use BODE as a prognostic index as it was very similar to FEV1 alone based on c-statistics, while the hazard ratios for BODE indicated a slight increase in risk of all-cause mortality and hospitalisations, which was less than or comparable to GOLD 2007 (used in the absence of FEV1 data). In addition, they commented that BODE was not currently being used in practice in the NHS, in part because it contains an exercise component that would need to be assessed in primary care. This data is not routinely available and it would be prohibitively time consuming to collect during a general practice consultation. Based on the evidence presented and their experience in the clinic,

they agreed that BODE was not useful in practice in predicting outcomes for people with COPD.

In the absence of a suitable prognostic multidimensional test, the committee considered individual factors that are known to have prognostic value for COPD and compiled a list of these factors, including FEV1, which could prove useful in predicting outcomes for a person with COPD. Frailty was included as it is associated with worse prognosis, it can be assessed by the e-frailty index and data for this test is available in primary care records. Hospital admissions, multimorbidity, the presence of chronic hypoxia and or cor pulmonale, prescription of and need for long term oxygen therapy (LTOT) and/or domiciliary non-invasive ventilation (NIV), severity and frequency of exacerbations, and smoking status were also included as they are strongly predictive of poor outcomes in COPD. The committee recommended that these factors should be used to inform a discussion about disease prognosis or treatment decisions with the person with COPD.

The prognostic factors were added to an existing list in a recommendation from the 2010 guideline. This list also included a reference to breathlessness being measured by the MRC scale. The committee noted that the MRC scale also exists in a modified form (mMRC) that is scored from 0-4, rather than 1-5, and that healthcare professionals could under-record a person's breathlessness status if they were unaware of this.

The committee noted that there was an absence of specific evidence for prognosis at the point of diagnosis, but there was no reason to think that this situation would be different to predicting outcomes in people with an established diagnosis of COPD. As a result they agreed that the recommendations made for people already diagnosed with COPD should also be applied to people with COPD at diagnosis.

The committee decided to include a research recommendation to attempt to stimulate research to develop an effective prognostic index for COPD based on data that is useful in a primary care setting. They noted that pulse oximetry data is now routinely available and that this could be included in a new prognostic index along with additional sources of data that might be obtained for people with COPD from the assessments done as part of process for pulmonary rehabilitation.

Cost effectiveness and resource use

The committee discussed the cost effectiveness of multidimensional indices in assessing disease severity. It was determined that their generally poor prognostic ability compared with FEV1, and required time for administration means that such indices are unlikely to be cost effective.

The committee considered the cost effectiveness of thinking about individual prognostic factors when discussing an individual's treatment. It was determined that attributes such as FEV1, breathlessness, and severity and frequency of exacerbations are routinely measured when assessing patients' COPD. Therefore consideration of such factors when planning treatment is likely to be cost effective, as it carries a minimal opportunity cost, and may improve the quality of patients' care.

It was determined that the recommendations are unlikely to have a significant impact on resource use, as multidimensional indices are currently infrequently used to assess disease prognosis, whereas individual disease attributes are commonly used in informing prognosis and treatment decisions.

Other factors the committee took into account

The committee commented on the importance of discussing the problems surrounding predicting disease prognosis in people with COPD, including sharing information on the lack of accurate predictive tests. They acknowledged that this uncertainty can be frightening and challenging to cope with for people living with COPD and their families. They further noted that even where rough predictions of prognosis could be made they were derived from population level data and that this did not translate directly into risk at the individual level and that this needed to be clearly explained.

Appendices

Appendix A – Review protocols

In people with suspected COPD, what is the most accurate and appropriate test (for example imaging or biomarkers) to confirm the diagnosis?

Review protocol for confirming COPD diagnosis

Field (based on PRISMA-P)	Content
Review question	In people with suspected COPD, what is the most accurate and appropriate test (for example imaging or biomarkers) to confirm the diagnosis?
Type of review question	Diagnostic
Objective of the review	To determine the diagnostic accuracy of tests for the diagnosis of COPD in people with a positive/negative spirometry result or without spirometry results.
Eligibility criteria – population	People with suspected COPD

Eligibility criteria – assessment tools	• Imaging, including:
	  ○ Chest CT
	  ○ Chest X-ray
	  ○ 18F-FDG-PET
	  ○ Lung MRI with or without O2, 3He or 129Xe
	  ○ Pulmonary Scintigraphy (Ventilation-Perfusion (V̇/Q̇) SPECT)
	
	• Other tests:
	  ○ Full Reversibility of airways obstruction in response to bronchodilator on spirometry (adjusted for BMI)
	  ○ Sputum culture
	  ○ Serial peak flow measurements (peak expiratory flow rate (PEFR), or peak expiratory flow (PEF). Both L/min).
	  ○ Alpha-1 antitrypsin
	  ○ Transfer factor for carbon monoxide (TLCO)
	  ○ ECG
	  ○ Echocardiogram
	  ○ Pulse oximetry (peripheral oxygen saturation, SpO2)

	○ Arterial blood gas analysis ○ Sputum myeloperoxidase and serum interleukin-6 ○ Systemic inflammatory markers including eosinophil count ○ Full blood count
Eligibility criteria – reference standard	• Clinical diagnosis of COPD by any means including Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, guideline; American Thoracic Society criteria for COPD; European Respiratory Society criteria) • Post-bronchodilator spirometry in a stable patient • CT demonstration of emphysema • Histopathology grading of emphysema
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio
Eligibility criteria – study design	• Cross-sectional studies • Systematic reviews of cross-sectional studies
Other exclusion criteria	• Retrospective studies • Non-English language publications
Proposed sensitivity/sub-group analysis, or meta-regression	Subgroups: • Multimorbidities (including COPD with asthma, bronchopulmonary dysplasia, bronchiectasis, anxiety or depression) • Smoking status (smokers versus non-smokers or, data permitting, never smoked, ex-smokers and current smokers). • Spirometry status- positive, negative, unknown. Subgroup analyses will only be conducted if the majority of trials report data for the listed categories in an accessible format.
Selection process – duplicate screening/selection/Analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If

	meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer. This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See Appendix B
Information sources – databases and dates	See Appendix C Main Searches: • Cochrane Database of Systematic Reviews – CDSR (Wiley) • Cochrane Central Register of Controlled Trials – CENTRAL (Wiley)Database of Abstracts of Reviews of Effects – DARE (Wiley) • Health Technology Assessment Database – HTA (Wiley) • EMBASE (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) Citation searching will be carried out in addition on analyst/committee selected papers. The search will not be date limited as it will cover multiple review questions and the 2004 recommendations were not based on a systematic literature search. Additional terminology will be included. Economics: • NHS Economic Evaluation Database – NHS EED (Wiley) • Health Economic Evaluations Database – HEED (Wiley) • EconLit (Ovid)

	• Embase (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) The economics search will cover all questions and will be date limited from the previous search January 2009-May 2017.
Identify if an update	Update of 2004 COPD guideline question: What are the most appropriate tests in a patient with suspected COPD to confirm the diagnosis?
Author contacts	Guideline update
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix E (clinical evidence tables) or I (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix E (clinical evidence tables) or I (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B
Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B
Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B

Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Damien Longson until September 2017 and then by Andrew Molyneux in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.

Review protocol for predicting COPD severity

Field (based on PRISMA-P)	Content
Review question	In people with suspected COPD, which tests (for example imaging or biomarkers) are the most accurate to identify whether they are at risk of poor outcomes and whether they will develop mild, moderate or severe COPD?
Type of review question	Prognostic
Objective of the review	To determine the prognostic accuracy of tests to predict outcomes in people with COPD at the point of diagnosis.

Eligibility criteria – population	People with a new diagnosis of COPD (by any means including Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, guideline; American Thoracic Society criteria for COPD; European Respiratory Society criteria)

Eligibility criteria – assessment tools	• Imaging including CT, 18F-FDG-PET
	• BMI
	• Biomarkers
	• MRC dyspnoea (breathlessness) tool/Borg dyspnoea (breathlessness) score
	• Multidimensional assessment indices including:
	  ○ BODE
	  ○ CAT (self-administered COPD assessment test)
	  ○ GOLD
	  ○ DECAF (hospital based for acute exacerbations and pneumonia in COPD)
	  ○ DOSE
	  ○ COPD Diagnostic Questionnaire
	• Polycythaemia (full blood count, FBC)
	• Oxygen saturation (SaO2)
	• 6 minute walk distance (6MWD)
	• Tests for anxiety (e.g. General anxiety disorder 7, GAD7; Hospital Anxiety and Depression Scale, HADS)
	• Tests for depression (e.g. patient health questionnaire 9, PHQ9; Hospital Anxiety and Depression Scale, HADS)

Eligibility criteria – outcomes	• Mortality • Hospitalisations (no hospitalisation versus hospitalisation) • Exacerbations (exacerbations versus no exacerbations) • Severity of COPD (as defined by Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, 2017 and NICE clinical guideline 101 (2010), based on predicted airflow limitation (FEV1 %) in patients with FEV1/FVC <0.70. This will be scored as mild versus not mild (moderate and severe), moderate versus severe and severe

	versus not severe (mild and moderate) as the data permits.
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio
Eligibility criteria – study design	• Prospective cohort studies • Systematic reviews of prospective cohort studies
Other exclusion criteria	• Retrospective studies • Non-English language publications
Proposed sensitivity/sub-group analysis, or meta-regression	Subgroups: • Disease stage at diagnosis (mild, moderate or severe COPD based on predicted airflow limitation (FEV% predicted). Mild >=80%, moderate 50-79%, severe 30-49%, very severe <30%) • Exacerbations: • Frequency (no exacerbations, 1-2 exacerbations per year, and 3 or more per year) • Severity of exacerbation, stratifying by moderate versus severe exacerbations. Moderate exacerbation is defined as worsening of respiratory status that requires treatment with systemic corticosteroids and/or antibiotics; severe exacerbation is defined as rapid deterioration that requires hospitalisation. • Length of stay in hospital (stratified into short, moderate and long stay, with short 0-1 days, moderate 2-6 days and long >6) • Multimorbidities (including COPD with asthma, bronchopulmonary dysplasia, bronchiectasis, anxiety or depression) • Smoking status (smokers versus non-smokers or, data permitting, never smoked, ex-smokers and current smokers). • Age (<35, 35-65, >65 years old) • Cannabis, shisha, heroin use

	Subgroup analyses will only be conducted if the majority of trials report data for the listed categories in an accessible format.
Selection process – duplicate screening/selection/Analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer. This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See Appendix B
Information sources – databases and dates	See Appendix C Main Searches: • Cochrane Database of Systematic Reviews – CDSR (Wiley) • Cochrane Central Register of Controlled Trials – CENTRAL (Wiley) • Database of Abstracts of Reviews of Effects – DARE (Wiley) • Health Technology Assessment Database – HTA (Wiley) • EMBASE (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) Citation searching will be carried out in addition on analyst/committee selected papers. The search will not be date limited as it will cover multiple review questions and the 2004 recommendations were not based on a systematic literature search. Additional terminology will be included.

	Economics: • NHS Economic Evaluation Database – NHS EED (Wiley) • Health Economic Evaluations Database – HEED (Wiley) • EconLit (Ovid) • Embase (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) The economics search will cover all questions and will be date limited from the previous search January 2009-May 2017.
Identify if an update	New review question for the 2017 COPD guideline update.
Author contacts	Guideline update
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix E (clinical evidence tables) or I (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix E (clinical evidence tables) or I (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B
Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B

Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Damien Longson until September 2017 then Andrew Molyneux in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.

Review protocol for predicting outcomes in people with COPD using multidimensional severity assessment indices

Field (based on PRISMA-P)	Content
Review question	In people with stable COPD, does routine assessment using a multidimensional severity assessment index (such as BODE [BMI, airflow obstruction, dyspnoea/ breathlessness and exercise capacity]) better predict outcomes than forced expiratory volume in 1 second (FEV1) alone?

Type of review question	Prognostic
Objective of the review	To determine the prognostic ability of these indices in comparison to FEV1 in people with an existing diagnosis of COPD
Eligibility criteria – population	People diagnosed with COPD (by any means including Global Strategy for the Diagnosis, Management and Prevention of COPD, GOLD, guideline; American Thoracic Society criteria for COPD; European Respiratory Society criteria)

Eligibility criteria – assessment tools	• FEV1 alone
	• Multidimensional assessment indices including:
	  ○ BODE
	  ○ CAT (self-administered COPD assessment test)
	  ○ GOLD
	  ○ DECAF (hospital based for acute exacerbations)
	  ○ DOSE
	  ○ COPD Diagnostic Questionnaire

Eligibility criteria – outcomes	• Mortality • Hospitalisations • Exacerbations • Change in FEV1
Outcomes	• Sensitivity/specificity (preferred outcomes) • c-statistic, • Hazard ratios • Model fit (e.g. r-squared)
Eligibility criteria – study design	• Prospective cohort studies • Systematic reviews of prospective cohort studies
Other exclusion criteria	• Retrospective studies • Univariate analyses • Multivariate analysis if it did not adjust for age and smoking and comorbidities. • Any index, apart from FEV1, that was not multidimensional (i.e. it must include measures of different outcome combinations such as quality of life + symptoms, not just multiple dimensions of one type of outcome measure such as quality of life)

	• Non-English language publications
Proposed sensitivity/sub-group analysis, or meta-regression	Subgroups: • Setting (primary care versus specialist care assessment) • Exacerbations: • Frequency (no exacerbations, 1-2 exacerbations per year, and 3 or more per year) • Severity of exacerbation, stratifying by moderate versus severe exacerbations. Moderate exacerbation is defined as worsening of respiratory status that requires treatment with systemic corticosteroids and/or antibiotics; severe exacerbation is defined as rapid deterioration that requires hospitalisation. • Length of stay in hospital (stratified into short, moderate and long stay, with short 0-1 days, moderate 2-6 days and long >6) • Multimorbidities (including COPD with asthma, bronchopulmonary dysplasia, bronchiectasis, anxiety or depression) • Smoking status (smokers versus non-smokers or, data permitting, never smoked, ex-smokers and current smokers). Subgroup analyses will only be conducted if the majority of trials report data for the listed categories in an accessible format.
Selection process – duplicate screening/selection/Analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer.

	This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See Appendix B
Information sources – databases and dates	See Appendix C Main Searches: • Cochrane Database of Systematic Reviews – CDSR (Wiley)Cochrane Central Register of Controlled Trials – CENTRAL (Wiley) • Database of Abstracts of Reviews of Effects – DARE (Wiley) • Health Technology Assessment Database – HTA (Wiley) • EMBASE (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) Citation searching will be carried out in addition on analyst/committee selected papers. The search will not be date limited as it will cover multiple review questions and the 2004 recommendations were not based on a systematic literature search. Additional terminology will be included. Economics: • NHS Economic Evaluation Database – NHS EED (Wiley) • Health Economic Evaluations Database – HEED (Wiley) • EconLit (Ovid) • Embase (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid)

	The economics search will cover all questions and will be date limited from the previous search January 2009-May 2017.
Identify if an update	Update of 2010 COPD guideline question: Is routine assessment using multidimensional severity assessment indices (e.g. BODE) more predictive of outcomes compared with FEV1 alone?
Author contacts	Guideline update
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix E (clinical evidence tables) or I (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix E (clinical evidence tables) or I (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B
Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B
Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.

Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Damien Longson until September 2017 then Andrew Molyneux in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.

Appendix B – Methods

Priority screening

The reviews undertaken for this guideline all made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. This uses a machine learning algorithm (specifically, an SGD classifier) to take information on features (1, 2 and 3 word blocks) in the titles and abstract of papers marked as being 'includes' or 'excludes' during the title and abstract screening process, and re-orders the remaining records from most likely to least likely to be an include, based on that algorithm. This re-ordering of the remaining records occurs every time 25 additional records have been screened.

Research is currently ongoing as to what are the appropriate thresholds where reviewing of abstract can be stopped, assuming a defined threshold for the proportion of relevant papers it is acceptable to miss on primary screening. As a conservative approach until that research has been completed, the following rules were adopted during the production of this guideline:

- In every review, at least 50% of the identified abstract (or 1,000 records, if that is a greater number) were always screened.
- After this point, screening was only terminated if a pre-specified threshold was met for a number of abstracts being screened without a single new include being identified. This threshold was set according to the expected proportion of includes in the review (with reviews with a lower proportion of includes needing a higher number of papers without an identified study to justify termination), and was always a minimum of 250.

As an additional check to ensure this approach did not miss relevant studies, the included studies lists of included systematic reviews were searched to identify any papers not identified through the primary search.

Incorporating published systematic reviews

For all review questions where a literature search was undertaken looking for a particular study design, systematic reviews containing studies of that design were also included. All included studies from those systematic reviews were screened to identify any additional relevant primary studies not found as part of the initial search.

Quality assessment

Individual systematic reviews were quality assessed using the ROBIS tool, with each classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.
- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.
- Low quality – It is possible that relevant and important studies have been missed by the review.

Each individual systematic review was also classified into one of three groups for its applicability as a source of data, based on how closely the review matches the specified review protocol in the guideline. Studies were rated as follows:

- Fully applicable – The identified review fully covers the review protocol in the guideline.
- Partially applicable – The identified review fully covers a discrete subsection of the review protocol in the guideline.
- Not applicable – The identified review, despite including studies relevant to the review question, does not fully cover any discrete subsection of the review protocol in the guideline.

Using systematic reviews as a source of data

If systematic reviews were identified as being sufficiently applicable and high quality, and were identified sufficiently early in the review process, they were used as the primary source of data, rather than extracting information from primary studies. The extent to which this was done depended on the quality and applicability of the review, as defined in Table 7. When systematic reviews were used as a source of primary data, any unpublished or additional data included in the review which is not in the primary studies was also included. Data from these systematic reviews was then quality assessed and presented in GRADE/CERQual tables as described below, in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were cross-referenced to ensure none of the data had been double counted through this process.

Table 7 Criteria for using systematic reviews as a source of data

Quality	Applicability	Use of systematic review
High	Fully applicable	Data from the published systematic review were used instead of undertaking a new literature search or data analysis. Searches were only done to cover the period of time since the search date of the review.
High	Partially applicable	Data from the published systematic review were used instead of undertaking a new literature search and data analysis for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the systematic review, searches were undertaken as normal.
Moderate	Fully applicable	Details of included studies were used instead of undertaking a new literature search. Full-text papers of included studies were still retrieved for the purposes of data analysis. Searches were only done to cover the period of time since the search date of the review.
Moderate	Partially applicable	Details of included studies were used instead of undertaking a new literature search for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the systematic review, searches were undertaken as normal.

Diagnostic test accuracy evidence

In this guideline, diagnostic test accuracy (DTA) data are classified as any data in which a feature – be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features – is observed in some people who have the condition of interest at the time of the test and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly have the condition) and false positives and true negatives (in people who, according to the reference standard, do not).

The 'raw' 2x2 data can be summarised in a variety of ways. Those that were used for decision making in this guideline are as follows:

- **Positive likelihood ratios** describe how many times more likely positive features are in people with the condition compared to people without the condition. Values greater than 1 indicate that a positive result makes the condition more likely.
  - LR^+ = (TP/[TP+FN])/(FP/[FP+TN])
- **Negative likelihood ratios** describe how many times less likely negative features are in people with the condition compared to people without the condition. Values less than 1 indicate that a negative result makes the condition less likely.
  - LR^- = (FN/[TP+FN])/(TN/[FP+TN])
- **Sensitivity** is the probability that the feature will be positive in a person with the condition.
  - sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person without the condition.
  - specificity = TN/(FP+TN)

The following schema (Table 8), adapted from the suggestions of Jaeschke et al. (1994), was used to interpret the likelihood ratio findings from diagnostic test accuracy reviews.

Table 8: Interpretation of likelihood ratios

Value of likelihood ratio	Interpretation
LR ≤ 0.1	**Very large** decrease in probability of disease
0.1 < LR ≤ 0.2	**Large** decrease in probability of disease
0.2 < LR ≤ 0.5	**Moderate** decrease in probability of disease
0.5 < LR ≤ 1.0	**Slight** decrease in probability of disease
1.0 < LR < 2.0	**Slight** increase in probability of disease
2.0 ≤ LR < 5.0	**Moderate** increase in probability of disease
5.0 ≤ LR < 10.0	**Large** increase in probability of disease
LR ≥ 10.0	**Very large** increase in probability of disease

The schema above has the effect of setting a minimal important difference for positive likelihoods ratio at 2, and a corresponding minimal important difference for negative likelihood ratios at 0.5. Likelihood ratios (whether positive or negative) falling between these thresholds were judged to indicate no meaningful change in the probability of disease.

Quality assessment

Individual studies were quality assessed using the QUADAS-2 tool, which contains four domains: patient selection, index test, reference standard, and flow and timing. Each individual study was classified into one of the following three groups:

- Low risk of bias – Evidence of non-serious bias in zero or one domain.
- Moderate risk of bias – Evidence of non-serious bias in two domains only, or serious bias in one domain only.
- High risk of bias – Evidence of bias in at least three domains, or of serious bias in at least two domains.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, index features and/or reference standard in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, index feature and/or reference standard.
- Partially indirect – Important deviations from the protocol in one of the population, index feature and/or reference standard.
- Indirect – Important deviations from the protocol in at least two of the population, index feature and/or reference standard.

Methods for combining diagnostic test accuracy evidence

Meta-analysis of diagnostic test accuracy data was conducted with reference to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

Where applicable, diagnostic syntheses were stratified by:

- Presenting symptomatology (features shared by all participants in the study, but not all people who could be considered for a diagnosis in clinical practice).
- The reference standard used for true diagnosis.

Where five or more studies were available for all included strata, a bivariate model was fitted using the `mada` package in R v3.4.0, which accounts for the correlations between positive and negative likelihood ratios, and between sensitivities and specificities. Where sufficient data were not available (2-4 studies), separate independent pooling was performed for positive likelihood ratios, negative likelihood ratios, sensitivity and specificity, using Microsoft Excel. This approach is conservative as it is likely to somewhat underestimate test accuracy, due to failing to account for the correlation and trade-off between sensitivity and specificity (see Deeks 2010).

Random-effects models (der Simonian and Laird) were fitted for all syntheses, as recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Deeks et al. 2010).

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses

where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Modified GRADE for diagnostic test accuracy evidence

GRADE has not been developed for use with diagnostic studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for positive and negative likelihood ratios, as the MIDs used to assess imprecision were based on these outcomes, but results for sensitivity and specificity are also presented alongside those data.

Cross-sectional and cohort studies were initially rated as high-quality evidence if well conducted, and then downgraded according to the standard GRADE criteria (risk of bias, inconsistency, imprecision and indirectness) as detailed in Table 9 below.

Table 9: Rationale for downgrading quality of evidence for diagnostic questions

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.

GRADE criteria	Reasons for downgrading quality
Imprecision	If the 95% confidence interval for a positive likelihood ratio spanned 2, the outcome was downgraded one level, as the data were deemed to be consistent with a meaningful increase in risk and no meaningful predictive value. Similarly, negative likelihood ratios that spanned 0.5 led to downgrading for serious imprecision. Any likelihood ratios that spanned both 0.5 and 2 were downgraded twice, as suffering from very serious imprecision. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if either of the following conditions were met:

- Data showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts or protocols without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements

The evidence statements based on likelihood ratios were written with reference to the size of the likelihood ratios in the GRADE tables in appendix G, using the interpretation detailed in the methods section on diagnostic test accuracy (Table 8) for both point estimates and confidence intervals. For example, positive likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate an increase in the probability of death and negative likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate a decrease in the probability of death in people with COPD. Evidence statements were grouped according to the size of the increase or decrease.

Prognostic test accuracy evidence

In this guideline, prognostic test accuracy data are classified as any data in which a feature – be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features – is observed in some people who go on to develop the condition of interest and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly develop the condition) and false positives and true negatives (in people who, according to the reference standard, do not). This category would include studies classed as prediction models under the TRIPOD statement, provided the data were reported a 2x2 classification data.

The 'raw' 2x2 data can be summarised in a variety of ways. Those that were used for decision making in this guideline are as follows:

- **Positive likelihood ratios** describe how many times more likely positive features are in people who develop the condition compared to people who do not. Values greater than 1 indicate that a positive result makes the condition more likely.
  - LR^+ = (TP/[TP+FN])/(FP/[FP+TN])
- **Negative likelihood ratios** describe how many times less likely negative features are in people who develop the condition compared to people who do not. Values less than 1 indicate that a negative result makes the condition less likely.
  - LR^- = (FN/[TP+FN])/(TN/[FP+TN])
- **Sensitivity** is the probability that the feature will be positive in a person who goes on to develop the condition.
  - sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person who does not go on to develop the condition.
  - specificity = TN/(FP+TN)

The following schema (Table 10), adapted from the suggestions of Jaeschke et al. (1994), was used to interpret the findings from prognostic test accuracy reviews.

Table 10: Interpretation of likelihood ratios

Value of likelihood ratio	Interpretation
LR ≤ 0.1	**Very large** decrease in probability of disease
0.1 < LR ≤ 0.2	**Large** decrease in probability of disease
0.2 < LR ≤ 0.5	**Moderate** decrease in probability of disease
0.5 < LR ≤ 1.0	**Slight** decrease in probability of disease
1.0 < LR < 2.0	**Slight** increase in probability of disease
2.0 ≤ LR < 5.0	**Moderate** increase in probability of disease
5.0 ≤ LR < 10.0	**Large** increase in probability of disease
LR ≥ 10.0	**Very large** increase in probability of disease

The schema above has the effect of setting a minimal important difference for positive likelihoods ratio at 2, and a corresponding minimal important difference for negative likelihood ratios at 0.5. Likelihood ratios (whether positive or negative) falling between these thresholds were judged to indicate no meaningful change to probability of disease.

Quality assessment

Individual studies were quality assessed using the PROBAST tool[a], which contains five domains: participant selection, predictors, outcome, sample size and participant flow, analysis.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, predictive features and/or reference standard in

[a] Wolff R, Moons K, Riley R, Whiting P, Westwood M, Collins G, Reitsma J, Kleijnen J, Mallett S. PROBAST – A risk-of-bias tool for prediction-modelling studies. Abstracts of the Global Evidence Summit, Cape Town, South Africa. Cochrane Database of Systematic Reviews 2017, Issue 9 (Suppl 1). https://doi.org/10.1002/14651858.CD201702.

the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, predictive feature and/or reference standard.
- Partially indirect – Important deviations from the protocol in one of the population, predictive feature and/or reference standard.
- Indirect – Important deviations from the protocol in at least two of the population, predictive feature and/or reference standard.

Methods for combining prognostic test accuracy evidence

Where applicable, prognostic test accuracy syntheses were stratified by:

- Presenting symptomatology (features shared by all participants in the study, but not all people in the full relevant clinical population).
- The length of time between the measurement of the predictive feature and the final outcome.
- The reference standard used for categorising true positives.

Where five or more studies were available for all included strata, a bivariate model was fitted using the `mada` package in R v3.4.0, which accounts for the correlations between positive and negative likelihood ratios, and between sensitivities and specificities. Where sufficient data were not available (2-4 studies), separate independent pooling was performed for positive likelihood ratios, negative likelihood ratios, sensitivity and specificity, using Microsoft Excel. This approach is likely to somewhat underestimate test accuracy (see Deeks 2001).

Random-effects models (der Simonian and Laird) were fitted for all syntheses, due to the expected level of between study heterogeneity in prognostic reviews.

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Modified GRADE for prognostic test accuracy evidence

GRADE has not been developed for use with prognostic test accuracy studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for positive and negative likelihood ratios, as the MIDs used to assess imprecision were based on these outcomes.

Cross-sectional and cohort studies were initially rated as high-quality evidence if well conducted, and then downgraded according to the standard GRADE criteria (risk of bias, inconsistency, imprecision and indirectness) as detailed in Table 11 below.

Table 11: Rationale for downgrading quality of evidence for prognostic questions

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded.

GRADE criteria	Reasons for downgrading quality
	Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If the 95% confidence interval for a positive likelihood ratio spanned 2, the outcome was downgraded one level, as the data were deemed to be consistent with a meaningful increase in risk and no meaningful predictive value. Similarly, negative likelihood ratios that spanned 0.5 led to downgrading for serious imprecision. Any likelihood ratios that spanned both 0.5 and 2 were downgraded twice, as suffering from very serious imprecision. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if either of the following conditions were met:

- Data showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts or protocols without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements

The evidence statements based on likelihood ratios were written with reference to the size of the likelihood ratios in the GRADE tables in appendix G, using the interpretation detailed in the methods section on diagnostic test accuracy (Table 8) for both point estimates and confidence intervals. For example, positive likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate an increase in the probability of death and negative likelihood ratios, and their associated 95% confidence intervals, were used to determine which tests indicate a decrease in the probability of death in people with COPD. Evidence statements were grouped according to the size of the increase or decrease.

Other prognostic evidence

Other prognostic studies were also included if they reported outcomes of c-statistics, hazard ratios or model fit statistics. These studies were also quality assessed using the PROBAST checklist, as in the prognostic test accuracy section above.

Methods for combining other prognostic evidence

Hazard ratios

Where appropriate, hazard ratios were pooled using the inverse-variance method. Adjusted hazard ratios from multivariate models were only pooled if the same set of predictor variables were used across multiple studies and they were on the same scale. For hazard ratios, a range of 0.8, 1.25 was used to assess imprecision in the absence of a more clinically meaningful MID.

In the absence of hazard ratio data that could be meta-analysed, data was pooled to obtain single GRADE ratings per index using the following decision rules:

1. Risk of bias and indirectness were assessed as detailed in Table 11 for other prognostic evidence, but % of study population was used instead of the weight in a meta-analysis.
2. Imprecision:
    a. In cases where a single or multiple per point increase hazard ratios are presented, the level of imprecision was calculated for each study using the MID of 0.8, 1.25.If >33% of the studies by population weight have 95% CI that span one side of the MID then the index was rated as serious, if >33% have 95% CI that span both MID values then the overall index was rated as at very serious risk of imprecision.
    b. In cases where several hazard ratios are presented compared to a reference category then the most extreme category was assessed using the MID and a

single pooled estimate was determined as in 2a. If the reference categories were in opposite directions then the high reference category data was reversed (1/value) and then included in the analysis as before.

c. In cases where there is a mix of data then the imprecision was calculated for each study and then merged based on population weight as in 2a.

3. Inconsistency:
    a. For a single study this was judged to be not applicable (N/A).
    b. For multiple studies with single HRs this was judged using I^2 calculated using Review Manager v5.3 and assessed following the rules in Table 11.
    c. In cases with multiple studies each presenting several hazard ratios compared to the same reference category, the HR data for the most severe category was pooled in RevMan and inconsistency was assessed using the I^2 value following the rules in Table 11.
    d. If hazard ratio data for a single index was reported in several ways (per point increase, with reference to high and/or low categories) then inconsistency for this outcome was determined to be serious as the results were not comparable.

Assessing c-statistics

c-statistics were assessed in a similar manner to likelihood ratios using categories agreed by the committee and specified in the Table 12 below.

Table 12 Interpretation of c-statistics

Value of c-statistic	Interpretation
c-statistic <0.6	**Poor classification accuracy**
0.6 ≤ c-statistic <0.7	**Adequate classification accuracy**
0.7 ≤ c-statistic <0.8	**Good classification accuracy**
0.8 ≤ c-statistic <0.9	**Excellent classification accuracy**
0.9 ≤ c-statistic < 1.0	**Outstanding classification accuracy**

Meta-analyses could not be carried out as the data included large numbers of studies without 95% CI. In the absence of meta-analyses, the following decision rules were used to assess risk of bias, indirectness, imprecision and inconsistency for each outcome:

1. Risk of bias and indirectness were calculated as normal, but using the study weight by population, rather than weight in the meta-analysis.
2. Imprecision
    a. Single study with 95% CI: the 95% CI boundaries were examined and if they crossed 2 categories of test classification accuracy then the study was downgraded once (imprecision rated as serious); if the boundaries crossed 3 categories then the study was downgraded twice (very serious imprecision).
    b. Multiple studies with 95% CI: the individual studies were rated as in a. and then if >33.3% of the studies by population weight were rated serious then the analysis was downgraded once; if > 33.33% were rated very serious the analysis was downgraded twice.
    c. Single study or multiple studies without 95% CI: the mean sample size was calculated and if this was < 250 then the analysis was downgraded twice (very serious); if it was >250, but > 500 the analysis was downgraded once (serious); if

the mean was > 500 people/study then the analysis was not downgraded (not serious).

d. Multiple studies with and without 95% CI: the studies without 95% CI were analysed as in 2c; those with 95% CI were analysed as in 2b. The results were averaged, but the number of studies in each group were also taken into account with the result that if there were a lot more studies in one group compared to the other then that group rating would be used. In general, not serious and serious or not serious and very serious were averaged to serious; serious and very serious resulted in a very serious rating.

3. Inconsistency
    a. Single study with or without 95% CI: N/A
    b. Multiple studies with or without 95% CI: the highest and lowest point estimates were examined. If they spanned < 2 categories of c-statistic classification accuracy the analysis was rated as not serious for inconsistency; if they spanned 2 categories this was rated as serious and ≥ 3 categories was rated as very serious.

Modified GRADE for association studies

GRADE has not been developed for use with association studies; therefore a modified approach was applied using the GRADE framework. Data from cohort studies was initially rated as high quality, and data from case-control studies as low quality, with the quality of the evidence for each outcome then downgraded or not from this initial point (Table 13).

Table 13: Rationale for downgrading quality of evidence for association studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias. In addition, unadjusted odds ratio outcomes from univariate analyses were downgraded one level, in addition to any downgrading for risk of bias in individual studies. Adjusted odds ratios from multivariate analyses were not similarly downgraded, provided they included all important confounding variables in the model.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.

GRADE criteria	Reasons for downgrading quality
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity). This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	If MIDs (1 corresponding to a meaningful increase; 1 corresponding to a meaningful decrease) were defined for the outcome, the outcome was downgraded once if the 95% confidence interval for the effect size crossed 1 MID, and twice if it crosses both the upper and lower MIDs. If the line of no effect was defined as an MID for the outcome, it was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if either of the following conditions were met:

- Data showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts or protocols without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements

c-statistics

The evidence statements based on c-statistics were written with reference to the size of the c-statistics in the GRADE tables in appendix G, using the interpretation detailed in the methods section on prognostic test accuracy (Table 12). Indices were listed by median classification accuracy (for example, good median classification accuracy) and then in

alphabetical order within that level of classification accuracy, with the quality and range of classification accuracy in brackets.

Hazard ratios

Evidence statements were divided into 2 groups: indices where the HR was reported as per unit increase and indices where the HR was reported in 1 or more severity levels compared to a reference category. In each group, the indices were ordered from the largest to smallest HR with the quality in brackets. The comparison group evidence statements were ranked based on the median most extreme category hazard ratio, with data reversed where necessary so the comparisons go in the same direction.

Health economics

Literature reviews seeking to identify published cost–utility analyses of relevance to the issues under consideration were conducted for all questions. In each case, the search undertaken for the clinical review was modified, retaining population and intervention descriptors, but removing any study-design filter and adding a filter designed to identify relevant health economic analyses. In assessing studies for inclusion, population, intervention and comparator, criteria were always identical to those used in the parallel clinical search; only cost–utility analyses were included. Economic evidence profiles, including critical appraisal according to the Guidelines manual, were completed for included studies.

Economic studies identified through a systematic search of the literature are appraised using a methodology checklist designed for economic evaluations (NICE guidelines manual; 2014). This checklist is not intended to judge the quality of a study per se, but to determine whether an existing economic evaluation is useful to inform the decision-making of the committee for a specific topic within the guideline.

There are 2 parts of the appraisal process. The first step is to assess applicability (that is, the relevance of the study to the specific guideline topic and the NICE reference case); evaluations are categorised according to the criteria in Table 14.

Table 14 Applicability criteria

Level	Explanation
Directly applicable	The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness
Partially applicable	The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness
Not applicable	The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. These studies are excluded from further consideration

In the second step, only those studies deemed directly or partially applicable are further assessed for limitations (that is, methodological quality); see categorisation criteria in Table 15.

Table 15 Methodological criteria

Level	Explanation
Minor limitations	Meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness
Potentially serious limitations	Fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness
Very serious limitations	Fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration

Studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available, then other less relevant studies may not have been included. Where selective exclusions were made on this basis, this is noted in the relevant section.

Where relevant, a summary of the main findings from the systematic search, review and appraisal of economic evidence is presented in an economic evidence profile alongside the clinical evidence.

Appendix C – Literature search strategies

Main searches

Sources searched for this review question:

- Cochrane Database of Systematic Reviews – CDSR (Wiley)
- Cochrane Central Register of Controlled Trials – CENTRAL (Wiley)
- Database of Abstracts of Reviews of Effects – DARE (Wiley)
- Health Technology Assessment Database – HTA (Wiley)
- EMBASE (Ovid)
- MEDLINE (Ovid)
- MEDLINE In-Process (Ovid)

Identification of evidence

The population terms have been updated from the original guideline to include potential co-morbidities such as asthma, bronchopulmonary dysplasia and bronchiectasis. These were excluded in the original strategy.

In this update, several lines of the strategy have been focused with the use of the term 'chronic' to reduce retrieval of articles focusing on acute signs or symptoms.

Additional acronyms for COPD have been included and on recommendation from the guideline committee, terms around 'breathlessness' have been added.

Searches were re-run in February 2018 and also included searching Medline epub ahead of print.

Review question search strategy

- In people with stable COPD, does routine assessment using a multidimensional severity assessment index (such as BODE [BMI, airflow obstruction, dyspnoea/ breathlessness and exercise capacity]) better predict outcomes than forced expiratory volume in 1 second (FEV1) alone?

- In people with suspected COPD, what is the most accurate and appropriate test (for example imaging or biomarkers) to confirm the diagnosis?

- In people with suspected COPD, which tests (for example imaging or biomarkers) are the most accurate to identify whether they are at risk of poor outcomes and whether they will develop mild, moderate or severe COPD?

The MEDLINE search strategy is presented below. This was translated for use in all of the other databases.

Search strategy

Medline Strategy, searched 8th August 2017 **Database: Ovid MEDLINE(R) 1946 to July Week 4 2017** **Search Strategy:**
1 lung diseases, obstructive/
2 exp pulmonary disease, chronic obstructive/
3 (copd or coad or cobd or aecb).tw.
4 emphysema*.tw.
5 (chronic* adj4 bronch*).tw.
6 (chronic* adj3 (airflow* or airway* or bronch* or lung* or respirat* or pulmonary) adj3 obstruct*).tw.
7 (pulmonum adj4 (volumen or pneumatosis)).tw.
8 pneumonectasia.tw.
9 *Dyspnea/
10 (chronic* adj3 (breath* or respirat*) adj3 (difficult* or labor* or labour* or problem* or short*)).tw.
11 (chronic* adj3 (dyspnea* or dyspnoea* or dyspneic or breathless*)).tw.
12 or/1-11
13 *Respiratory Function Tests/
14 ((lung* or pulmonary or respirat* or ventilat*) adj2 function* adj2 test*).tw.
15 exp *Spirometry/
16 (spirometr* or bronchospirometr*).tw.
17 *Forced Expiratory Volume/
18 exp *Forced Expiratory Flow Rates/
19 (peak* adj2 flow*).tw.
20 (FEV or FEVT or FEV1 or "FEV(1)" or PEFR).tw.
21 (forced adj2 expir* adj2 (flow* or volume* or test* or index*)).tw.
22 ((lung* or timed or forced) adj2 vital capacit*).tw.
23 (maxim* adj2 (breath* or lung*) adj2 (volume* or capacit*)).tw.
24 (maxim* adj2 (expir* or inspirat* or respirat* or ventilat*)).tw.
25 or/13-24
26 *body mass index/
27 ((body mass or quetelet*) adj2 index*).tw.
28 BMI.tw.
29 or/26-28
30 *Airway Obstruction/
31 ((airflow* or airway* or respirat*) adj2 (closure or obstruct* or occlu*)).tw.
32 or/30-31
33 (DLCO or TLCO).tw.
34 transfer factor for carbon monoxide.tw.
35 or/33-34
36 *Biomarkers/
37 (biomarker* or bioindicator*).tw.
38 (biolog* adj2 indicator*).tw.
39 ((biochemical* or biolog* or clinical* or disease* or immun* or inflamm* or laborator* or serum* or surrogate* or viral) adj2 marker*).tw.
40 *Eosinophils/

Medline Strategy, searched 8th August 2017
Database: Ovid MEDLINE(R) 1946 to July Week 4 2017
Search Strategy:

41 (eosinophil* or eosinophyl*).tw.
42 (acidophil* adj2 (granulocyte* or leucocyte* or leukocyte*)).tw.
43 or/36-42
44 *Sputum/
45 sputum*.tw.
46 or/44-45
47 *Oximetry/
48 (pulse* adj2 oximetr*).tw.
49 ((oxygen* or O2) adj2 saturation).tw.
50 or/47-49
51 *Polycythemia/
52 (erythrocythemia* or polycytemia* or polycythemia* or polycythaemia* or polyerythemia* or polyglob*).tw.
53 or/51-52
54 *Interleukin-6/
55 (b cell* adj2 (differentiat* or stimulat*) adj2 factor*).tw.
56 (bsf-2 or ifn-beta 2 or il-6 or (interferon adj2 (beta-2 or beta2)) or interleukin-6 or mgi-2).tw.
57 (interleukin adj2 (b or hp1)).tw.
58 (protein* adj2 26*).tw.
59 (myeloid adj2 differentiat*).tw.
60 ((hepatocyte* or hybridoma* or plasmacytoma*) adj3 factor*).tw.
61 (a1pi or prolastin or zemaira).tw.
62 ((antipro* or antitrypsin or pi or protease or proteinase or trypsin) adj2 alpha).tw.
63 serpin a1.tw.
64 or/54-63
65 *Diagnostic Imaging/
66 (diagnos* adj2 (imag* or scan* or tomograph*)).tw.
67 exp *Tomography, X-Ray Computed/
68 ((CT or CAT) adj2 (imag* or diagnos* or scan* or detect* or exam* or tomograph*)).tw.
69 (cine-ct or tomodensitometr*).tw.
70 ((compute* or electro*) adj2 tomograph*).tw.
71 *X-Rays/
72 (x-ray* or xray* or x-radiation*).tw.
73 ((radiation* or ray*) adj2 roentgen).tw.
74 exp *Positron-Emission Tomography/
75 ((positron or PET or FDG) adj2 (imag* or scan* or tomograph*)).tw.
76 (PET adj2 FDG).tw.
77 exp *Echocardiography/ or exp *Electrocardiography/
78 ((cardi* or heart* or myocardium) adj2 scan*).tw.
79 (cardiog* or cardioecho* or (cardi* adj2 echo*) or ecg or ekg or echocardiog* or echog* or electrocardiog* or (electro adj2 cardiog*) or electromyocardiog* or polycardiog*).tw.
80 or/65-79
81 *Exercise Tolerance/
82 exp *Exercise Test/

Medline Strategy, searched 8th August 2017 Database: Ovid MEDLINE(R) 1946 to July Week 4 2017 Search Strategy:
83 (exercis* adj2 (capacit* or tolerance or test*)).tw.
84 ((fitness or step* or treadmill* or walk*) adj2 test*).tw.
85 ((six or shuttle* or "6") adj2 walk*).tw.
86 or/81-85
87 (ADO or BODE or BODEx or mBODE or CAT or CCQ or CODEX or DECAF or GOLD or SGRQ).tw.
88 (COPD adj2 assess* adj2 test*).tw.
89 "Global Initiative for Chronic Obstructive Lung Disease".tw.
90 (DOSE adj2 (index* or indice*)).tw.
91 ((COPD or St George*) adj4 questionnaire*).tw.
92 ((dyspnea* or dyspnoea* or dyspneic or breathless*) adj4 (borg or mrc or mmrc or medical research council or scale* or scor* or tool*)).tw.
93 *"Predictive Value of Tests"/
94 *"Severity of Illness Index"/
95 *"Surveys and Questionnaires"/
96 ((severity or assess* or multicomponent or multi-component or multidimensional or multi-dimensional or prognos*) adj2 (index* or indice* or survey* or tool* or questionnaire* or grad* or rate or rating or scale* or scor*)).tw.
97 (severity adj2 assess*).tw.
98 ((scor* or grad* or rate or rating) adj2 (scale* or system*)).tw.
99 or/87-98
100 25 or 29 or 32 or 35 or 43 or 46 or 50 or 53 or 64 or 80 or 86 or 99
101 12 and 100
102 sensitiv:.mp.
103 predictive value:.mp.
104 accurac:.tw.
105 or/102-104
106 prognos:.tw.
107 first episode.tw.
108 cohort.tw.
109 or/106-108
110 105 or 109
111 101 and 110
112 limit 111 to english language
113 limit 112 to (letter or historical article or comment or editorial or news or case reports)
114 112 not 113

Note: McMaster optimal diagnosis and specific prognosis filters were appended. This was adapted for Wiley database.

Study design filters and limits

The McMaster optimal diagnosis and specific prognosis filters were appended to the search strategies and are presented below. They were translated for use in the MEDLINE In-Process, Embase and Wiley databases.

Study design filters

The MEDLINE McMaster optimal diagnosis and specific prognosis filters are presented below.
McMaster optimal diagnosis sensitiv:.mp. OR predictive value:.mp. OR accurac:.tw.
McMaster specific prognosis prognos:.tw. OR first episode.tw. OR cohort.tw.

An English language limit has been applied and certain publication types (letters, historical articles, comments, editorials, news and case reports) have been excluded.

The search is not date limited as it covers multiple review questions and the 2004 recommendations were not based on a systematic literature search.

Health Economics search strategy

Economic evaluations and quality of life data

Sources searched:

- NHS Economic Evaluation Database – NHS EED (Wiley) (legacy database)
- Health Technology Assessment (HTA Database)
- EconLit (Ovid)
- Embase (Ovid)
- MEDLINE (Ovid)
- MEDLINE In-Process (Ovid)

Search filters to retrieve economic evaluations and quality of life papers were appended to population search terms in MEDLINE, MEDLINE In-Process and EMBASE to identify relevant evidence and can be seen below. Searches were carried out on 5th May 2017 with a date limit from the previous search of January 2009 – May 2017. Searches were re-run in February 2018.

An English language limit has been applied. Animal studies and certain publication types (letters, historical articles, comments, editorials, news and case reports) have been excluded.

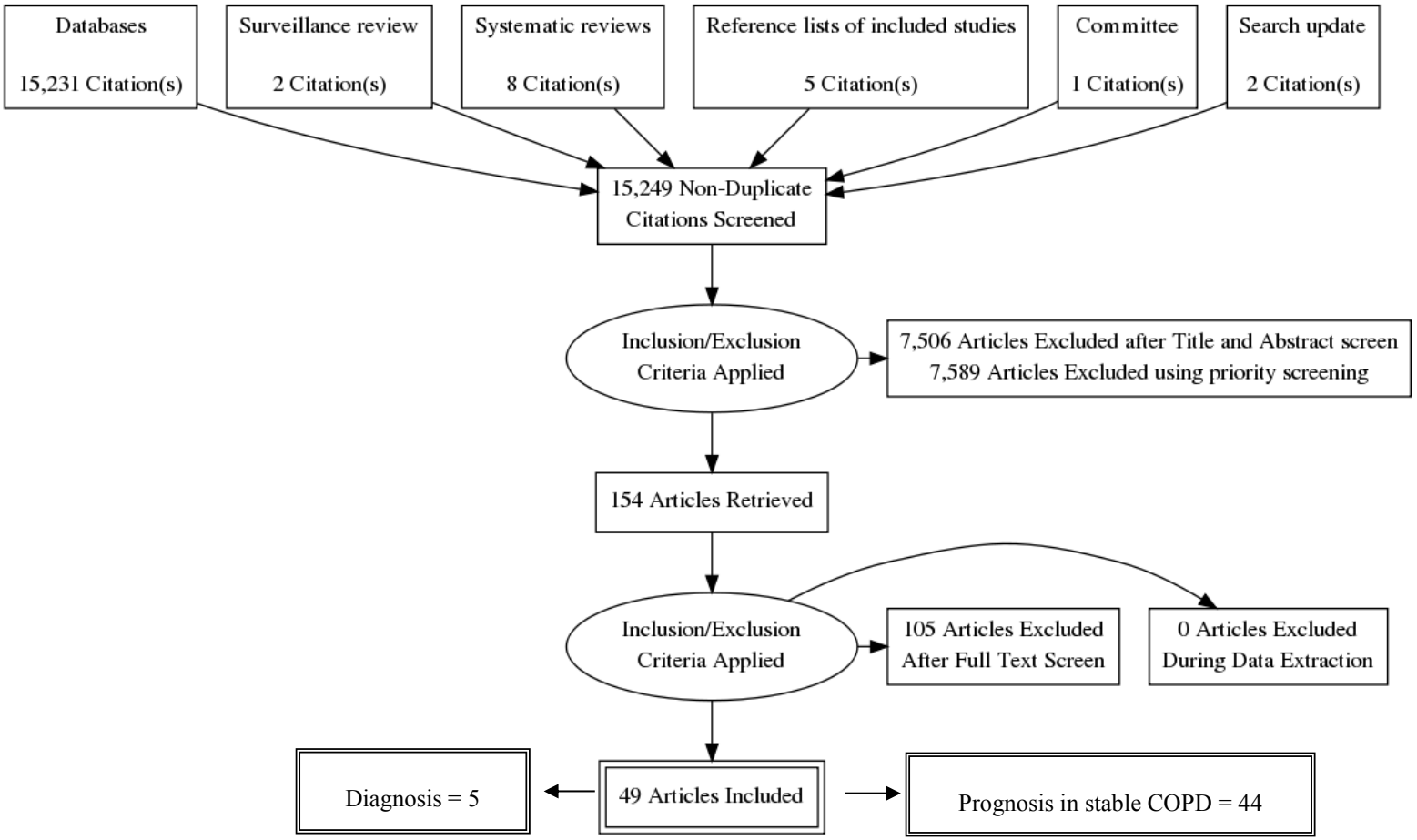
Health economics filters

The MEDLINE economic evaluations and quality of life search filters are presented below. They were translated for use in the MEDLINE In-Process and Embase databases.	
Economic evaluations	
1	Economics/
2	exp "Costs and Cost Analysis"/
3	Economics, Dental/

The MEDLINE economic evaluations and quality of life search filters are presented below. They were translated for use in the MEDLINE In-Process and Embase databases.
Economic evaluations
4 exp Economics, Hospital/
5 exp Economics, Medical/
6 Economics, Nursing/
7 Economics, Pharmaceutical/
8 Budgets/
9 exp Models, Economic/
10 Markov Chains/
11 Monte Carlo Method/
12 Decision Trees/
13 econom$.tw.
14 cba.tw.
15 cea.tw.
16 cua.tw.
17 markov$.tw.
18 (monte adj carlo).tw.
19 (decision adj3 (tree$ or analys$)).tw.
20 (cost or costs or costing$ or costly or costed).tw.
21 (price$ or pricing$).tw.
22 budget$.tw.
23 expenditure$.tw.
24 (value adj3 (money or monetary)).tw.
25 (pharmacoeconomic$ or (pharmaco adj economic$)).tw.
26 or/1-25
Quality of life
1 "Quality of Life"/
2 quality of life.tw.
3 "Value of Life"/
4 Quality-Adjusted Life Years/
5 quality adjusted life.tw.
6 (qaly$ or qald$ or qale$ or qtime$).tw.
7 disability adjusted life.tw.
8 daly$.tw.
9 Health Status Indicators/
10 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
11 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
12 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
13 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw.
14 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
15 (euroqol or euro qol or eq5d or eq 5d).tw.
16 (qol or hql or hqol or hrqol).tw.
17 (hye or hyes).tw.

The MEDLINE economic evaluations and quality of life search filters are presented below. They were translated for use in the MEDLINE In-Process and Embase databases.	
Economic evaluations	
18	health$ year$ equivalent$.tw.
19	utilit$.tw.
20	(hui or hui1 or hui2 or hui3).tw.
21	disutili$.tw.
22	rosser.tw.
23	quality of wellbeing.tw.
24	quality of well-being.tw.
25	qwb.tw.
26	willingness to pay.tw.
27	standard gamble$.tw.
28	time trade off.tw.
29	time tradeoff.tw.
30	tto.tw.
31	or/1-30

Appendix D – Clinical evidence study selection



Chronic obstructive pulmonary disease in over 16s: diagnosis and management: evidence reviews for Diagnosing COPD and predicting outcomes [December, 2018]

Appendix E – Clinical evidence tables

Confirming diagnosis of COPD

Systematic review

Author (year)	Title	Study characteristics	Quality assessment
Li (2012)	Diagnostic value of computed tomography in chronic obstructive pulmonary disease: a systematic review and meta-analysis	**Study details** • Dates searched *All of the databases were searched from their inception to October 2011.* • Databases searched *PUBMED, EMBASE, CNKI, VIP, CBM, WANFANG, The Cochrane Library.* • Sources of funding *Not stated.* **Study inclusion criteria** • Type of research was a diagnostic test that assessed the diagnostic accuracy of CT, HRCT, LDCT, or MDST for COPD • Sensitivity and specificity were reported or a 2 x 2 contingency table could be (re-) constructed • Diagnostic method for evaluation of test was CT imaging diagnosis, and reference standard was PTF • The publication was a full report	**Study eligibility criteria** • Low risk of bias **Identification and selection of studies** • Low risk of bias **Data collection and study appraisal** • High risk of bias *Study characteristics were insufficient to interpret the results* **Synthesis and findings** • Low risk of bias **Overall quality** • Moderate

Author (year)	Title	Study characteristics	Quality assessment
		Study exclusion criteria • None reported **Participant inclusion criteria** • None reported **Participant exclusion criteria** • None reported **Index test(s)** • Chest CT **Reference standard(s)** • Other *Pulmonary function tests.* **Outcomes** • Sensitivity • Specificity • Positive likelihood ratio • Negative likelihood ratio	*Study characteristics were insufficient to interpret the results* **Applicability as a source of data** • Partially applicable *Participants inclusion/exclusion criteria were not reported. It might be possible that participants did not match the target population of this review question.*

Author (year)	Title	Study characteristics	Quality assessment
		Included studies from the systematic review • Kurashima 2005 **Excluded studies from the systematic review** • Li 2008 *Chinese* • Chen 2009 *Chinese* • Long 2008 *Chinese* • Miao 2010 *Chinese* • Tsushima 2010 *Does not contain a population of people with suspected COPD* • Marsh 2007 *Does not contain a population of people with suspected COPD* • Mets 2011 *Does not contain a population of people with suspected COPD*	

Observational studies

Author (year)	Title	Study characteristics	Quality assessment

Garcia-Pachon (2004)	Can pulse oximetry select patients for screening spirometry?	**Study type** • Cross-sectional study **Study details** • Study location *Spain.* • Study setting *Outpatient Pulmonary Clinic.* • Study dates *Not stated.* • Loss to follow-up *Not applicable.* • Sources of funding *Not stated.* **Inclusion criteria** • Primary care physicians referral for evaluation of respiratory problems including sleep-disordered breathing **Exclusion criteria** • Referral because of dyspnea/ breathlessness • Patients presenting with basal dyspnea/ breathlessness score higher than 1 in the Medical Research Council Scale • Patients unable to perform spirometry	**Patient selection** • Low risk of bias **Index test** • High risk of bias *A pre-specified threshold was not used.* **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • High *A pre-specified threshold was not used* **Directness** • Directly applicable

		• Patients with haemoptysis, or with suspicion of tuberculosis **Sample characteristics** • Sample size *210* • %female *27%* • Mean age (SD) *62 years (11)* • Smoking status and history *History of smoking of more than 20 pack–years in 110 participants* • FEV1, % predicted (mean, SD) *103 participants had FEV1 value <80%* **Index test(s)** • Pulse oximetry (peripheral oxygen saturation, SpO2) *% of arterial oxygen saturation: <96 <97 <98* **Reference standard(s)** • Post-bronchodilator spirometry in a stable patient *COPD was defined as FEV1/FVC <0.70.* **Outcomes** • Sensitivity • Specificity	

		• Positive likelihood ratio • Negative likelihood ratio	
Kurashima (2005)	High resolution CT and bronchial reversibility test for diagnosing COPD	**Study type** • Cross-sectional study **Study details** • Study location *Japan* • Study setting *Respiratory clinic* • Study dates *January 2002 to June 2003* • Loss to follow-up *Not applicable* • Sources of funding *Not stated* **Inclusion criteria** • Respiratory symptoms for at least 2 months • FEV1/FVC <70%, 30 minutes after inhaled salbutamol **Exclusion criteria** • None reported	**Patient selection** • Low risk of bias **Index test** • Unclear risk of bias *Unclear whether the index test results were interpreted without knowledge of the results of the reference standard* **Reference standard** • Unclear risk of bias *Unclear whether the reference standard results were interpreted without knowledge of the results of the index test* **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Unclear whether the reference standard/index test results were*

		Sample characteristics • Sample size *516* • %female *10.5%* • Mean age (SD) *69.0 years (0.1)* • Smoking status and history *Never smoked 10.9% Ex-smoker 79.3% Current smoker 9.8%* • FEV1, % predicted (mean, SD) *58.6 (1.0)* **Index test(s)** • Chest CT *High resolution thoracic CT* **Reference standard(s)** • Clinical diagnosis of COPD *GOLD* **Outcomes** • Sensitivity • Specificity	*interpreted without knowledge of the results of the index test/reference standard* **Directness** • Directly applicable

Miniati (2011)	Computer-aided recognition of emphysema on digital chest radiography	**Study type** • Cross-sectional study **Study details** • Study location *Italy.* • Study setting *Institute of Clinical Physiology.* • Study dates *June 2007 to July 2008.* • Loss to follow-up *Not applicable.* • Sources of funding *National Research Council of Italy. Department of Medical and Surgical Critical Care, University of Florence, Italy.* **Inclusion criteria** • None reported **Exclusion criteria** • None reported **Sample characteristics** • Sample size *225*	**Patient selection** • Unclear risk of bias *Inclusion/exclusion criteria of participants were not reported* **Index test** • Low risk of bias **Reference standard** • Low risk of bias **Flow and timing** • Low risk of bias **Overall risk of bias** • Moderate *Inclusion/exclusion criteria of participants were not reported* **Directness** • Partially applicable *Pulmonary arterial hypertension was suspected in 15% of the sample before CT scan was*

		• %female *Derivation sample= 19%* *Validation sample= 44%* • Mean age (SD) *Median age (interquartile range [IQR])* *Derivation sample= 65 years (46 to 70)* *Validation sample= 66 years (57 to 73)* • Smoking status and history *Not reported* • FEV1, % predicted (mean, SD) *Not reported* **Index test(s)** • Chest X-ray *Computer-aided procedure to recognise emphysema on digital chest X-ray* **Reference standard(s)** • CT demonstration of emphysema **Outcomes** • Sensitivity • Specificity	*performed*

Tilemann (2011)	Differences in local and systemic inflammatory markers in patients with obstructive airways disease	**Study type** • Cross-sectional study **Study details** • Study location *Germany* • Study setting *General Practice and University Medical Hospital* • Study dates *Not stated* • Loss to follow-up *Not applicable* • Sources of funding *Federal Ministry of Education and Research (BMBF), Germany* **Inclusion criteria** • Respiratory symptoms for at least 2 months *Breathlessness, coughing and/or expectoration* **Exclusion criteria** • Respiratory tract infections in the 6 weeks prior to investigation *Namely, pregnancy, untreated hyperthyroidism, unstable coronary artery disease, and cardiac arrhythmia*	**Patient selection** • Low risk of bias **Index test** • High risk of bias *It is unclear whether index test results were interpreted without knowledge of the results of the reference standard. Thresholds were not pre-specified* **Reference standard** • High risk of bias *It is unclear whether reference standard results were interpreted without knowledge of the results of the index test* **Flow and timing** • Unclear risk of bias *Unclear whether all participants were included to calculate sensitivity, specificity, positive and negative predictive values because table 2 x 2 was to*

		Sample characteristics • Sample size *210* • %female *Asthma 64%* *COPD 52.8%* *Partial reversibility 46.2%* *No obstructive airways disease (OAD) 58.7%* • Mean age (SD) *Asthma 38.0 years (14.6)* *COPD 56.8 years (11.7)* *Partial reversibility 57.9 years (11.2)* *No OAD 42.3 years (14.4)* • Smoking status and history *Asthma* *Current smokers 19.8%* *Past smokers 12.8%* *Never smokers 67.4%* *COPD* *Current smokers 47.2%* *Past smokers 36.1%* *Never smokers 16.7%* *Partial reversibility* *Current smokers 61.5%* *Past smokers 23.1%* *Never smokers 15.4%* *No OAD* *Current smokers 28.0%* *Past smokers 12.0%* *Never smokers 60.0%*	*reported* **Overall risk of bias** • High *It is unclear whether index test/reference standard results were interpreted without knowledge of the results of the reference standard/index test. Thresholds for index test were not pre-specified.* *Unclear whether all participants were included to calculate sensitivity, specificity, positive and negative predictive values because table 2 x 2 was to reported* **Directness** • Directly applicable

		• FEV1, % predicted (mean, SD) *Asthma 99.7 (12.0)* *COPD 69.1 (17.1)* *Partial reversibility 67.6 (17.2)* *No OAD 106.3 (12.8)* **Index test(s)** • Systemic inflammatory markers including eosinophil count *High-sensitivity C-reactive protein concentrations (hs-CRP). The best cut-off values to discriminate between COPD and no COPD were hs-CRP concentrations of 2.39mg/L and 3.5mg/L* **Reference standard(s)** • Post-bronchodilator spirometry in a stable patient *Patients with forced expiratory volume in one second (FEV1) <80% of predicted received a bronchodilation test with an additional whole body plethysmography 20 mins after inhaling 400µg salbutamol. An OAD was diagnosed if FEV1/vital capacity (VC) was ≤0.7. The obstruction was classified as irreversible (indicating COPD) if the post-bronchodilator FEV1 was less than 12% compared with baseline and was below 200mL.* **Outcomes** • Sensitivity • Specificity	

Predicting outcomes using multidimensional severity assessment indices for people with an existing diagnosis of COPD

Author (year)	Title	Study characteristics	Quality assessment
Andrianopoulos (2015)	Prognostic value of variables derived from the six-minute walk test in patients with COPD: Results from the ECLIPSE study	• Named study cohort *ECLIPSE study* **Study type** • Prospective cohort study **Study details** • Study location *Bulgaria, Canada, Czech Republic, Denmark, Netherlands, New Zealand, Norway, Slovenia, Spain, Ukraine, United Kingdom, United States.* • Study setting *Not stated* • Study dates *December 2005 to February 2010* • Duration of follow-up *3 years* • Loss to follow-up *There was no loss at follow-up* • Sources of funding *Not stated* **Inclusion criteria** • Age *40 to 75 years* • Smoking history	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study characteristics	Quality assessment
		≥10 pack-years • Diagnosis of COPD **Exclusion criteria** • None reported **Sample characteristics** • Sample size *2010* • % female *35%* • Mean age (SD) *63.4 years (7.1)* • Smoking details *Current smokers 36%* • FEV1 %, predicted (mean (SD)) *48.5 (15.5)* **Relevant prognostic factor(s)** • BODE index (dyspnea/breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Multivariate regression model adjusted covariates** • Age • Smoking status • Gender • Body Mass Index (BMI) • FEV1 %, predicted • SGRQ (St George's Respiratory Questionnaire total score)	**Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• FEV1/FVC ratio • Emphysema **Measures** • c-statistic • Sensitivity and specificity • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality* • Hospitalisations	
Ansari (2016)	Body mass index, airflow obstruction and dyspnea and body mass index, airflow obstruction, dyspnea scores, age and pack years-predictive properties of new multidimensional prognostic indices of	**Study type** • Prospective cohort study **Study details** • Study location *UK* • Study setting *Primary care.* • Study dates *September 1999 to December 2010.* • Duration of follow-up *Median of 10 years.* • Loss to follow-up *The whole cohort was analysed.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow**

Author (year)	Title	Study characteristics	Quality assessment
	chronic obstructive pulmonary disease in primary care	• Sources of funding *Higher Education Commission, Pakistan and Sunderland Royal Hospital.* **Inclusion criteria** • Clinically stable COPD • Age *>40 years* • Diagnosis of COPD *Based on GOLD criteria* • Pulmonary function test results • COPD symptoms *Chronic cough (with or without sputum), breathlessness (with or without exertion), wheezing, and chronic airway obstruction* • COPD treatment *Stable COPD treatment* **Exclusion criteria** • Reversible airflow obstruction *>15% and >200 ml post-bronchodilator increase in FEV1* **Sample characteristics** • Sample size *458* • % female *51%* • Mean age (SD) *64.7 years (9.7)* • Smoking details	• Low risk of bias **Analysis** • High risk of bias *A pre-specified threshold was not used* **Overall risk of bias** • High *A pre-specified threshold was not used* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		87% of the cohort was either current or ex-smokers. Mean pack year history was 33.3 years (SD 18.9). Among current smokers: 53% were women and 29% were men. • FEV1 %, predicted (mean (SD)) *Survivors: 63.3 (20.3). Non-survivors: 55.8 (19.9).* **Relevant prognostic factor(s)** • BODS index • BODAS index • BOD index (dyspnea/ breathlessness, FEV1 and BMI) **Measures** • c-statistic • Sensitivity and specificity **Outcome(s)** • Mortality	
Casanova (2005)	Inspiratory-to-total lung capacity ratio predicts mortality in patients with chronic obstructive pulmonary disease. American	**Study type** • Prospective cohort study **Study details** • Study location *USA and Spain.* • Study setting *Pulmonary clinics in Boston, USA, and Tenerife and Zaragoza, Spain.* • Study dates *Participants were enrolled from December 1995 to August 2003.* • Duration of follow-up	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	journal of respiratory and critical care medicine	*Median follow-up of 34 months (range 1-62 months)* • Loss to follow-up *No loss to follow-up reported.* • Sources of funding *Not stated, but authors have no conflicts of interest.* **Inclusion criteria** • Clinically stable COPD *For at least 6 weeks prior to participation in the study and receiving optimal medical therapy.* • Smoking history *> 20 pack-years* • Pulmonary function test results *FEV1/FVC< 0.7* **Exclusion criteria** • Asthma or history of asthma *Defined as a change in FEV1 of > 200ml after bronchodilator treatment.* • Uncontrolled comorbidities *Those likely to affect mortality within 3 years such as malignancies or cardiovascular disease.* • Inability to perform the required tests **Sample characteristics** • Sample size *689* • % female *5.4*	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Mean age (SD) *Median age of survivors: 65 (range 48-79) Non-survivors: 68 (range 54-81)* • Comorbidities *Charlson index Survivors: median 4 (range 1-9) Non-survivors: median 5 (range 2-12)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Measures** • c-statistic • Sensitivity and specificity • Risk ratios **Outcome(s)** • Mortality **Additional comments** *Data for IC/TLC was not extracted as this is not a multidimensional index.*	
Casanova (2015)	Differential Effect of Modified Medical Research Council Dyspnea, COPD Assessment	• Named study cohort *CHAIN* • Multivariate regression model adjusted covariates **Study type** • Prospective cohort study	**Participant selection** • Unclear risk of bias *The confounding comorbidities used to exclude participants are not stated.*

Author (year)	Title	Study characteristics	Quality assessment
	Test, and Clinical COPD Questionnaire for Symptoms Evaluation Within the New GOLD Staging and Mortality in COPD	**Study details** • Study location *Spain* • Study setting *University hospitals in Spain* • Study dates *Participants were recruited from January 2010 to March 2012 and followed up until May 2014 for the current study.* • Duration of follow-up *Up to 5 years; time varies depending on date of recruitment (38 months on average).* • Loss to follow-up *No loss to follow-up mentioned in paper; mortality data was available for 768/768 (100%) of participants.* • Sources of funding *Astra Zeneca* **Inclusion criteria** • Clinically stable COPD *Stable for at least 8 weeks and receiving optimal medical therapy.* • Smoking history *≥ 10 pack -years* • Diagnosis of COPD *FEV1/FCV <0.7 after 400 micrograms of inhaled albuterol.* **Exclusion criteria** • Uncontrolled comorbidities	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • High risk of bias *Multivariate analysis was not adjusted for confounding variables such as smoking status and comorbidities. Data is only presented for some prognostic factors in the multivariate analysis; CAT and CCQ data is not shown.* • Low risk of bias *For c-statistic data.*

Author (year)	Title	Study characteristics	Quality assessment
		Such as malignancy at baseline or other confounding diseases that could interfere with the study. **Sample characteristics** • Sample size *768* • % female *17.5* • Mean age (SD) *68 years (9)* • Smoking details *Active smokers: 30%* • Comorbidities *Charlson Index: median 1 (range 0-5)* • FEV1 %, predicted (mean (SD)) *60 (20)* **Relevant prognostic factor(s)** • CCQ (Clinical COPD Questionnaire score) • GOLD 2011 • CAT (COPD Assessment Test) • FEV1 **Multivariate regression model adjusted covariates** • No adjustments made for covariates	**Overall risk of bias** • High *For HR due to lack of adjustment for confounding variables and selective data presentation.* • Low *For c-statistic data.* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Measures • c-statistic • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality*	
Celli (2004)	The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease	• Named study cohort *May be part of the BODE cohort, not clear from paper.* • Multivariate regression model adjusted covariates **Study type** • Prospective cohort study **Study details** • Study location *United States, Spain, and Venezuela* • Study setting *Unspecified clinics* • Study dates *Not stated, but participants were recruited between January 1997 and June 2002* • Duration of follow-up *Median follow-up of 28 months (range 4 to 68).* • Loss to follow-up *598/625 (95.7%) of participants completed the trial.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Unclear risk of bias *The model was validated in a*

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Not stated* **Inclusion criteria** • Clinically stable COPD *Also receiving appropriate therapy. If on oxygen then a stable dose for 6 months was required.* • Smoking history *> 20 pack-years* • Pulmonary function test results *FEV1/FVC <0.7 measured 20 mins after the administration of albuterol.* **Exclusion criteria** • Asthma or history of asthma *Defined as an increase in the FEV 1 of more than 15 percent above the base-line value or of 200 ml after the administration of a bronchodilator.* • Inability to perform the required tests • Illness, other than COPD, that is likely to cause death *Within 3 years* **Sample characteristics** • Sample size *Validation cohort of 625 Spain: 223 Venezuela: 54 USA: 348* • % female *Not stated* • Mean age (SD) *Spain: 66 (8) Venezuela: 64 (10) USA: 67 (9)* • Comorbidities	*separate cohort to the derivation one and bootstrapping was used during the analysis. However, the regression model was only adjusted for comorbidities, leaving other potential confounders such as age and smoking status unaccounted for.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Charlson index, mean (SD) Spain: 2.9 (1.3) Venezuela: 3.9 (1.5) USA: 5.3 (3.1) • FEV1 %, predicted (mean (SD)) *Spain: 47 (17) Venezuela: 47 (19) USA: 39 (15)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • FEV1 **Multivariate regression model adjusted covariates** • Comorbidities *Using the Charlson index* **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality	
Chan (2016)	Prognostic utility of the 2011 GOLD classification and other multidimensional tools in Asian COPD patients:	**Study type** • Prospective cohort study+ **Study details** • Study location *Singapore* • Study setting *Unspecified university hospital* • Study dates	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	a prospective cohort study	*March 2008 and March 2013* • Duration of follow-up *5 years* • Loss to follow-up *Not stated so assuming 1110/1110 (100%) of participants completed the study.* • Sources of funding *No funding was received for this study.* **Inclusion criteria** • Diagnosis of COPD *Patients with dyspnea, chronic cough, and/or sputum production with at least 10 pack-years of smoking and persistent airflow limitation as evidenced by a post-bronchodilator FEV1/forced vital capacity ratio of <0.7.* **Exclusion criteria** • None reported **Sample characteristics** • Sample size *1,110* • % female *11.2* • Mean age (SD) *71.7 years (9.3)* • FEV1 %, predicted (mean (SD)) *GOLD A: 64.3 (11.4) GOLD B: 61.9 (10.9) GOLD C: 41.4 (11.8) GOLD D: 38.2 (11.1)*	**Outcome** • Low risk of bias **Sample size and participant flow** • Unclear risk of bias *Unclear whether all participants completed the trial.* **Analysis** • Unclear risk of bias *The confounding variables that were adjusted for in the analysis were not specified.* **Overall risk of bias** • Moderate *Due to the lack of information regarding the number of participants lost to follow up and the confounding variables adjusted for in the*

Author (year)	Title	Study characteristics	Quality assessment
		Relevant prognostic factor(s) • BOD index (dyspnea/ breathlessness, FEV1 and BMI) • GOLD 2011 • GOLD 2007 **Multivariate regression model adjusted covariates** • Unspecified *Possible co-founding co-variates were adjusted for, but the study does not mention which were used.* **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality • Exacerbations *Defined as an increase in symptoms not relieved by usual reliever medications and requiring either emergency department attendance and/or admission into the hospital.*	*statistical analysis.* **Directness** • Directly applicable
Chan (2017)	Role of BMI, airflow obstruction, St George's Respiratory Questionnaire and age index in	**Study type** • Prospective cohort study **Study details** • Study location *Singapore* • Study setting *University hospital*	**Participant selection** • Unclear risk of bias *Inclusion/exclusion criteria of participants were not reported*

Author (year)	Title	Study characteristics	Quality assessment
	prognostication of Asian COPD	• Study dates *March 2008 to March 2013* • Duration of follow-up *Median of 4.07 years* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Not stated* **Inclusion criteria** • None reported **Exclusion criteria** • None reported **Sample characteristics** • Sample size *772* • % female *BOSA group 1: 7.4%* *BOSA group 2: 9.7%;* *BOSA group 3: 14.4%* *BOSA group 4: 17.9%* • Mean age (SD) *BOSA group 1: 65.6 years (7.4)* *BOSA group 2: 71.8 years (9.0)* *BOSA group 3: 73.1 years (8.5)*	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Moderate *Inclusion/exclusion criteria of participants were not reported* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		BOSA group 4: 78.3 years (7.3) • FEV1 %, predicted (mean (SD)) *BOSA group 1: 58.0% (14.9)* *BOSA group 2: 47.7% (15.3)* *BOSA group 3: 41.2% (12.0)* *BOSA group 4: 36.2% (11.4)* **Relevant prognostic factor(s)** • BOSA (BMI, FEV1 %, SGRQ, age) **Multivariate regression model adjusted covariates** • Gender • Race **Measures** • c-statistic • Sensitivity and specificity • Hazard ratios **Outcome(s)** • Mortality	
Chen (2015a)	Validation of the GOLD 2013 classification in predicting exacerbations and mortality in	**Associated studies** *Chen Chiung-Zuei, Ou Chih-Ying, Yu Chun-Hsiang, Yang Szu-Chun, Chang Han-Yu, and Hsiue Tzuen-Ren (2015) Comparison of global initiative for chronic obstructive pulmonary disease 2013 classification and body mass index, airflow obstruction, dyspnea, and exacerbations index in predicting mortality and exacerbations in elderly adults with*	**Participant selection** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	Taiwanese patients with chronic obstructive pulmonary disease	*chronic obstructive pulmonary disease. Journal of the American Geriatrics Society 63, 244-50* *Ou Chih-Ying, Chen Chiung-Zuei, Yu Chun-Hsiang, Shiu Chih-Hui, and Hsiue Tzuen-Ren (2014) Discriminative and predictive properties of multidimensional prognostic indices of chronic obstructive pulmonary disease: a validation study in Taiwanese patients. Respirology (Carlton, and Vic.) 19, 694-9* **Study type** • Prospective cohort study **Study details** • Study location *Taiwan* • Study setting *National Cheng Kung University Medical Center, Tainan.* • Study dates *Not stated, but participants were diagnosed with COPD between January 2006 and December 2012 at the hospital.* • Duration of follow-up *Chen 2015a: median follow-up of 2.9 years (range 1.4-4.1).* *Chen 2015b: median follow-up of 2.8 years (range 0.2-6.9).* • Loss to follow-up *Chen 2015a: data was analysed for 471/518 (90.0%) of participants for mortality; 338/518 (65.3%) for exacerbations.* *Chen 2015b: data was analysed for 354/429 (80.4%) for mortality; 262/429 (61.1%) for exacerbations.*	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • High risk of bias *Chen 2015a: data was only available for 65% of participants for exacerbations* *Chen 2015b: data was only available for 61% of participants for exacerbations; 80% for mortality.* • Low risk of bias *Data was available for 90% of participants for mortality.* **Analysis** • Unclear risk of bias *OR data was not adjusted for*

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Grants NSC 99-2314-B-006-040 and NSC 102- 2314-B-006-044 from the National Science Council and NCKUH 9903016 from the National Cheng Kung University Hospital.* **Inclusion criteria** • Clinically stable COPD • Age *≥ 40 years* • Pulmonary function test results *FEV1/FVC <0.7* **Exclusion criteria** • Current malignancy *Likely to result in death within 2 years.* • Illness, other than COPD, that is likely to cause death *Within 2 years (such as advanced malignancy or end-stage idiopathic pulmonary fibrosis).* **Sample characteristics** • Sample size *Chen 2015a: 518* *Chen 2015b: 429* • % female *Chen 2015a: 6.8 Chen 2015b: 7.0* • Mean age (SD) *Chen 2015a: 71.1 years (10.0) for the 471 analysed participants* *Chen 2015b: 75.7 years (5.9) for the 354 analysed participants.*	*confounding variables* • Low risk of bias *For AUC data* **Overall risk of bias** • Moderate *For exacerbations in both Chen 2015a and Chen 2015b for mortality also as there was high loss to follow-up in both studies* • Low *For mortality in Chen 2015a as most participants were included in the analysis. The measures used in the evidence review (OR not included) were also at low risk of bias.* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Smoking details *Chen 2015a: Current smoker: 63.8% History of smoking: 89.2%* *Chen 2015b: Current smoker: 70.8%* • FEV1 %, predicted (mean (SD)) *Chen 2015a: 61.6 (18.0)* **Relevant prognostic factor(s)** • BODEx index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exacerbations) • GOLD 2013 • GOLD 2007 **Measures** • c-statistic • Odds ratios **Outcome(s)** • Mortality *All-cause and respiratory mortality* • Exacerbations *Total (moderate and severe) and severe exacerbations (exacerbations requiring hospitalisation). Moderate exacerbations were defined as exacerbations requiring medical intervention with steroid or antibiotics.* **Additional comments** *To avoid double counting, data on GOLD 2013 was not extracted from Chen 2015b as it*	

Author (year)	Title	Study characteristics	Quality assessment
		was unclear whether the same participants had already been included in the data taken from Chen 2015a.	
Chen (2015b)	Comparison of global initiative for chronic obstructive pulmonary disease 2013 classification and body mass index, airflow obstruction, dyspnea, and exacerbations index in predicting mortality and exacerbations in elderly adults with chronic obstructive pulmonary disease	**Associated studies** *Chen Chiung-Zuei, Ou Chih-Ying, Hsu Chih-Hui, and Hsiue Tzuen-Ren (2015a) Validation of the GOLD 2013 classification in predicting exacerbations and mortality in Taiwanese patients with chronic obstructive pulmonary disease. Journal of the Formosan Medical Association = Taiwan yi zhi 114, 1258-66.* *Ou Chih-Ying, Chen Chiung-Zuei, Yu Chun-Hsiang, Shiu Chih-Hui, and Hsiue Tzuen-Ren (2014) Discriminative and predictive properties of multidimensional prognostic indices of chronic obstructive pulmonary disease: a validation study in Taiwanese patients. Respirology (Carlton, and Vic.) 19, 694-9* **Additional comments** *The sample characteristics of this study are recorded in the associated study record for Chen 2015a.* *To prevent double counting, data for GOLD 2013 was not extracted as it was unclear whether some of the population were also included in Chen 2015a.*	**Please refer to Chen 2015a**

Author (year)	Title	Study characteristics	Quality assessment
Cote (2008)	The modified BODE index: validation with mortality in COPD.	**Named study cohort** *May be part of the BODE cohort (not stated, but same enrolment criteria as BODE study and same countries involved).* **Associated studies** *Celli Bartolome R, Cote Claudia G, Marin Jose M, Casanova Ciro, Montes de Oca Maria, Mendez Reina A, Pinto Plata Victor, and Cabral Howard J (2004) The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. The New England journal of medicine 350, 1005-12* **Study type** • Prospective cohort study **Study details** • Study location *USA and Spain* • Study setting *Participants were recruited at the Bay Pines Veterans Administration Health Care System (Bay Pines, FL, USA) and St Elizabeth's Medical Centre (Boston, MA,UA) and Miguel Servet Hospital (Zaragoza, Spain).* • Study dates *Participants were recruited between 1996 and 2006 and followed until June 2008.* • Duration of follow-up *Not stated, but approximately 2 -12 years from study dates.* • Loss to follow-up *Not stated, but it appears all participants were included in the analysis.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Not stated.* **Inclusion criteria** • Clinically stable COPD *≥ 6 weeks with no exacerbations. Patients who were receiving inhaled oxygen had to have been taking a stable dose for at least six months before study entry.* • Smoking history *>20 pack-years* • Diagnosis of COPD • Pulmonary function test results *FEV1/FVC < 0.7* **Exclusion criteria** • Asthma or history of asthma *Defined as an increase in the FEV1 of more than 15 percent above the base-line value or of 200 ml after the administration of a bronchodilator.* • Uncontrolled comorbidities *Unstable angina* • Inability to perform the required tests • Congestive heart failure • Myocardial infarction • Illness, other than COPD, that is likely to cause death *Within 3 years* **Sample characteristics** • Sample size *444*	**Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• % female *13* • Mean age (SD) *66 years (8)* • Smoking details *Smoking history pack-yrs, mean (SD): 80 (44)* • FEV1 %, predicted (mean (SD)) *41 (15)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • mBODE% (BMI, Obstruction, Dyspnoea/ breathlessness, oxygen uptake measured at peak exercise (V'O2)) **Measures** • c-statistic • Odds ratios **Outcome(s)** • Mortality **Additional comments** *To prevent double counting, data was not analysed for the BODE index as it was unclear whether the participants had already been included in an earlier study looking at BODE (Celli 2004).*	

Author (year)	Title	Study characteristics	Quality assessment
de Torres (2008)	C-reactive protein levels and survival in patients with moderate to very severe COPD	**Named study cohort** *BODE cohort* **Study type** • Prospective cohort study **Study details** • Study location *USA and Spain.* • Study setting *Pulmonary clinics at the Hospital Universario Ntr Sra de Candelaria, Tenerife, Spain and Caritas St Elizabeth's Medical Centre, Boston, USA.* • Study dates • Duration of follow-up *24-50 months* • Loss to follow-up *Data was available for 203/218 (93.1%) participants.* • Sources of funding *Canarian Research and Health Foundation* **Inclusion criteria** • Clinically stable COPD *No exacerbations for 2 months* • Smoking history *≥ 20 pack-years* • Diagnosis of COPD *FEV1/FEVC <0.7 after 400 micrograms of inhaled albuterol*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study characteristics	Quality assessment
		Exclusion criteria • Asthma or history of asthma • Bronchiectasis • History of tuberculosis • History of malignancy • Inflammatory bowel disease • Connective tissue disorders **Sample characteristics** • Sample size *218* • % female *36.7* • Mean age (SD) *65 years (9)* • Smoking details *Pack-years: median 55 (range 40-79)* • Comorbidities *Charlson scale median 2 (range 1-4)* • FEV1 %, predicted (mean (SD)) *46 (19)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Multivariate regression model adjusted covariates** • Age • Smoking (pack years)	**Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Presence of cardiovascular risk factors or disease • Treatment with inhaled corticosteroids **Measures** • Hazard ratios **Outcome(s)** • Mortality	
de Torres (2014)	Prognostic evaluation of COPD patients: GOLD 2011 versus BODE and the COPD comorbidity index COTE.	**Named study cohort** *BODE cohort* **Study type** • Prospective cohort study **Study details** • Study location *USA and Spain* • Study setting *Pulmonary clinics* • Study dates *November 1997- March 2012* • Duration of follow-up *Not stated, but from enrolment until March 2012.* • Loss to follow-up *Not stated, data appears to be for the full 707/707 participants.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Not stated.* **Inclusion criteria** • Clinically stable COPD *Also receiving standard therapy.* • Smoking history *> 10 pack-years* • Pulmonary function test results *FEV1/FEVC <0.7 after 400 micrograms of inhaled albuterol* • Availability of data on previous exacerbations in the last year **Exclusion criteria** • A primary diagnosis other than COPD as the main respiratory disease *Asthma* • Inability to perform the required tests *Any condition that could unacceptably increase the subject's risk of performing any of the testing.* **Sample characteristics** • Sample size *707* • % female *20* • Mean age (SD) *64 years (9)* • Smoking details *Current smoking: 33%*	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Comorbidities *Charlson index, mean (SD): 1.8 (1.2)* • FEV1 %, predicted (mean (SD)) *55 (21)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • GOLD 2011 • BODE and COTE (Copd cO-morbidity TEst) combined **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality **Additional comments** To prevent double counting, data for BODE and BODE with COTE was not extracted as it was unclear whether some of the population were also included in Divo 2012 (BODE and COTE) and other BODE cohort studies.	
Divo (2012)	Comorbidities and risk of mortality in patients with chronic obstructive	**Named study cohort** *BODE cohort* **Study type** • Prospective cohort study	**Participant selection** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	pulmonary disease.	**Study details** • Study location *Spain and USA* • Study setting *Pulmonary clinics in Spain and USA.* • Study dates *November 1997-March 2009.* • Duration of follow-up *Median follow up 51 months (interquartile range 28-78 months)* • Loss to follow-up *Data was available for 1659/1664 (99.7%) of participants.* • Sources of funding *Not stated* **Inclusion criteria** • Clinically stable COPD *Also receiving appropriate therapy. If on oxygen then a stable dose for 6 months was required.* • Smoking history *> 10 pack-years* • Pulmonary function test results *FEV1/FVC <0.7 measured 20 mins after the administration of albuterol.* **Exclusion criteria** • Asthma or history of asthma *As the primary diagnosis* • Inability to perform the required tests	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Or any condition that could unacceptably increase the subject's risk of performing any of the testing. **Sample characteristics** • Sample size *1664 (data for 1659)* • % female *11* • Mean age (SD) *66 years (9)* • Comorbidities *Average (SD): 6 (3)* • FEV1 %, predicted (mean (SD)) *49 (20)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • BODE and COTE (Copd cO-morbidity TEst) combined **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality **Additional comments** *To prevent double counting, data was not analysed for the BODE index as it was unclear*	

Author (year)	Title	Study characteristics	Quality assessment
		whether the participants had already been included in an earlier study looking at BODE (Celli 2004). *COTE alone was not included in the analysis as it is not a multidimensional index.*	
Eisner (2010)	Measurement of COPD severity using a survey-based score: validation in a clinically and physiologically characterized cohort	**Named study cohort** *The Function, Living, Outcomes, and Work Study of COPD cohort* **Study type** • Prospective cohort study **Study details** • Study location *USA* • Study setting *Not stated, but the study participants were Kaiser Permanente Medical Care Programme members who lived in San Francisco.* • Study dates *Not stated* • Duration of follow-up *Not stated* • Loss to follow-up *It is unclear if there was any loss to follow-up as no information is provided.* • Sources of funding *National Heart, Lung, and Blood Institute [Grant R01 HL077618], National Institutes of Health and Flight Attendants Medical Research Institute, UCSF Bland Lane Centre of Excellence in Secondhand Smoke. Dr Eisner was also supported by the National Heart, Lung, and Blood Institute [K24 HL 097245]*	**Participant selection** • Unclear risk of bias *The study provides little information about the inclusion and exclusion criteria used to establish the cohort and in particular, does not state whether the participants all had stable COPD at baseline.* **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • High risk of bias *Due to a lack of information*

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Age *40-65 years* • Diagnosis of COPD *Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for COPD* • History of exacerbations *1 or more ambulatory visits, emergency department visits, or hospitalisations with a principal International Classification of Disease (ICD-9) diagnosis code for COPD (chronic bronchitis, emphysema, or COPD during a recent 12-month time period.* • Medication requirements *Two or more prescriptions for a COPD-related medication during a 12-month window beginning 6 months before the index utilisation date and ending 6 months after index date.* **Exclusion criteria** • Inability to perform the required tests *Such as not being able to perform spirometry due to previous tracheostomy placement.* **Sample characteristics** • Sample size *1,202* • % female *57* • Mean age (SD) *58 years (6)* • FEV1 %, predicted (mean (SD)) *62 (23)*	*about study duration, the number of hospitalisations that occurred and whether there was any loss to follow-up.* **Analysis** • Unclear risk of bias *Model was not adjusted for all the confounding variables required by our review protocol.* **Overall risk of bias** • High *Due to the lack of information provided about patient inclusion/exclusion criteria; study duration; loss to follow-up and the number of exacerbation and hospitalisation events occurring.*

Author (year)	Title	Study characteristics	Quality assessment
		Relevant prognostic factor(s) • COPD severity score **Multivariate regression model adjusted covariates** • Age • Gender • Race • Smoking history • Educational attainment **Measures** • Hazard ratios • Odds ratios **Outcome(s)** • Hospitalisations *Together with emergency room visits this was used as a proxy measure for acute exacerbations.* • Exacerbations	**Directness** • Directly applicable
Esteban (2006)	A simple score for assessing stable chronic obstructive pulmonary disease	**Study type** • Prospective cohort study **Study details** • Study location *Spain* • Study setting *Out-patient clinics affiliated with a teaching hospital in the interior district of Bizkaia*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		• Study dates *February 1998 to February 1999* • Duration of follow-up *3 years* • Loss to follow-up *26/611 (4.25%) were lost to follow-up* • Sources of funding *Fondo de Investigacion Sanitaria* **Inclusion criteria** • Clinically stable COPD *No increase in respiratory symptoms or changes in treatment for the 6 weeks prior to inclusion* • Age *Less than 80 years* • Diagnosis of COPD *For at least 6 months* • FEV1 <80% of predicted value, with FEV1/FVC <70% and negative bronchodilation test, with a change in FEV1 <200 ml and <15% of the baseline value **Exclusion criteria** • Asthma or history of asthma • History of tuberculosis *Old or ongoing concomitant pulmonary tuberculosis* • Current malignancy *Neoplastic processes* • Problems with communication	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Hearing or other problems • Psychiatric or neurological problems **Sample characteristics** • Sample size *611* • % female *2.3%* • Mean age (SD) *67.2 years (8.4)* • Smoking details *Cigarette pack-years mean (SD) 47.7 (28.7)* • Comorbidities *Hypertension 28% Diabetes 17.3% Heart problems 24.4% Back disorders 36.8% Osteoarthritis/rheumatism 43.4% Psychiatric problems 12.8%* • FEV1 %, predicted (mean (SD)) *49.7 (14.56)* **Relevant prognostic factor(s)** • HADO score (Health, Activity, Dyspnoea/ breathlessness, Obstruction Score) **Multivariate regression model adjusted covariates** • Age • Smoking status	

Author (year)	Title	Study characteristics	Quality assessment
		Measures • c-statistic • Hazard ratios **Outcome(s)** • Mortality	
Esteban (2010)	BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice?	**Associated studies** *Esteban C, Quintana JM, Aburto M, Moraza J, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. Respiratory medicine 2011; 105: 1662-70* **Study type** • Prospective cohort study **Study details** • Study location *Spain* • Study setting *Participants were recruited from an outpatient clinic affiliated with the Hospital Galdakao-Usansolo.* • Study dates *Participants were recruited between January 2003 and January 2004 and studied for 3 years.* • Duration of follow-up *3 years* • Loss to follow-up *All participants were followed-up for the 3 years.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Departamento de Sanidad del Gobierno Vasco [grant number 200111002] and by Fondo de Investigación Sanitaria [grant number PI020510].* **Inclusion criteria** • Clinically stable COPD *No changes in respiratory symptoms or treatment for at least 6 weeks.* • Diagnosis of COPD *≥ 6 months and under treatment for ≥ 6 months.* • Pulmonary function test results *FEV1 < 80% of the predicted value, with an FEV1/FVC quotient <70% and a negative bronchodilation test with FEV1 change <15% of the baseline value or <200 ml.* **Exclusion criteria** • Asthma or history of asthma • History of tuberculosis *Extensive pulmonary tuberculosis.* • Current malignancy • History of malignancy • Problems with communication *Such as hearing problems.* • Psychiatric or neurological problems *That might prevent full participation in the study.* **Sample characteristics** • Sample size *543*	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• % female *4* • Mean age (SD) *68.3 years (8.3)* • Smoking details *Smoking habit Current smokers 114 (20.9%) Ex-smokers 414 (76.2%) Never smoked 15 (2.8%) Mean pack/year (SD) 48.2 (26.5)* • Comorbidities *Charlson index: mean 2.4 (SD 1.4)* • FEV1 %, predicted (mean (SD)) *55.0 years (13.3)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • HADO score (Health, Activity, Dyspnoea/ breathlessness, Obstruction Score) **Multivariate regression model adjusted covariates** • Age • Smoking (pack years) • Comorbidities *Charlson comorbidity index* • Number of hospitalisations in the previous year	

Author (year)	Title	Study characteristics	Quality assessment
		Measures • c-statistic • Odds ratios **Outcome(s)** • Mortality *All-cause and respiratory mortality.* **Additional comments** *There was a possible overlap between study populations with Esteban 2011. As a result data was only extracted for respiratory mortality as this is not presented in the other paper.*	
Esteban (2011)	The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients	**Associated studies** *Esteban C, Quintana JM, Moraza J et al BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? BMC Medicine 2010; 8:28* **Study type** • Prospective cohort study **Study details** • Study location *Spain* • Study setting *Participants were recruited from outpatient clinics affiliated with a teaching hospital.* • Study dates	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow**

Author (year)	Title	Study characteristics	Quality assessment
		Recruited from January 2003-January 2004 and then followed for 5 years. • Duration of follow-up *5 years* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis.* • Sources of funding *Grants from the Fondo de Investigacion Sanitaria of Spain; Departamento de Sanidad del Gobierno Vasco and the Research Committee Hospital Galdakao-Usansolo.* **Inclusion criteria** • Clinically stable COPD *No changes in respiratory symptoms or treatment for the 6 weeks prior to inclusion.* • Diagnosis of COPD *≥ 6 months beforehand* • Pulmonary function test results *FEV1< 80%, FEV1/FVC <0.7 and a negative bronchodilation test with a change in FEV1 smaller than 15% and less than 200ml of the baseline value.* • Under treatment at the outpatient clinic *≥ 6 months* **Exclusion criteria** • Asthma or history of asthma • History of tuberculosis *Residual pulmonary tuberculosis* • Current malignancy	• Unclear risk of bias *Unclear whether the people included in the analysis (348 validation cohort) were all of the people included at baseline.* **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Too physically ill or mentally incapacitated to participate *Psychiatric or other problems that could prevent participation.* **Sample characteristics** • Sample size *348* • % female *4.9* • Mean age (SD) *68.0 years (8.49)* • Smoking details *Pack-years, mean (SD): 48.5 (27.3)* • FEV1 %, predicted (mean (SD)) *55.3 (13.9)* **Relevant prognostic factor(s)** • BODS index • ADO index (age, dyspnea/ breathlessness and FEV1) • HADO score (Health, Activity, Dyspnoea/ breathlessness, Obstruction Score) • HADO-AH index (HADO plus age and hospitalisation for severe COPD exacerbations) **Measures** • c-statistic **Outcome(s)** • Mortality *All-cause mortality at 3 and 5 years.*	

Author (year)	Title	Study characteristics	Quality assessment
		Additional comments *Data from the derivation cohort is not analysed here. Data for mortality was extracted for the latest time point (5 years). There was a possible overlap between study populations with Esteban 2010. As a result data was only extracted for all-cause mortality from this paper for the BODE and HADO indices.*	
Faganello (2010)	BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease.	**Study type** • Prospective cohort study **Study details** • Study location *Brazil* • Study setting *Botucata Medical School University Hospital, Sao Paolo.* • Study dates *Participants were recruited from July 2004-August 2006 and followed up for one year.* • Duration of follow-up *1 year* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Research grant from Fundacao de Amparo a Pesquisa do Estado de Sao Paolo* **Inclusion criteria** • Clinically stable COPD *≥ 6 weeks since the last exacerbation and no changes in medication.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias *Univariate analysis was used to select variables for the*

Author (year)	Title	Study characteristics	Quality assessment
		• Age *≥ 40 years* • Smoking history *≥ 10 pack-years* • Diagnosis of COPD *According to GOLD 2003 and the Brazilian Thoracic Society.* • Pulmonary function test results *FEV1/FVC <0.7* **Exclusion criteria** • Other respiratory diseases or respiratory related diseases *Interstitial fibrosis; sleep apnea/hypopnea syndrome; lung cancer.* • Asthma or history of asthma *History of asthma and/or FEV1 >12% or 200ml post-bronchodilator* • History of tuberculosis • Congestive heart failure *or unstable angina* • Myocardial infarction *Within the preceding 4 months* **Sample characteristics** • Sample size *120* • % female *29* • Mean age (SD) *65 years (9.5)* • Smoking details	*multivariable analysis, but this data was presented as OR and not used in this review as a result.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Active smoking: 29.2% Smoking history. mean (SD): 53.3 pack years (29.1) • Comorbidities *Charlson index, median (range): 3 (3-4)* • FEV1 %, predicted (mean (SD)) *60.9 (25.2)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Multivariate regression model adjusted covariates** • Age • Smoking status • Smoking (pack years) • GOLD stage • 6 MWD (6 minute walk distance) • mMRC dyspnoea/ breathlessness • SGRQ (St George's Respiratory Questionnaire total score) • SpO2 (Peripheral oxygen saturation) **Measures** • c-statistic • Sensitivity and specificity • Odds ratios **Outcome(s)** • Exacerbations *Defined as an in increase in dyspnea/ breathlessness, sputum purulence, and increased sputum volume. Classified as moderate if a visit to a physician was required (emergency*	

Author (year)	Title	Study characteristics	Quality assessment
		room or primary care) and treatment with antibiotics or systemic steroids followed. Severe exacerbations required hospital admission.	
Goossens (2014)	Does the 2013 GOLD classification improve the ability to predict lung function decline, exacerbations and mortality: a post-hoc analysis of the 4-year UPLIFT trial	**Named study cohort** *Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT) trial. Participants received 18µg of tiotropium or matching placebo once daily.* **Study type** • Prospective cohort study *Post-hoc analysis of the 4-year UPLIFT trial* **Study details** • Study location *Multinational in 37 countries* • Study setting *Not stated* • Study dates *Recruitment took place from 2003 to 2004* • Duration of follow-up *4 years* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Boehringer Ingelheim GmbH*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Age *≥40 years* • Smoking history *Currently or formerly smoking patients (≥10 pack-years)* • GOLD *Moderate to severe COPD according to the old GOLD classification system (stages 2 to 4, post-bronchodilator FEV1 of 70% or less of the predicted value)* **Exclusion criteria** • Asthma or history of asthma • Recent history of exacerbation *A COPD exacerbation or respiratory infection within 4 weeks before screening* • Lung volume reduction surgery • Supplemental oxygen for more than 12 hours per day • Coexisting illnesses that could preclude participation in the study or interfere with the study results **Sample characteristics** • Sample size *5,630* • % female *GOLD A 18.6%; GOLD B 27.3%; GOLD C; 14.2%; GOLD D 26.8%* • Mean age (SD) *GOLD A 64.9 years; GOLD B 64.6 years; GOLD C 64.6 years; GOLD D 64.5 years* • Smoking details *Current smoker GOLD A 29.2%; GOLD B 34.3%; GOLD C 24.0%; GOLD D 28.5%;* *Pack-years GOLD A 37.7%; GOLD B 40.4%; GOLD C 37.8%; GOLD D 40.7%*	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Comorbidities *Mean number of comorbidities GOLD A 3.32; GOLD B 3.71; GOLD C 3.03; GOLD D 3.71* • FEV1 %, predicted (mean (SD)) *GOLD A 60.3%; GOLD B 58.6%; GOLD C 46.5%; GOLD D 41.9%* **Relevant prognostic factor(s)** • GOLD 2011 *Stages 2 to 4* • GOLD 2013 *Stages A to D* **Measures** • c-statistic **Outcome(s)** • Mortality *All-cause mortality*	
Imfeld (2006)	The BODE index after lung volume reduction surgery correlates with survival	**Study type** • Prospective cohort study **Study details** • Study location *Switzerland* • Study setting *Pulmonary Division and Division of Thoracic Surgery, University Hospital, Zurich.* • Study dates *Participants were recruited between 1994 and 2004 to take part in a prospective study of*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		LVRS outcomes, with data collected for 40 months (median). • Duration of follow-up *Median follow-up 40 months (range 3 to 116 months)* • Loss to follow-up *Data was collected for 186/186 (100%) of the people who survived for > 3 months post-surgery.* • Sources of funding *Swiss National Foundation (grant 3200-063709.00) and the Zurich Lung League.* **Inclusion criteria** • Diagnosis of COPD • Pulmonary function test results *Severe airflow obstruction and hyperinflation (FEV1< 40% predicted, total lung capacity > 120% of predicted).* • Suitable for Lung volume reduction surgery • Severe pulmonary emphysema **Exclusion criteria** • Current malignancy *That could affect survival adversely.* • Uncontrolled comorbidities *Comorbidities likely to result in unacceptable postoperative mortality, such as symptomatic cardiovascular disease* • An extremely low functional reserve *FEV1< 20% predicted* • CT evidence of very advanced pulmonary parenchymal destruction	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Unclear risk of bias *HR data was not adjusted for confounders such as age, smoking status and comorbidities.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Sample characteristics • Sample size *186* • % female *40.9* • Mean age (SD) *63.9 years (8.2)* • FEV1 %, predicted (mean (SD)) *27.7 (7.8)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • FEV1 **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality **Additional comments** *Cohort consisted of people who had undergone lung volume reduction surgery (LVRS) at baseline.*	

Author (year)	Title	Study characteristics	Quality assessment
Johannessen (2013)	Comparison of 2011 and 2007 Global Initiative for Chronic Obstructive Lung Disease guidelines for predicting mortality and hospitalization.	**Named study cohort** *GenKOLS* **Study type** • Prospective cohort study *Nested from a case-control study* **Study details** • Study location *Norway* • Study setting *Hospital registry and general population* • Study dates *January 2003 to June 2011* • Duration of follow-up *8 years* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Not stated* **Inclusion criteria** • Age *>40 years* • Smoking history *>2.5 pack-years of smoking history*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study characteristics	Quality assessment
		• Pulmonary function test results *FEV1/FVC <0.7 and FEV1 <80%* **Exclusion criteria** • No informed consent • Severe alpha-1 antitrypsin deficiency • Previous chronic pulmonary disorder (other than COPD) • HIV • Hepatitis B or C • Dementia • Severe anaemia • Previous organ transplantation • Lung volume reduction surgery • Antibiotics for respiratory disease within 1 month of the visit • Respiratory infection within 6 weeks of the visit **Sample characteristics** • Sample size *912* • % female *GOLD 2007 classification: 2: 39%; 3: 40%; 4: 37%* *GOLD 2011 classification: A: 34%; B: 40%; C: 44% D: 40%* • Mean age (SD) *GOLD 2007 classification: 2: 64 years (10); 3: 67 years (10); 4: 65 years (10)* *GOLD 2011 classification: A: 61 years (9); B: 66 years (10); C: 63 years (10) D: 67 years (10)* • Smoking details *Current smokers %*	**Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		GOLD 2007 classification: 2: 52%; 3: 44%; 4: 36% *GOLD 2011 classification: A: 60%; B: 47%; C: 61%; D: 40%* *Pack-years mean (SD)* *GOLD 2007 classification: 2: 31 (17); 3: 33 (20); 4: 33 (18);* *GOLD 2011 classification: A: 31 (16); B: 31 (18); C: 29 (15); D: 32 (19)* • Comorbidities *Diabetes % GOLD 2007 classification: 2: 6%; 3: 5%; 4: 7%* *GOLD 2011 classification: A: 5%; B: 6%; C: 2%; D: 7%* *Heart attack/angina % GOLD 2007 classification: 2: 17%; 3: 24%; 4: 13%* *GOLD 2011 classification: A: 9%; B: 22%; C: 19%; D: 21%* *High blood pressure % GOLD 2007 classification: 2: 30%; 3: 30%; 4: 21%;* *GOLD 2011 classification: A: 29%; B: 31%; C: 25%; D: 28%* • FEV1 %, predicted (mean (SD)) *GOLD 2007 classification: 2: 64 (8); 3: 40 (6); 4: 22 (6)* *GOLD 2011 classification: A: 66 (8); B: 63 (8); C: 42 (15); D: 37 (13)* **Relevant prognostic factor(s)** • GOLD 2011 • GOLD 2007 **Multivariate regression model adjusted covariates** • Age • Smoking status • Comorbidities *Diabetes, heart attack/angina, high blood pressure.* • Gender • Body Mass Index (BMI)	

Author (year)	Title	Study characteristics	Quality assessment
		Measures • c-statistic • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality Respiratory mortality Cardiovascular mortality* • Hospitalisations *All-cause hospitalisations Respiratory hospitalisations*	
Lee (2014)	**T**he COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients.	**Study type** • Prospective cohort study **Study details** • Study location *Australia, China, Korea and Taiwan* • Study setting *Outpatient clinics across 19 hospitals.* • Study dates *Participants were recruited between August 2010 and April 2011.* • Duration of follow-up *6 months* • Loss to follow-up *495/545 (90.8%) of participants who completed the CAT questionnaire at baseline were included in the study.* • Sources of funding *GlaxoSmithKline*	**Participant selection** • Unclear risk of bias *545 participants were recruited, but only 495 were included at baseline- unclear why the remaining people were excluded.* **Predictors** • Low risk of bias **Outcome** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Age *≥ 40 years old* • Smoking history *> 10 pack- years* • Diagnosis of COPD *Diagnosed at least 6 months earlier. FEV1/FVC <0.7.* • History of exacerbations *That required additional treatment in the last 12 months.* **Exclusion criteria** • Asthma or history of asthma *Current asthma diagnosis* **Sample characteristics** • Sample size *495* • % female *12.1* • Mean age (SD) *69.4 years (8.8)* • Smoking details *Current smokers: 107/ 495 (22%)* *Smoking history (pack years): median 40 (range 10.0-196)* • Duration of COPD symptoms, months (mean (SD)) *Median 36.0 months (range 6.0-379.0)* • Comorbidities *0: 283 (57%); 1-2: 154 (31%); ≥ 3: 58 (12%)*	**Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Cardiovascular comorbidities: 189 (38%) • FEV1 %, predicted (mean (SD)) *median 47.0 (range 13.0-121.0)* **Relevant prognostic factor(s)** • GOLD (not specified) *Stages 1-4* • CAT (COPD Assessment Test) *Categories: 0-9, 10-19, 20-29, 30-40.* **Multivariate regression model adjusted covariates** • Age • Smoking status • Number of comorbidities • Country • Number of exacerbations in the previous year • Gender • Body Mass Index (BMI) • Influenza vaccination • Duration of COPD • GOLD stage **Measures** • c-statistic • Sensitivity and specificity • Hazard ratios	

Author (year)	Title	Study characteristics	Quality assessment
		Outcome(s) • Exacerbations *Defined as a worsening of symptoms of COPD for at least 2 consecutive day. Classified as mild if patients did not need to use systemic corticosteroids and/or antibiotics; moderate if treatment was required and severe if hospitalisation or a visit to the emergency room was needed. A separate exacerbation was recorded if symptoms re-occurred after >7 days of improvement.*	
Leivseth (2013)	GOLD classifications and mortality in chronic obstructive pulmonary disease: the HUNT Study, Norway	**Named study cohort** *Nord-Trøndelag Health Study (HUNT2)* **Study type** • Prospective cohort study **Study details** • Study location *Norway* • Study setting *Nord-Trøndelag County* • Study dates *August 1995 to 24 May 2012.* • Duration of follow-up *Participants were recruited from August 1995 to June 1997 and followed up to the date of death or emigration, or the end of follow-up, 24 May 2012, whichever came first. Median of 14.6 years follow-up.* • Loss to follow-up *The only reported losses to follow-up came from emigration, but the number of people involved was not stated.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Unclear risk of bias *It is unclear how many participants were lost to follow-up as the number of people lost as a result of emigration is not stated. In*

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding **Inclusion criteria** • Age *≥ 19 years old* • Diagnosis of COPD *Post-bronchodilator FEV1/ VC <0.70* • Nord-Trøndelag Health Study (HUNT2) Lung study participants *Comprised of a 5% random sample and a symptom sample from the main cohort. The symptom sample included participants reporting attacks of wheezing or breathlessness during the last 12 months, having ever had asthma and/or having ever used asthma medication, and who were not included in the random sample. Participants underwent pulmonary function tests and those with COPD were included in this study.* **Exclusion criteria** • Post-bronchodilator FEV1/FVC >0.7 **Sample characteristics** • Sample size *1540* • % female *37.9* • Mean age (SD) *63.6 years (12.5)* • Smoking details *Ever smokers: 1202/1540 (78.1%)*	*addition, people were excluded from the analysis if they had data missing.* **Analysis** • Unclear risk of bias *Model was not adjusted for all the confounding variables required by our review protocol.* **Overall risk of bias** • Moderate *Due to the lack of adjustment for all of the confounding variables required by our review protocol and the uncertainty surrounding loss to follow-up.* **Directness** • Partially directly applicable *Unclear what proportion of participants had an existing diagnosis of COPD and what*

Author (year)	Title	Study characteristics	Quality assessment
		Relevant prognostic factor(s) • GOLD (not specified) *1-4 severity grouping* • GOLD 2011 *A-D grouping* **Multivariate regression model adjusted covariates** • Age • Smoking status *Never, current, former, unknown.* • Educational attainment *<10 years, ≥ 10 years, unknown.* **Measures** • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality*	*proportion were newly diagnosed.*
Marin (2009)	Prediction of risk of COPD exacerbations by the BODE index	**Study type** • Prospective cohort study **Study details** • Study location *Spain* • Study setting *Outpatient pulmonary clinics of two tertiary teaching hospitals in Tenerife and Zaragoza.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		• Study dates *Participants were recruited between January 1997 and June 2002.* • Duration of follow-up *Up to 8 years (median 5.1 years)* • Loss to follow-up *275/275 (100%) of the participants were followed up for the duration of the study.* • Sources of funding *Not stated.* **Inclusion criteria** • Clinically stable COPD *≥ 8 weeks prior to enrolment* • Smoking history *> 20 pack-years* • Diagnosis of COPD • Pulmonary function test results *A maximal ratio of FEV1/FVC < 0.7 measured 20 min after the administration of inhaled salbutamol.* **Exclusion criteria** • Asthma or history of asthma *History of asthma and an increase in the FEV1 greater than 15% or more than 200 ml from baseline after the administration of inhaled salbutamol.* • Uncontrolled comorbidities *Likely to result in death* • Inability to perform the required tests	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Sample characteristics • Sample size *275* • % female *Not specified* • Mean age (SD) *65.1 years (8.2)* • Smoking details *Pack-years, mean (SD) Zaragoza site: 56.2 (25.4) Tenerife: 48.5 (21.8)* • Comorbidities *CHARLSON index, mean (SD) Zaragoza site: 2.7 (1.4) Tenerife site: 2.5 (1.3)* • FEV1 %, predicted (mean (SD)) *Zaragoza site: 49.6 (17.6) Tenerife site: (48.5 (19.2)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Measures** • c-statistic • Sensitivity and specificity • Odds ratios **Outcome(s)** • Exacerbations *Exacerbations were defined as events characterized by a sustained worsening of baseline respiratory symptoms that lasted for at least 3 days and that required treatment intervention with antibiotics, and/or systemic corticosteroids. The number of COPD*	

Author (year)	Title	Study characteristics	Quality assessment
		exacerbations leading to primary care visits, emergency room visits and hospitalisations were recorded. **Additional comments** *Data could not be extracted for a 2x2 table based on the sensitivity and specificity of BODE in predicting exacerbations that needed primary care, emergency room visits or hospitalisations due to the lack of information regarding the numbers of people who had each type of exacerbation. The FEV1 c-statistic for predicting exacerbations was not provided and so could not be used in our analyses.*	
Marin (2011)	Prognostic assessment in COPD: health related quality of life and the BODE index	**Named study cohort** *BODE cohort* **Study type** • Prospective cohort study **Study details** • Study location *Spain, USA, Venezuela.* • Study setting *Unspecified clinics in the host countries.* • Study dates *Participants were recruited between January 1997 and September 2006* • Duration of follow-up *Until August 2007; mean 53 months (SD 28)* • Loss to follow-up *1398/1398 (100%) of people were included in the analysis.*	**Participant selection** • Low risk of bias **Predictors** • Unclear risk of bias *The prognostic tests may have not have been assessed in the same way for all participants because the study was carried out at sites across 3 countries.*

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *Not stated* **Inclusion criteria** • Clinically stable COPD *All patients were in clinically stable condition and receiving appropriate therapy. Patients who were receiving inhaled oxygen had to have been taking a stable dose for at least six months before study entry.* • Smoking history *20 pack-years* • Pulmonary function test results *FEV1/FVC <0.7 measured 20 minutes after the administration of albuterol.* **Exclusion criteria** • Asthma or history of asthma *Defined as an increase in the FEV 1 of more than 15 percent above the base-line value or of 200 ml after the administration of a bronchodilator.* • Uncontrolled comorbidities *Illness other than COPD that was likely to result in death within three years; unstable angina.* • Inability to perform the required tests • Congestive heart failure • Myocardial infarction *Within the last 4 months* **Sample characteristics** • Sample size *1398*	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• % female *15* • Mean age (SD) *66 years (9)* • Comorbidities *Charlson (points): 4.16 (2.4)* • FEV1 %, predicted (mean (SD)) *Post-bronchodilator 46 (18)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Measures** • c-statistic • Model fit (e.g. r-squared) *Pearson's correlation coefficients* **Outcome(s)** • Mortality	
Marin (2013)	Multicomponent indices to predict survival in COPD: The COCOMICS study	**Named study cohort** *COCOMICS* **Study type** • Prospective cohort study	**Participant selection** • Unclear risk of bias *Participant inclusion and exclusion criteria varied across the cohorts.*

Author (year)	Title	Study characteristics	Quality assessment
		Study details • Study location *Spain* • Study setting *Data from cohorts recruited in Galdakao, Pamplona, Requena, Seville, Tenerife, Terrassa and Zaragoza from out-patient clinics and interventional clinical trials of people with COPD.* • Study dates *Individual studies ran from between 1997 for the earliest up to 2010 for the latest.* • Duration of follow-up *Galdako: 7 years; Pamplona: 5 years; Requena I: 7 years; Requena II: 6 years; Seville: 12 years; Tenerife: 14 years; Zaragoza I: unclear 1998-?; Zaragoza II: 11 years* • Loss to follow-up *Unclear as data was obtained for multiple cohorts and loss to follow-up was not detailed.* • Sources of funding *The COCOMICS initiative received a group coordination grant of the Spanish Society of Pneumology and Thoracic Surgery BECA SEPAR 2012 coded with number 057|2012. Pablo Martı´nez-Camblor was supported by the research Grant MTM2011-23204 of the Spanish Ministerio de Ciencia e Innovacion.* **Inclusion criteria** • Clinically stable COPD *No exacerbation for at least 4 weeks before enrolment, apart from the Terrassa cohorts which recruited people with exacerbations.* • Age *< 80 years for the Galdakao cohort; no age limits for the other cohorts.* • Smoking history *Pamplona, Tenerife and Requena: ≥20 pack-years*	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Unclear risk of bias *It is unclear how many participants were lost to follow-up across the cohorts.* **Analysis** • High risk of bias *Data was only presented for selected indices with no explanation.* **Overall risk of bias** • Moderate *Due to selective reporting of test data and the lack of information about loss to*

Author (year)	Title	Study characteristics	Quality assessment
		• Diagnosis of COPD *Seville cohort: according to the in the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria.* • History of exacerbations *Recent exacerbations required for inclusion in the Terressa cohorts.* • Pulmonary function test results *Pamplona: post-bronchodilator FEV 1 /FVC of < 0.70 after the administration of 400 mg of inhaled albuterol; Requena: FEV1/FVC <0.7 with a change in FEV1 of less than 200 ml and 12% in the bronchodilator test; Tenerife: post-bronchodilator FEV1/FVC ratio < 0.7* • No previous anti-pneumococcal vaccination *Seville cohort* **Exclusion criteria** • Other respiratory diseases or respiratory related diseases *Requena cohort: cystic fibrosis, upper airways obstruction, or bronchiolitis related to systemic pathology* • Asthma or history of asthma *Pamplona, Tenerife and Requena cohorts* • Bronchiectasis *Pamplona and Requena cohorts* • History of tuberculosis *Pamplona cohort* • Uncontrolled comorbidities *Tenerife cohort: comorbidities likely to affect mortality within 3 years* • Inability to perform the required tests *Tenerife cohort* • Other potentially confounding diseases *Pamplona cohort*	*follow-up.* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• A change in FEV1 of more than 200 ml after bronchodilator treatment *Tenerife cohort* • Pregnant or immunosuppressed, known neoplasia, renal insufficiency in dialysis, HIV infection, hypo gammaglobulinemia or anatomical or functional asplenia *Seville cohort* **Sample characteristics** • Sample size *3,633 for all 10 cohorts. Galdekao: 543; Palmplona: 190; Requena I: 174; Requena II: 186; Seville: 595; Tenerife: 275; Zaragoza I:137; Zaragoza II: 1150 (Terassa excluded)* • % female *Galdakao: 4 Pamplona: 16; Requena I: 1; Requena II: 1; Seville: 5; Tenerife: 21; Zaragoza I: 1; Zaragoza II: 7* • Mean age (SD) *Galdakao: 68.3 years (8.3); Pamplona: 65.2 (8.4); Requena I: 72.1 years (8.9); Requena II: 70.9 (8.0); Seville: 65.8 years (9.6); Tenerife: 62.9 years (9.9); Zaragoza I: 65.8 years (7.6); Zaragoza II: 63.4 years (9.4)* • Smoking details *Pack-years of smoking, mean (SD) Galdakao: 48 (26); Pamplona: 53 (27); Requena I: 61 (33); Requena II: 63 (36); Seville: 50 (22); Tenerife: 65 (27); Zaragoza I: 57 (25); Zaragoza II: 52 (25)* *Smoking status, mean (%)* *Former: Galdakao 414 (76); Pamplona 119 (63); Requena I 131 (76); Requena II 151 (81); Seville 453 (76); Tenerife 140 (57); Zaragoza I 99 (73); Zaragoza II 740 (66).* *Current: 114 (21); 71 (37); 39 (22); 32 (17); 143 (24); 103 (43); 384 (34).* *Never: 15 (3); 0 (0); 3 (2); 3 (2); 0 (0); 0 (0); 0 (0); 0 (0).* • Comorbidities *Charlson index, mean (SD)*	

Author (year)	Title	Study characteristics	Quality assessment
		Galdakao: 1.4 (1.4); Pamplona: 1.6 (1.3); Requena I: 1.1 (1.6); Requena II: 0.8 (0.9); Seville: 1.1 (1.8); Tenerife: 2.0 (1.7); Zaragoza I: 1.2 (1.3); Zaragoza II: 2.5 (1.1) • FEV1 %, predicted (mean (SD)) *Galdakao: 55 (13); Pamplona: 69 (19); Requena I: 48 (17); Requena II: 46 (17); Seville: 43 (13); Tenerife: 56 (21); Zaragoza I: 45 (14); Zaragoza II: 48 (18)* **Relevant prognostic factor(s)** • BODE index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • e-BODE (BODE plus exacerbations) • BODEx index (dyspnea/ breathlessness (mMRC), BMI, FEV1 and exacerbations) • DOSE index (Dyspnoea/ breathlessness, Obstruction, Smoking, Exacerbation Index) • ADO index (age, dyspnea/ breathlessness and FEV1) • HADO score (Health, Activity, Dyspnoea/ breathlessness, Obstruction Score) • SAFE index (quality of life by SGRQ, FEV1 and 6MWD) • TARDIS (age, BMI, dyspnea/ breathlessness, airflow obstruction, hospitalisations and influenza vaccination) • COPD Prognostic Index (CPI) (quality of life standardised by the CRQ or SGRQ, FEV1, age, sex, BMI, exacerbation history and cardiovascular disease history) • COPDSS-COPD severity score (respiratory symptoms, systemic corticosteroid use, other COPD medication use, previous hospitalisation or intubation for respiratory disease and home oxygen use) • FEV1 **Measures** • c-statistic	

Author (year)	Title	Study characteristics	Quality assessment
		Outcome(s) • Mortality **Additional comments** *Data for the Terrassa cohorts was excluded as these people did not have stable COPD at baseline. The Zaragoza cohorts were part of a study of sleep apnoea and had suspected sleep disorders at baseline.*	
Mattila (2015)	Association between all-cause and cause-specific mortality and the GOLD stages 1-4: A 30-year follow-up among Finnish adults	**Study type** • Prospective cohort study **Study details** • Study location *Finland* • Study setting *National representative sample of adult Finns* • Study dates *Sample was taken between 1978 and 1980* • Duration of follow-up *3 years* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *HUCS/Hyvinkaa Hospital, the Foundation of the Finnish Anti-Tuberculosis Association, and the Tuberculosis Foundation of Tampere.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Age *30 years and older* **Exclusion criteria** • Asthma or history of asthma **Sample characteristics** • Sample size *6636* • % female *54.0%* • Mean age (SD) *Not reported* • Smoking details *Non-smoker 55.4% Former smoker 21.0% Current smoker (1 to 9 cigarettes/day) 14.7% Current smoker (≥20 cigarettes/day) 8.9%* • Comorbidities *Diabetes 4.1%* **Relevant prognostic factor(s)** • GOLD (not specified) *Article has a reference but the reference does not have a date of publication.* **Multivariate regression model adjusted covariates** • Age	**Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Partially directly applicable *Participants without COPD were also included.*

Author (year)	Title	Study characteristics	Quality assessment
		• Smoking status • Gender **Measures** • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality Respiratory mortality*	
Moberg (2014)	Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD participating in pulmonary rehabilitation	**Study type** • Prospective cohort study **Study details** • Study location *Denmark* • Study setting *Pulmonary rehabilitation programme* • Study dates *March 2002 to March 2011* • Duration of follow-up *Mean follow-up was 66 months (range 11 to 118 months)* • Loss to follow-up *18 (2.6%) participants with missing values* • Sources of funding *TrygFonden*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Clinically stable COPD • Pulmonary function test results *FEV1/FVC <0.70 and FEV1 <80%* • Motivation • No previous participation in pulmonary rehabilitation **Exclusion criteria** • Significant cognitive problems • Significant musculo-skeletal problems • Significant cardiac problems **Sample characteristics** • Sample size *674* • % female *64.2%* • Mean age (SD) *69.2 years (9.2)* • Smoking details *Pack years 40.4 (range 0 to 150) Current smoking 26.1%* • FEV1 %, predicted (mean (SD)) *36.7% (13.3)* **Relevant prognostic factor(s)** • i-BODE (BODE plus incremental shuttle walking test [ISWT])	**Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Multivariate regression model adjusted covariates • Age • Smoking status • Smoking (pack years) • Gender • Oxygen saturation at rest • Desaturation >4% during shuttle walking test (SWT) • Maintenance prednisolone • LTOT **Measures** • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality* • Hospitalisations *All-cause hospitalisations* *Hospitalisations due to exacerbation in COPD*	
Motegi (2013)	A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations.	**Study type** • Prospective cohort study **Study details** • Study location *Japan* • Study setting *Outpatient Respiratory Care Clinic*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	International journal of chronic obstructive pulmonary disease	• Study dates *Enrolment was from April 2007 to October 2007. Follow-up was from November 2007 to October 2009.* • Duration of follow-up *2 years* • Loss to follow-up *23/206 (11.2%) were lost to follow-up* • Sources of funding *Not stated* **Inclusion criteria** • Age *≥40 years* • Smoking history *Current or former smoking history of at least 20 pack-years* • Diagnosis of COPD *By a chest physician* **Exclusion criteria** • Asthma or history of asthma • Bronchiectasis • History of tuberculosis *Active tuberculosis or any history of pulmonary fibrosis* • Current malignancy • Dementia • Bulluos lung disease • Withdrawal of consent	**Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Sample characteristics • Sample size *183* • % female *7.1%* • Mean age (SD) *71.4 years (8.7)* • Smoking details *Current smoking 4.4% Smoking pack-years 75.6 (45.3)* • Comorbidities *Charlson index mean (SD) 2.6 (1.0)* • FEV1 %, predicted (mean (SD)) *55.7 (20.7)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • DOSE index (Dyspnoea/ breathlessness, Obstruction, Smoking, Exacerbation Index) • ADO index (age, dyspnoea/ breathlessness and FEV1) • GOLD 2006 **Measures** • c-statistic • Odds ratios	

Author (year)	Title	Study characteristics	Quality assessment
		Outcome(s) • Exacerbations *Occurrence of exacerbation during the 1-year follow-up*	
Moy (2014)	An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease.	**Study type** • Prospective cohort study **Study details** • Study location *USA* • Study setting *Not stated* • Study dates *January 2009 to November 2011* • Duration of follow-up *Mean follow-up was 15 months* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Department of Veteran's Affairs, Veterans Health Administration, Rehabilitation Research and Development Services; Center for Integration of Medicine and Innovative Technology, Boston; and VA Rehabilitation Research and Development Service Merit Review.* **Inclusion criteria** • Clinically stable COPD • Age	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Over 40 years • Diagnosis of COPD *Define as a smoking history of at least 10 pack-years and either a FEV1/FVC ratio of <0.70 or evidence of emphysema on chest computed tomography* • At least 4 weeks had elapsed since previous acute exacerbations **Exclusion criteria** • Asthma or history of asthma **Sample characteristics** • Sample size *167* • % female *1.2%* • Mean age (SD) *71 years (8)* • Smoking details *Pack years: mean 68 (SD 37) Current cigarette smoker: 22%* • Comorbidities *Coronary artery disease: 38% Congestive heart failure: 14% Diabetes mellitus: 28%* • FEV1 %, predicted (mean (SD)) *54% (20)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD))	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Measures • c-statistic **Outcome(s)** • Hospitalisations *COPD related-hospitalisations* • Exacerbations *Acute exacerbations*	
Neo (2017)	Prediction of Poor Short-Term Prognosis and Unmet Needs in Advanced Chronic Obstructive Pulmonary Disease: Use of the Two-Minute Walking Distance Extracted from a Six-Minute Walk Test	**Study type** • Prospective cohort study **Study details** • Study location *Singapore* • Study setting *Respiratory outpatient clinics* • Study dates *The study commenced in early 2013* • Duration of follow-up *18 months* • Loss to follow-up *All subjects were followed-up* • Sources of funding *NHG-KTPH Small Innovative Grant (SIG) and the Lien Center of Palliative Care Extramural Research Awards.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Age *≥21 years old* • Diagnosis of COPD • Pulmonary function test results *FEV1/FVC <0.7* • GOLD *Stage 3 and 4 COPD* **Exclusion criteria** • Other respiratory diseases or respiratory related diseases *Active pulmonary tuberculosis, pulmonary fibrosis, pneumothorax, or lung cancer* • Asthma or history of asthma • Too physically ill or mentally incapacitated to participate • Active microbial infections • Hospitalisation for acute COPD exacerbations within the recent 2 weeks **Sample characteristics** • Sample size *124* • % female *11.5%* • Mean age (SD) *71.7 years (7.6)* • FEV1 %, predicted (mean (SD)) *35.9 (9.8)*	**Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Relevant prognostic factor(s) • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • ADO index (age, dyspnoea/ breathlessness and FEV1) *Updated-ADO index* **Measures** • c-statistic **Outcome(s)** • Mortality	
Omachi (2008)	The COPD severity score: a dynamic prediction tool for health-care utilization	**Study type** • Prospective cohort study **Study details** • Study location *USA* • Study setting *Population-based sample of US adults with COPD across 48 states in USA and in geographic COPD hotspots. Contacted by phone.* • Study dates *Not stated, but interviews at baseline took place in 2001.* • Duration of follow-up *1 year* • Loss to follow-up *Data was available for 65% of the derivation cohort of 267 people (interviewed in 2002) for the 2004 interview (end point). This corresponds to 173/204 (85%) of the people who were re-interviewed in 2003.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias *15% off the population was lost to follow-up, but data*

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *National Institutes of Health grant R01 HL067438 from the National Heart, Lung, and Blood Institute. Dr. Omachi was supported by National Heart, Lung, and Blood Institute, grant number T32 HL007185. Dr. Eisner was supported by R01HL077618 National Heart, Lung, and Blood Institute, National Institutes of Health, with co-funding by the Social Security Administration.* **Inclusion criteria** • Age *55-77 years* • Diagnosis of COPD *Subjects were classified as having COPD for the current analysis if they reported, in response to a specific query during the structured telephone interview, that a physician had diagnosed them with chronic bronchitis, emphysema, or specifically with COPD. People with a concomitant diagnosis of asthma were included.* **Exclusion criteria** • None reported **Sample characteristics** • Sample size *267 in the internal validation cohort* • % female *63* • Mean age (SD) *65.2 years (6.1)* • Smoking details *Cigarette Smoking, n (%) Never smoked 48 (18) Current smoker 82 (31) Former smoker*	*was calculated for the missing participants.* **Analysis** • Unclear risk of bias *15% off the population was lost to follow-up, but data was calculated for the missing participants. However, the study does not state how many events occurred in the 1 year follow-up period. Model was not adjusted for all the confounding variables required by our review protocol.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		136 (51) • Comorbidities *Comorbidities, n (%) Coronary Artery Disease 57 (21) Congestive Heart Failure 43 (16) Diabetes 61 (23) Sleep Apnea 23 (9)* **Relevant prognostic factor(s)** • COPD severity score **Multivariate regression model adjusted covariates** • Age • Smoking history • Comorbidities • Race • Educational attainment **Measures** • c-statistic **Outcome(s)** • Hospitalisations **Additional comments** *383 people were included in the initial cohort in 2001 and used to help derive the COPD severity score. Of these people, 267 were re-interviewed in 2002 and used as a validation cohort.*	

Author (year)	Title	Study characteristics	Quality assessment
Ou (2014)	Discriminative and predictive properties of multidimensional prognostic indices of chronic obstructive pulmonary disease: a validation study in Taiwanese patients.	**Associated studies** *Chen Chiung-Zuei, Ou Chih-Ying, Hsu Chih-Hui, and Hsiue Tzuen-Ren (2015) Validation of the GOLD 2013 classification in predicting exacerbations and mortality in Taiwanese patients with chronic obstructive pulmonary disease. Journal of the Formosan Medical Association = Taiwan yi zhi 114, 1258-66* *Chen Chiung-Zuei, Ou Chih-Ying, Yu Chun-Hsiang, Yang Szu-Chun, Chang Han-Yu, and Hsiue Tzuen-Ren (2015) Comparison of global initiative for chronic obstructive pulmonary disease 2013 classification and body mass index, airflow obstruction, dyspnea, and exacerbations index in predicting mortality and exacerbations in elderly adults with chronic obstructive pulmonary disease. Journal of the American Geriatrics Society 63, 244-50* **Study type** • Prospective cohort study **Study details** • Study location *Taiwan* • Study setting *Outpatient clinic at National Chen Kung University Hospital.* • Study dates *Participants were recruited between May 2006 and December 2011.* • Duration of follow-up *Not specified, but at least one year.* • Loss to follow-up *594/621 (95.6%) of participants were followed up for at least one year.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low

Author (year)	Title	Study characteristics	Quality assessment
		• Sources of funding *National Science Council of Taiwan.* **Sample characteristics** • Sample size *594* • % female *6.6* • Mean age (SD) *71.7 years (10.2)* • Smoking details *Current smokers: 20.4%* • FEV1 %, predicted (mean (SD)) *Survivors: 63.7 (22.1) Non-survivors: 54.5 (21.7)* **Relevant prognostic factor(s)** • BODEx index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exacerbations) • ADO index (age, dyspnoea/ breathlessness and FEV1) • COPD Prognostic Index (CPI) (quality of life standardised by the CRQ or SGRQ, FEV1, age, sex, BMI, exacerbation history and cardiovascular disease history) **Measures** • c-statistic **Outcome(s)** • Mortality	**Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Additional comments *To prevent double counting, data was not extracted for BODEx as it was unclear whether some of the same participants had been included in the studies by Chen 2015a and Chen 2015b.*	
Pedone (2014)	BODE index or geriatric multidimensional assessment for the prediction of very-long-term mortality in elderly patients with chronic obstructive pulmonary disease?	**Named study cohort** *SARA (Salute Respiratoria nell'Anziano – Respiratory Health in the Elderly) study* **Study type** • Prospective cohort study **Study details** • Study location *Italy* • Study setting *Pulmonary or geriatric outpatient clinics* • Study dates *Recruitment took place from January 1996 to July 1999* • Duration of follow-up *15 years* • Loss to follow-up *Information on vital status as of December 2010 was obtained for 468/563 (82%) participants* • Sources of funding *Not stated*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Diagnosis of COPD *FEV1/FVC <0.7* **Exclusion criteria** • Asthma or history of asthma *FEV1 ≥80% of predicted and a history of wheezing in the last year, or a FEV1 <80% of predicted and with FEV1 increase ≥12% after inhalation of fenoterol* **Sample characteristics** • Sample size *468* • % female *23.3%* • Mean age (SD) *73.3 years (5.51)* • Comorbidities *Ischaemic heart disease 12.7% Heart failure 6.9% Stroke 3.7%* • FEV1 %, predicted (mean (SD)) *69 (24.7)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) **Measures** • c-statistic	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Outcome(s) • Mortality *All-cause mortality* **Additional comments** *Hazard ratios were reported, but data was not extracted as it was not stated whether these were adjusted for potential confounding variables*.	
Pothirat (2015)	Detection of acute deterioration in health status visit among COPD patients by monitoring COPD assessment test score.	**Study type** • Prospective cohort study **Study details** • Study location *Thailand* • Study setting *Chest clinic of Chiang Mai University Hospital* • Study dates *Not stated, but participants were recruited from October 2010 to December 2011 and monitored for 15 months.* • Duration of follow-up *15 months* • Loss to follow-up *Data was available for 140/140 (100%) of participants.* • Sources of funding *Not stated*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • High risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Clinically stable COPD *≥ 6 weeks before enrolment.* • Age *≥ 40 years* • Smoking history *Smokers or ex-smokers with a smoking history of 10 or more pack-years.* **Exclusion criteria** • Asthma or history of asthma • Inability to perform the required tests *Unable to complete questionnaires* • Cardiopulmonary disease **Sample characteristics** • Sample size *140* • % female *43.6* • Mean age (SD) *71.1 (8.4)* • FEV1 %, predicted (mean (SD)) *47.4 (18.2)* **Relevant prognostic factor(s)** • CAT (COPD Assessment Test)	*The optimum cut-off for the CAT was determined and used to calculate the sensitivity and specificity of the test.* • Low risk of bias *For the c-statistic data* **Overall risk of bias** • Low *Data for sensitivity and specificity was not used as it was not in an accessible format.* **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Thai version • FEV1 **Measures** • c-statistic • Sensitivity and specificity **Outcome(s)** • Exacerbations *Moderate exacerbation was defined as a worsening of two or more of the following major symptoms for 2 or more consecutive days: breathlessness, sputum volume, or sputum purulence, requiring treatment with systemic corticosteroids and/or antibiotics. Mild exacerbation was defined as worsening of COPD symptoms more than the usual daily variations by patient's global assessment or worsening of symptoms requiring increased dosage, shortened dosage interval, or additional long-acting bronchodilators, but not systemic corticosteroids and/or antibiotics, by physician's global assessment.* **Additional comments** *A 2x2 table could not be calculated from the sensitivity and specificity data provided as the number of people having an exacerbation was not stated.*	
Puhan (2009)	Expansion of the prognostic assessment of patients with chronic obstructive	**Named study cohort** *Swiss Barmelweid cohort* **Study type** • Prospective cohort study	**Participant selection** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	pulmonary disease: the updated BODE index and the ADO index	**Study details** • Study location *Switzerland* • Study setting *Secondary care hospital* • Study dates *May 2004 to December 2005* • Duration of follow-up *At least 30 months* • Loss to follow-up *No information provided so it appears that 100% of participants were included in the analysis* • Sources of funding *Swiss National Science Foundation; Klinik Barmelweid; Fondo de Investigacion Sanitaria Ministry of Health, Spain; Agencia d'Avaluacio de Tecnologia i Recerca Mediques, Catalonia Government; Spanish Society of Pneumology and Thoracic Surgery; Catalan Foundation of Pneumology; Red RESPIRA; Red RCESP; Fondo de Investigacion Sanitaria; Fundacio La Marato de TV3; Novartis Farmaceutica, Spain.* **Inclusion criteria** • Pulmonary function test results *FEV1/FVC <0.7 and FEV1 <80% (GOLD stages II to IV)* • After a respiratory rehabilitation programme **Exclusion criteria** • No measurement of 6-min walk distance because of neurological or musculoskeletal comorbidities	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Language other than German • No informed consent **Sample characteristics** • Sample size *232* • % female *40%* • Mean age (SD) *72.2 years (9.1)* • Smoking details *Current smokers: 18%* • Comorbidities *Cardiovascular disease: 38% Diabetes: 18%* • FEV1 %, predicted (mean (SD)) *45.2% (16.2)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) *Original BODE and updated BODE in the Swiss cohort* **Measures** • c-statistic **Outcome(s)** • Mortality	

Author (year)	Title	Study characteristics	Quality assessment
		Additional comments *The study also had a Spanish cohort. However, the Spanish cohort did not meet our inclusion criteria because participants did not have a stable COPD as they were recruited when they admitted to hospital for their first exacerbation*	
Soler-Cataluna (2009)	Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients.	**Study type** • Prospective cohort study **Study details** • Study location *Spain* • Study setting *Outpatient clinics in Valencia, Spain.* • Study dates *Not stated, but participants were enrolled between January 1999 and June 2004.* • Duration of follow-up *Mean follow-up of 36 (SD 24).* • Loss to follow-up *Not stated; it appears that data was analysed for 185/185 participants.* • Sources of funding *Not stated.* **Inclusion criteria** • Clinically stable COPD *No exacerbations in the month before enrolment in the study.* • Smoking history *> 10 pack-years* • Diagnosis of COPD	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		According to the latest GOLD criteria. • Pulmonary function test results *FEV1/FVC <0.7* **Exclusion criteria** • Other respiratory diseases or respiratory related diseases *Bronchiectasis, cystic fibrosis, upper airway obstruction or bronchiolitis related to systemic pathology.* • Asthma or history of asthma **Sample characteristics** • Sample size *185* • % female *0.0* • Mean age (SD) *71 years (9)* • Smoking details *Current smoker, mean (SD): 33% (17.8)* • Comorbidities *Charlson index, mean (SD): 0.80 (0.96)* • FEV1 %, predicted (mean (SD)) *47.9 (15.5)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD))	**Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• e-BODE (BODE plus exacerbations) • BODEx index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exacerbations) **Multivariate regression model adjusted covariates** • Comorbidities • Exacerbation frequency • Blood gases *PaO2, PaCO2* **Measures** • Hazard ratios **Outcome(s)** • Mortality	
Stolz (2014a)	Mortality risk prediction in COPD by a prognostic biomarker panel.	**Named study cohort** *Pro-ProCOLD (procalcitonin-guided antibiotic therapy in acute exacerbations of COPD: a randomised trial) study. Predicting Outcome using Systemic Markers in Severe Exacerbations of COPD (PROMISE-COPD) study.* **Associated studies** *Stolz 2014b Stolz Daiana, Kostikas Kostantinos, Blasi Francesco, Boersma Wim, Milenkovic Branislava, Lacoma Alicia, Louis Renaud, Aerts Joachim G, Welte Tobias, Torres Antoni, Rohde Gernot G. U, Boeck Lucas, Rakic Janko, Scherr Andreas, Hertel Sabine, Giersdorf Sven, and Tamm Michael (2014) Adrenomedullin refines mortality prediction by the BODE index in COPD: the "BODE-A" index. The European respiratory journal 43, 397-408.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
		Study type • Prospective cohort study **Study details** • Study location *Switzerland* • Study setting *University Hospital Basel, Basel, Switzerland.* • Study dates *Not stated* • Duration of follow-up *5 years* • Loss to follow-up *243/257 (94.6%) of the participants completed the trial or data was available in the case of death.* • Sources of funding *The cohort studies were supported by funding from the Clinic of Pulmonary Medicine and Respiratory Cell Research of the University Hospital Basel, Switzerland. D. Stolz was supported by the Swiss National Foundation (PP00P3_128412/1).* **Inclusion criteria** • Clinically stable COPD *Moderate-to-very severe COPD based on anamnesis, physical examination and spirometry performed ≥4 weeks after resolution of the latest exacerbation.* • Age *≥ 40 years old*	**Sample size and participant flow** • Unclear risk of bias *Participants with missing biomarker data (n=14/243) were excluded from the analysis* **Analysis** • Unclear risk of bias *Model was not adjusted for all the confounding variables required by our review protocol.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Smoking history *≥10 pack-years* **Exclusion criteria** • A primary diagnosis other than COPD as the main respiratory disease *For example bronchiectasis, asthma or pulmonary fibrosis.* • Immunosuppression *Including AIDS or a history of organ transplantation, or current chronic steroid use (>20 mg prednisolone equivalent per day).* • A rapid fatal disease • A musculoskeletal disorder preventing walking **Sample characteristics** • Sample size *Validation cohort =257* • % female *19.8%* • Mean age (SD) *66.1 years (10.5)* • Smoking details • Duration of COPD symptoms, months (mean (SD)) *77 (76)* • Comorbidities *% (n) Arterial hypertension: 55.1 (134) Coronary arterial disease: 32.1 (78) Congestive heart failure: 24.7 (60) Myocardial infarction Pulmonary hypertension: 18.9 (46) Malignancy: 8.2 (20) Diabetes mellitus: 12.7 (31) Renal failure: 12.7 (31)*	

Author (year)	Title	Study characteristics	Quality assessment
		• FEV1 %, predicted (mean (SD)) *47.0 (16.6)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • FEV1 **Multivariate regression model adjusted covariates** • Smoking status • Comorbidities *Including arterial hypertension, cardiopathy, malignancy, diabetes mellitus and renal failure (derivation cohort), or age-adjusted Charlson score (validation cohort).* • Gender • FEV1 %, predicted **Measures** • c-statistic • Hazard ratios • Odds ratios **Outcome(s)** • Mortality	
Stolz (2014b)	Adrenomedullin refines mortality prediction by the BODE index in COPD: the	**Named study cohort** *Predicting Outcome using Systemic Markers in Severe Exacerbations of Chronic Obstructive Pulmonary Disease (PROMISE-COPD) Study.*	**Participant selection** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	"BODE-A" index.	**Associated studies** *Stolz Daiana, Meyer Anja, Rakic Janko, Boeck Lucas, Scherr Andreas, and Tamm Michael (2014a) Mortality risk prediction in COPD by a prognostic biomarker panel. The European respiratory journal 44, 1557-70.* **Study type** • Prospective cohort study **Study details** • Study location *Belgium, Germany, Greece, Italy, Netherlands, Serbia, Spain, and Switzerland.* • Study setting *11 European hospital pneumology departments* • Study dates *November 2008 to October 2011* • Duration of follow-up *2 years* • Loss to follow-up *549/638 (86.05%) participants were analysed* • Sources of funding *Pulmonary Medicine Clinic, University Hospital Basel, Basel, Switzerland; Swiss National Foundation; and Thermo Scientific Biomarkers (formerly Brahms AG), Hennigsdorf, Germany.* **Inclusion criteria** • Clinically stable COPD *Moderate to very severe COPD based on anamnesis, physical examination and spirometry ≥4 weeks after the latest exacerbation resolved*	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Age *≥40 years* • Smoking history *≥10 pack-years* **Exclusion criteria** • Immunosuppression *Including AIDS, organ transplantation or chronic steroids (>20 mg day-1 prednisolone equivalent)* • A musculoskeletal disorder preventing walking *or neuromuscular disorder* • Main respiratory disorder other than COPD • Death expected within 6 months **Sample characteristics** • Sample size *549* • % female *30.2%* • Mean age (SD) *66.0 years (11.4)* • Smoking details *Current smoker 33.3% Pack years mean 45.0 (SD 31.1)* • Comorbidities *Arterial hypertension 51.4% Coronary arterial disease 23.7% Congestive heart failure 14.4% Myocardial infarction 9.3% Pulmonary hypertension 9.7% Malignancy 3.8% Diabetes mellitus 12.0% Renal failure 6.0% Adjusted Charlson score median 4 (interquartile range: 3 to 5)*	

Author (year)	Title	Study characteristics	Quality assessment
		• FEV1 %, predicted (mean (SD)) *Post-bronchodilator FEV1 % pred 48.9 (18.3)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD)) • BOD index (dyspnoea/ breathlessness, FEV1 and BMI) **Multivariate regression model adjusted covariates** • Unspecified **Measures** • c-statistic • Sensitivity and specificity • Hazard ratios **Outcome(s)** • Mortality *All-cause mortality at 1 and 2 years* **Additional comments** *Data on BODE was not extracted because Stolz 2014a already reported c-statistic for BODE. It is unclear whether some of the same participants are included in both studies.*	
Suetomo (2014)	COPD assessment tests scores are associated with	**Study type** • Prospective cohort study	**Participant selection** • Low risk of bias

Author (year)	Title	Study characteristics	Quality assessment
	exacerbated chronic obstructive pulmonary disease in Japanese patients.	**Study details** • Study location *Japan* • Study setting *Hospital* • Study dates *September 2011 to August 2013* • Duration of follow-up *1 year* • Loss to follow-up *11.5%* • Sources of funding *No funding sources associated with this work* **Inclusion criteria** • Clinically stable COPD • History of exacerbations *No history of exacerbations while receiving systemic antibiotics and corticosteroids* • Hospitalisations for 4 weeks prior study entry **Exclusion criteria** • Asthma or history of asthma • Bronchiectasis • Current malignancy • Hepatitis B or C • Dementia • Main respiratory disorder other than COPD *Interstitial pneumonia. Pneumoconiosis.*	**Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		• Cerebro- or cardio-vascular disease • Cirrhosis • Chronic kidney disease • Psychological disease **Sample characteristics** • Sample size *123* • % female *High CAT 15.6% Low CAT 10.2%* • Mean age (SD) *High CAT 69.4 (5.3) Low CAT 65.1 (6.1)* • Smoking details *Current smoker High CAT 35.9% Low CAT 28.8% Smoking index, packs per year mean (SD) High CAT 60.1 (25.4) Low CAT 52.2 (27.8)* • Comorbidities *Hypertension High CAT 15.6% Low CAT 13.6% Hyperlipidaemia High CAT 4.7% Low CAT 5.1% Diabetes High CAT 26.6% Low CAT 27.1%* • FEV1 %, predicted (mean (SD)) *High CAT 46.0 (13.8) Low CAT 52.9 (13.0)* **Relevant prognostic factor(s)** • CAT (COPD Assessment Test) *High CAT >10 points Low CAT <10 points* • GOLD 2009 **Measures** • c-statistic	

Author (year)	Title	Study characteristics	Quality assessment
		• Sensitivity and specificity • Odds ratios **Outcome(s)** • Hospitalisations • Exacerbations *Exacerbation was defined on the basis of symptom-based diagnosis such as increased cough and sputum production, a change of sputum colour, and worsening of breathlessness from a stable state and beyond-normal day-to-day variations, i.e., showing acute onset and necessitating a change in regular medication, in accordance with a previous report. Moderate exacerbations required a prescription for antibiotics and/or systemic corticosteroids, and severe exacerbations required hospitalization. COPD-related death was also counted as severe exacerbation.*	
Sundh (2012a)	Clinical COPD Questionnaire score (CCQ) and mortality	**Study type** • Prospective cohort study **Study details** • Study location *Sweden* • Study setting *The population was recruited from the central hospital and one district hospital plus eight primary care centres in each of seven Swedish county councils.* • Study dates *2005 to 2010* • Duration of follow-up *5 years* • Loss to follow-up	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow**

Author (year)	Title	Study characteristics	Quality assessment
		For the CCQ analysis (Sundh 2012a) data was available for 970/ 1111 (87.3%) of participants. *For the DOSE analysis (Sundh 2012b) data was available for 562 /1111 (50.6%) of participants.* • Sources of funding *Grants from the county councils of the Uppsala-Örebro Health Care region, the Swedish Heart and Lung Association, the Swedish Asthma and Allergy Association, the Bror Hjerpstedts Foundation, and the Örebro Society of Medicine.* **Inclusion criteria** • Age *Between 18- 75 years old* • Diagnosis of COPD *ICD-10 code J44 in the medical records during the period of 2000–2003.* **Exclusion criteria** • None reported **Sample characteristics** • Sample size *Both studies: 1111 people consented to participate.* *Sundh 2012a CCQ study n= 970 with complete data.* *Sundh 2012b DOSE study n= 562 with complete data.* • % female *CCQ (n=970): 57.5 DOSE (n= 562): 57.1* • Mean age (SD) *CCQ (n=970): mean and median not stated. <50 years: 64 51–60 years: 194 61–70*	• Unclear risk of bias *Data was available for 87% of participants for the CCQ index. Data was only available for 50.6% participants to enable calculation of the DOSE index. Missing data was not imputed in either case.* **Analysis** • Unclear risk of bias *Model was not adjusted for all the confounding variables required by our review protocol.* **Overall risk of bias** • Moderate *For DOSE index due to low number of participants with complete data to enable calculation of the index.* • Low

Author (year)	Title	Study characteristics	Quality assessment
		years: 490 >70: 222 DOSE (n= 562): mean and median not stated. <50 years: 40 51–60 years: 122 61–70 years: 288 >70: 112 • Smoking details *DOSE: not stated. CCQ: Smoking never: 61 Ex- smoker: 574 Occasional smoker: 61 Current smoker: 273* • FEV1 %, predicted (mean (SD)) *CCQ: mean not stated. DOSE: mean not stated. FEV1 % predicted <80 = 412 people* **Relevant prognostic factor(s)** • CCQ (Clinical COPD Questionnaire score) **Multivariate regression model adjusted covariates** • Age • Comorbidities *Heart disease* • Gender **Measures** • Hazard ratios **Outcome(s)** • Mortality	*For CCQ index* **Directness** • Directly applicable
Sundh (2012b)	The Dyspnoea, Obstruction, Smoking, Exacerbation	**Associated studies** *Sundh J, Janson C, Lisspers K, Montgomery S, Stallberg B. Clinical COPD Questionnaire*	Please refer to Sundh 2012a

Author (year)	Title	Study characteristics	Quality assessment
	(DOSE) index is predictive of mortality in COPD	*score (CCQ) and mortality. International journal of chronic obstructive pulmonary disease 2012a; 7:833-842.* **Additional comments** Study characteristics are shown in Sundh 2012a	
Thabut (2014)	Performance of the BODE index in patients with alpha1-antitrypsin deficiency-related COPD.	**Study type** • Prospective cohort study **Study details** • Study location *France* • Study setting *Not stated, but the study aimed to recruit all French patients who fulfilled the inclusion criteria and thus was based at multiple sites across France.* • Study dates *January 2006 to December 2012* • Duration of follow-up *Median follow-up time was 31.4 months (range 1–91.3 months).* • Loss to follow-up *140/215 of the study participants were alive the end of the study. Twenty patients died, 22 underwent lung transplantation, 5 withdrew from the study and 4 were lost to follow-up. There is no information provided about the missing 24 people. Mortality data was available for 160/215 (74.4%) of the participants, assuming the 22 who had a lung transplant were excluded.* • Sources of funding *Laboratoire francais du Fractionnement et des Biotechnologies.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Low risk of bias *Although there was a relatively large loss to follow-up, the study used an inverse probability of censoring weighted survival estimator to fill in missing data for those people (22)*

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Pulmonary function test results *FEV1/FVC <0.7* • alpha1-antitrypsin level <0.5 gL • Emphysema diagnosed by computed tomography **Exclusion criteria** • None reported **Sample characteristics** • Sample size *215* • % female *37.3* • Mean age (SD) *50.8 years (10.9)* • Smoking details *Smoking status, n (%) Never-smokers 21 (11.0) Current smokers 7 (3.7) Smoking history pack-years, mean (SD): 18.2 (16.3)* • Comorbidities *Charlson index, mean (SD): 2.7 (1.3)* • FEV1 %, predicted (mean (SD)) *42.5 (19.9)* **Relevant prognostic factor(s)** • BODE index (dyspnoea/ breathlessness (mMRC), BMI, FEV1 and exercise (6MWD))	*who had undergone lung transplantation.* **Analysis** • Unclear risk of bias *The model was not adjusted for age, smoking and comorbidities.* **Overall risk of bias** • Low **Directness** • Directly applicable

Author (year)	Title	Study characteristics	Quality assessment
		Multivariate regression model adjusted covariates • Augmentation therapy **Measures** • c-statistic • Hazard ratios **Outcome(s)** • Mortality	
Varol (2014)	Assessing the effectiveness of the COPD Assessment Test (CAT) to evaluate COPD severity and exacerbation rates.	**Study type** • Prospective cohort study **Study details** • Study location *Turkey* • Study setting *One government hospital and 2 chest disease education and training hospitals.* • Study dates *April 2011 to February 2012* • Duration of follow-up *10 months* • Loss to follow-up *Data was analysed for 165/165 (100%) study participants.* • Sources of funding *Not stated.*	**Participant selection** • Low risk of bias **Predictors** • Low risk of bias **Outcome** • Low risk of bias **Sample size and participant flow** • Unclear risk of bias *It is unclear whether the study duration was long enough to detect a sufficient*

Author (year)	Title	Study characteristics	Quality assessment
		Inclusion criteria • Clinically stable COPD • Age *≥ 40 years old* • Smoking history *≥ 10 pack-years or a history of biomass exposure* • Pulmonary function test results *FEV1 <80% % predicted after bronchodilator use. FEV1/FVC <0.7* **Exclusion criteria** • A primary diagnosis other than COPD as the main respiratory disease *Asthma or other chronic respiratory diseases* • Uncontrolled comorbidities **Sample characteristics** • Sample size *165* • % female *9.7* • Mean age (SD) *65.0 years (9.9)* • Smoking details *Smoking history, number (%) Biomass: 11 (6) Ex-smoker: 63 (37) Current smoker: 92 (55) Pack-years, mean (SD): 52 (23.8)* • Comorbidities *79 (47%) had a comorbidity.*	*number of exacerbations for the analysis and the study does not state how many exacerbations occurred.* **Analysis** • Low risk of bias **Overall risk of bias** • Low **Directness** • Directly applicable

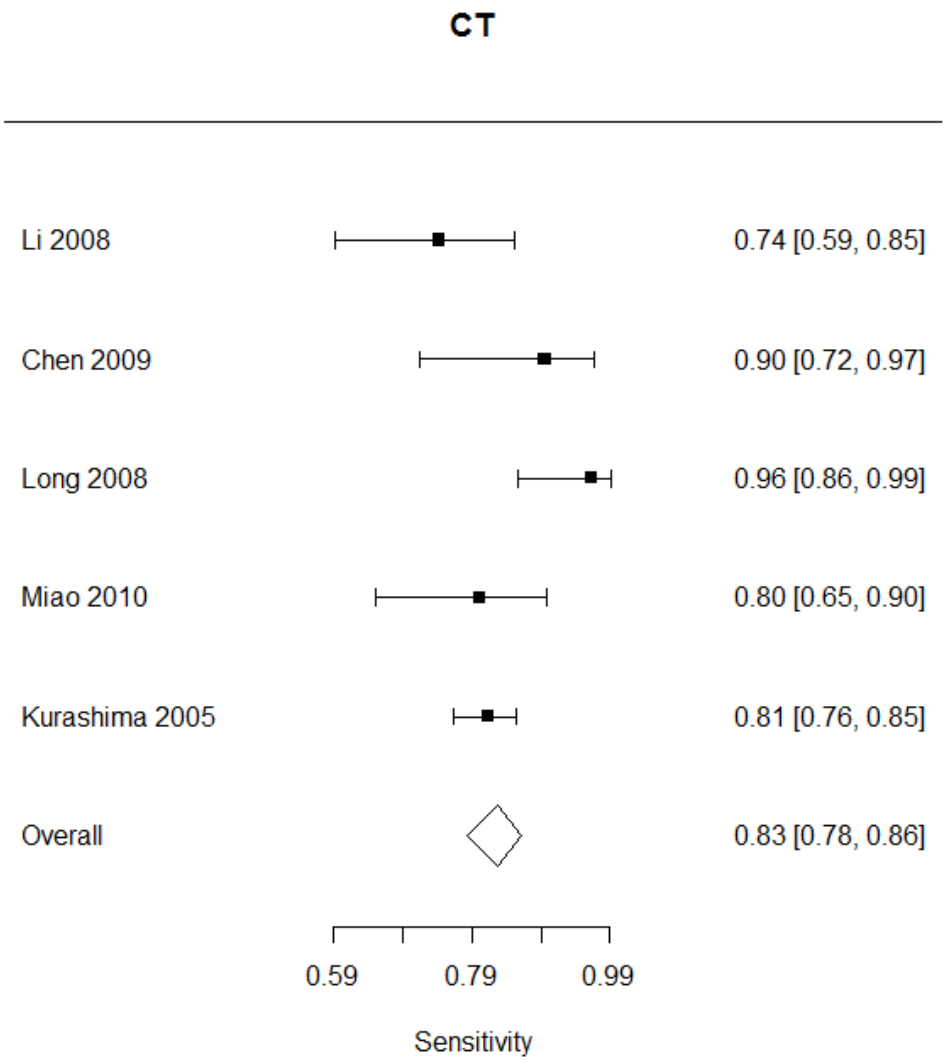
Author (year)	Title	Study characteristics	Quality assessment
		• FEV1 %, predicted (mean (SD)) *43.7 (14.8)* **Relevant prognostic factor(s)** • CAT (COPD Assessment Test) *Turkish version* **Measures** • c-statistic **Outcome(s)** • Exacerbations *Defined here as an increase in sputum purulence and volume, and worsening of breathlessness which results in an unexpected visit to a doctor or emergency department and causes a change in disease management.*	

Appendix F – Forest plots

Confirming COPD diagnosis

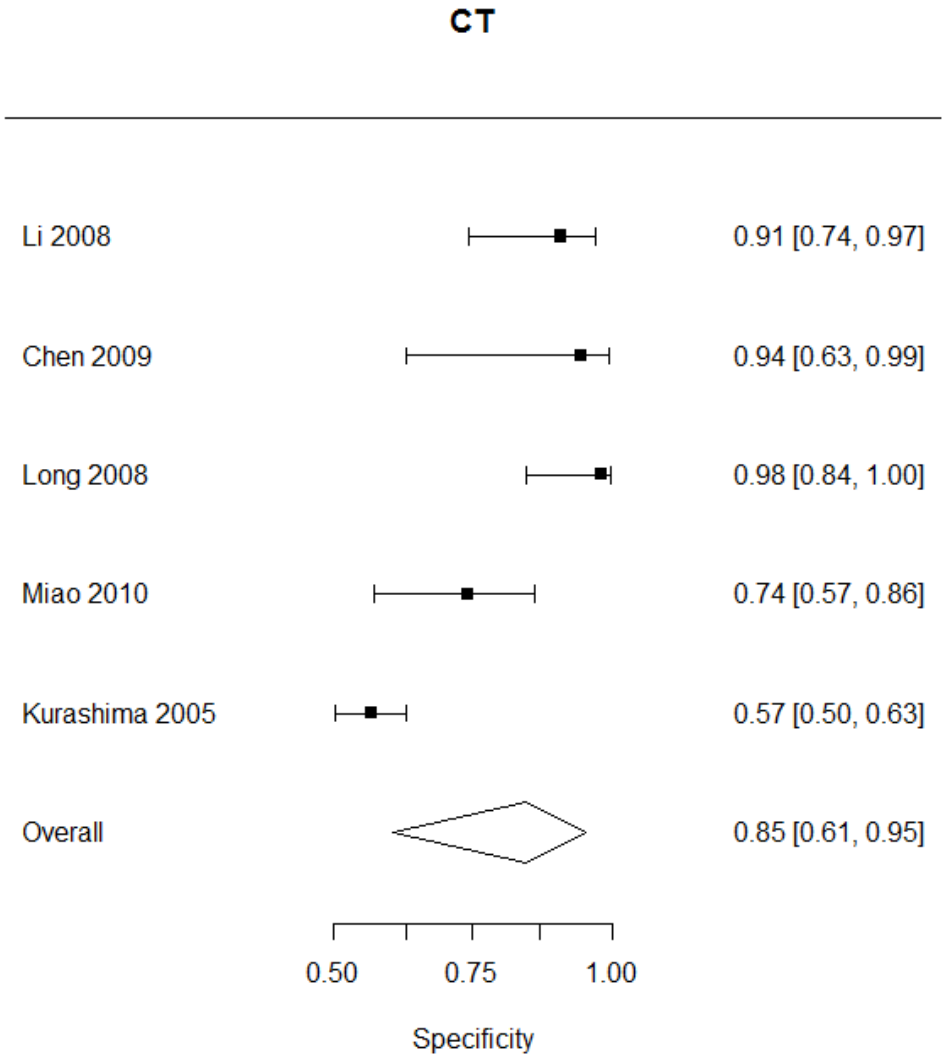
Computed tomography (reference standard: pulmonary function tests)

Sensitivity



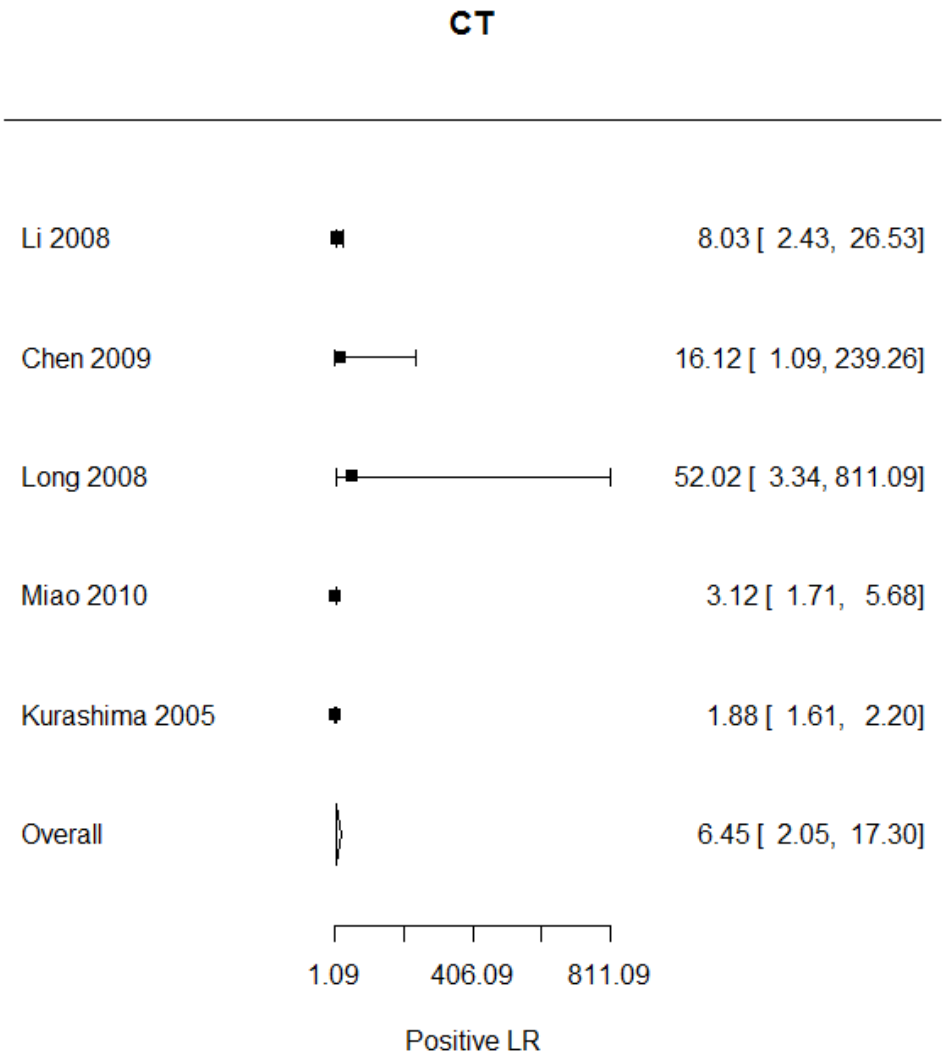
I^2 53.13%

Specificity

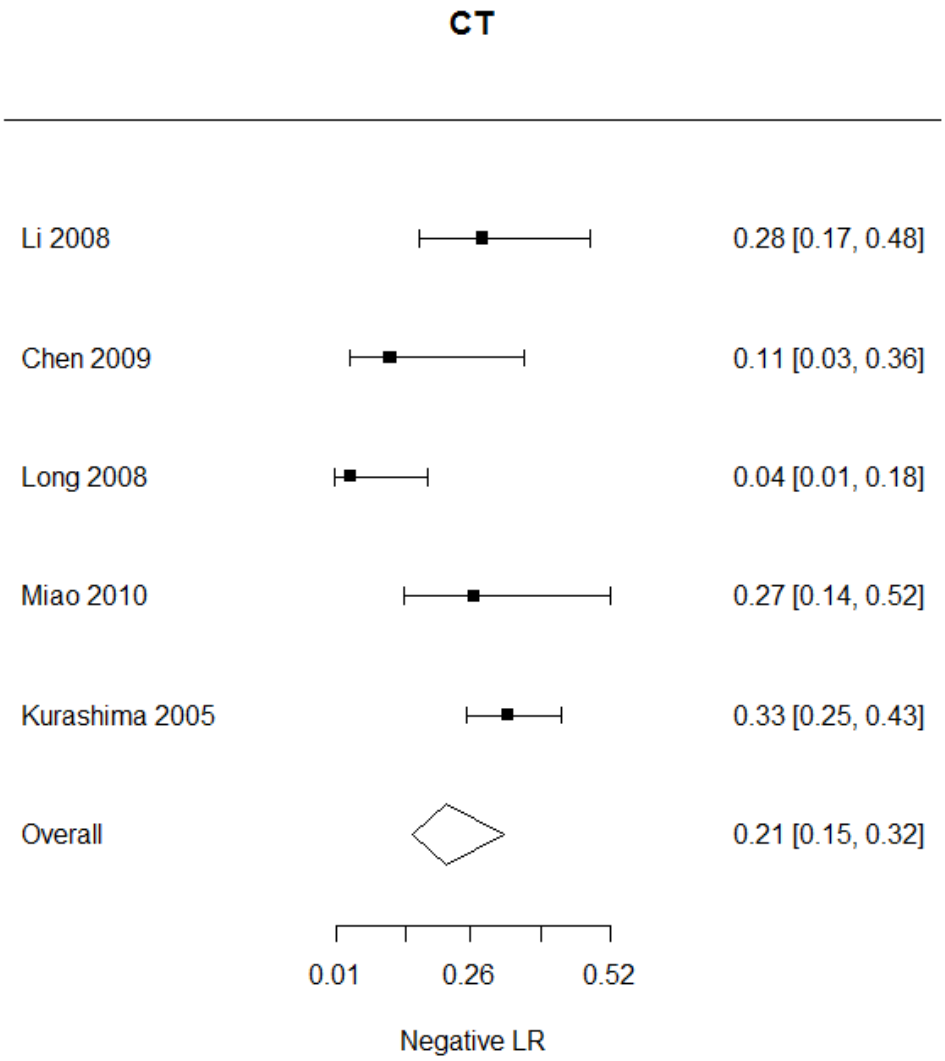


I^2 87.83%

Positive likelihood ratios



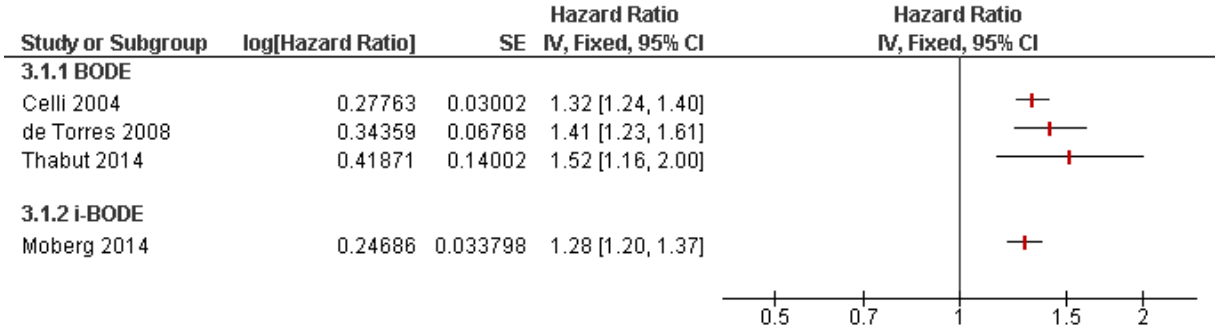
Negative likelihood ratios



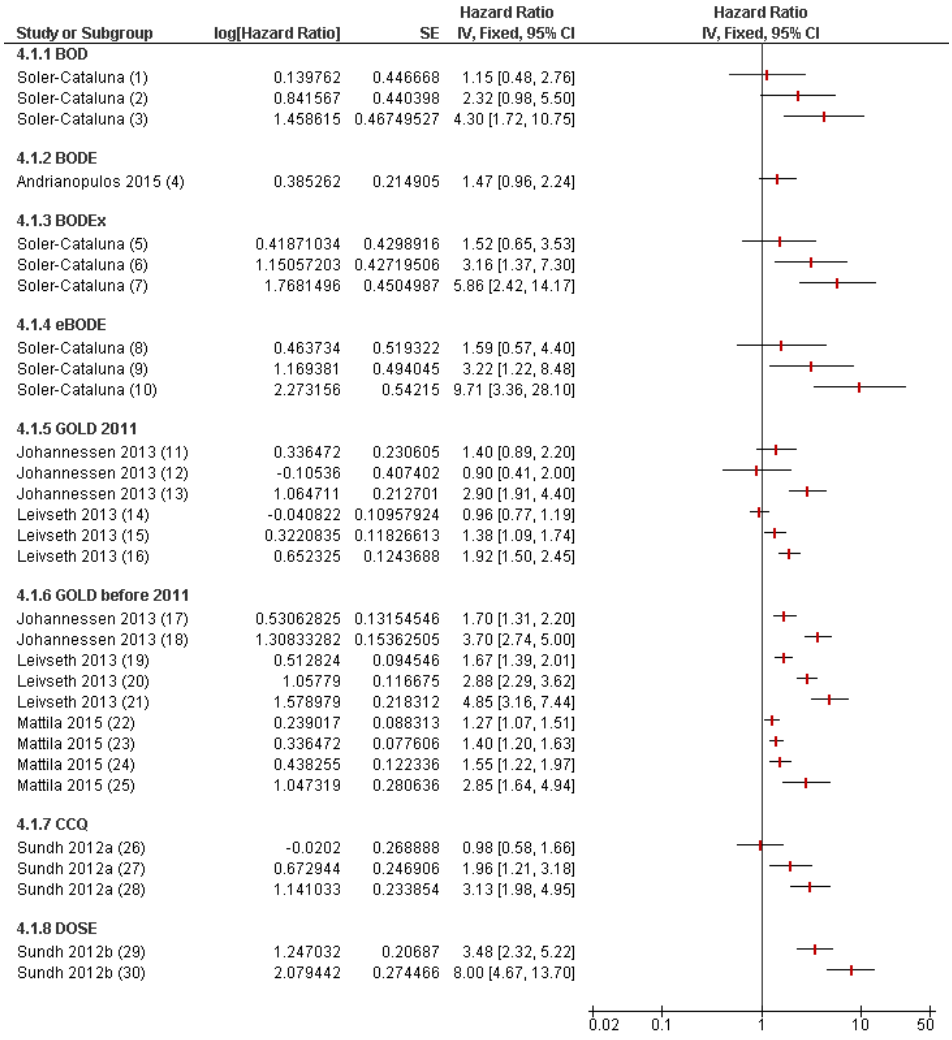
Predicting outcomes using multidimensional severity assessment indices for people with an existing diagnosis of COPD

All-cause mortality

Hazard ratios, per unit increase in index

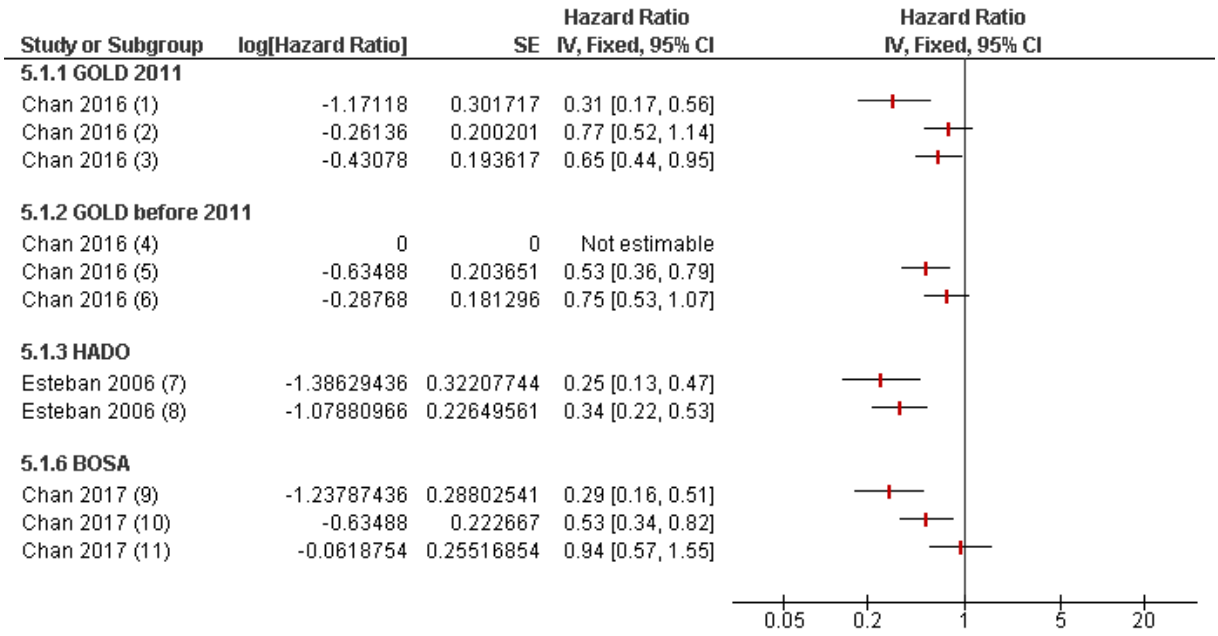


Hazard ratios by category (low risk reference)



Footnotes
(1) Quartile 2 (reference quartile 1)
(2) Quartile 3
(3) Quartile 4
(4) ≥4 (reference < 4)
(5) Quartile 2 (reference quartile1)
(6) Quartile 3
(7) Quartile 4
(8) Quartile 2 (reference quartile 1)
(9) Quartile 3
(10) Quartile 4
(11) Group B (reference group A)
(12) Group C
(13) Group D
(14) Group B (reference group A)
(15) Group C
(16) Group D
(17) GOLD 2007 Stage 3 (reference stage 2)
(18) GOLD 2007 stage 4
(19) Stage 2 (reference stage 1)
(20) Stage 3
(21) Stage 4
(22) Stage 1 (reference no COPD)
(23) Stage 2
(24) Stage 3
(25) Stage 4
(26) ≥1, <2 (reference <1)
(27) ≥ 2, <3
(28) ≥3
(29) 4-5 (reference 1-3)
(30) 6-7

Hazard ratios by category (high risk reference)

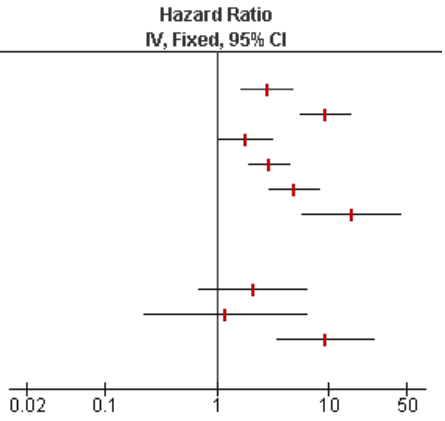


Footnotes
(1) Group A (versus group D)
(2) Group B
(3) Group C
(4) Stage 1 (reference stage 4)- no events
(5) Stage 2
(6) Stage 3
(7) Mild (reference severe)
(8) Moderate
(9) Score 0-3 (reference 7-12)
(10) Score 4-5
(11) Score 6

Mortality due to respiratory causes

Hazard ratios by category (low risk reference)

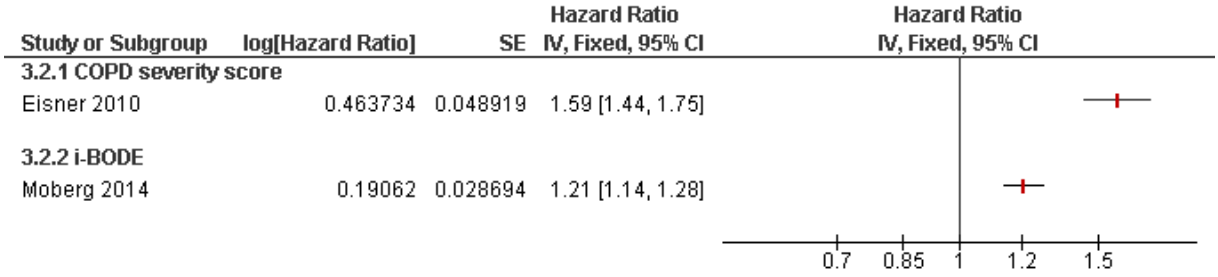
Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Fixed, 95% CI
4.2.1 GOLD 2007			
Johannessen 2013 (1)	1.02961942	0.26425668	2.80 [1.67, 4.70]
Johannessen 2013 (2)	2.2300144	0.26716649	9.30 [5.51, 15.70]
Mattila 2015 (3)	0.593327	0.284309	1.81 [1.04, 3.16]
Mattila 2015 (4)	1.071584	0.209194	2.92 [1.94, 4.40]
Mattila 2015 (5)	1.599388	0.266772	4.95 [2.93, 8.35]
Mattila 2015 (6)	2.769459	0.518994	15.95 [5.77, 44.11]
4.2.2 GOLD 2011			
Johannessen 2013 (7)	0.74193734	0.56855135	2.10 [0.69, 6.40]
Johannessen 2013 (8)	0.18232156	0.85406961	1.20 [0.23, 6.40]
Johannessen 2013 (9)	2.24070969	0.51315373	9.40 [3.44, 25.70]



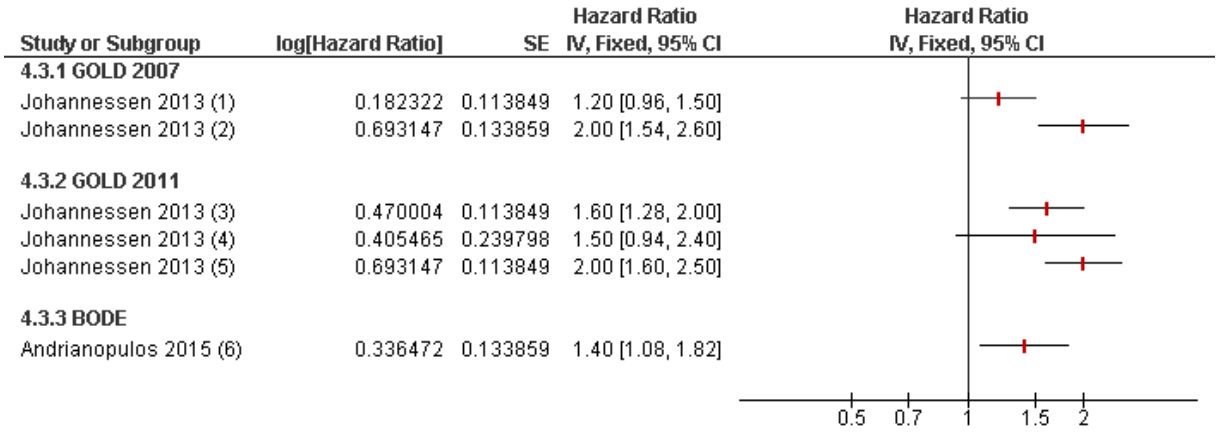
Footnotes
(1) Stage 3 (reference stage 2)
(2) Stage 4
(3) Stage 1 (reference no COPD)
(4) Stage 2
(5) Stage 3
(6) Stage 4
(7) Group B (reference group A)
(8) Group C
(9) Group D

All-cause hospitalisations

Hazard ratios, per unit increase in index



Hazard ratios by category (low risk reference)

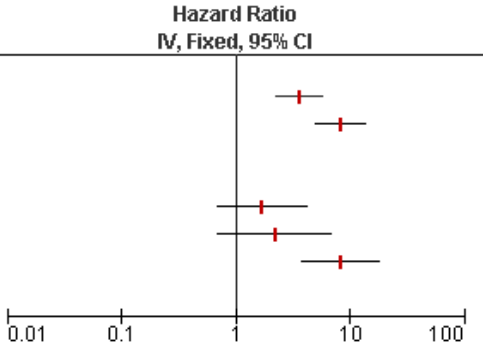


Footnotes
(1) Stage 3 (reference stage 2)
(2) Stage 4
(3) Group B (versus group A)
(4) Group C
(5) Group D
(6) ≥3 (reference <3)

Respiratory specific hospitalisations

Hazard ratios by category (low risk reference)

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Fixed, 95% CI
4.4.1 GOLD 2007			
Johannessen 2013 (1)	1.28093	0.24333	3.60 [2.23, 5.80]
Johannessen 2013 (2)	2.11626	0.25939	8.30 [4.99, 13.80]
4.4.2 GOLD 2011			
Johannessen 2013 (3)	0.53062825	0.461457	1.70 [0.69, 4.20]
Johannessen 2013 (4)	0.788457	0.583196	2.20 [0.70, 6.90]
Johannessen 2013 (5)	2.104134	0.395441	8.20 [3.78, 17.80]

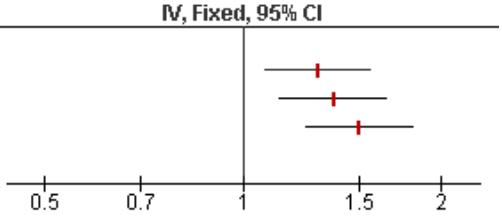


Footnotes
(1) Stage 3 (reference stage 2)
(2) Stage 4
(3) Group B (reference group A)
(4) Group C
(5) Group D

Exacerbations

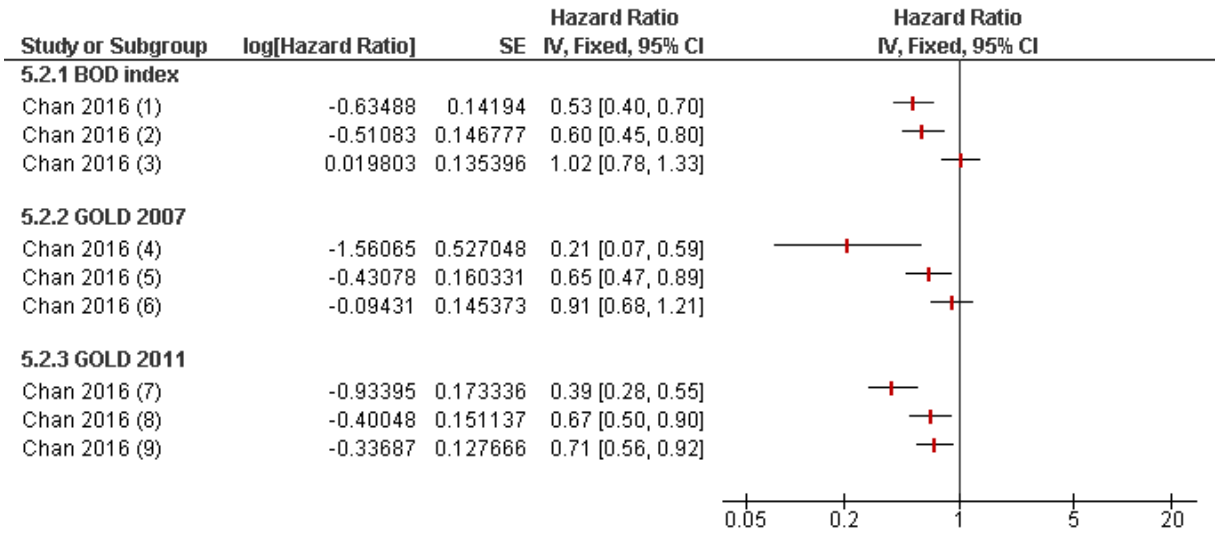
Hazard ratios by category (low risk reference)

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Fixed, 95% CI
4.5.1 CAT			
Lee 2014 (1)	0.262364	0.093021	1.30 [1.08, 1.56]
Lee 2014 (2)	0.314811	0.09488	1.37 [1.14, 1.65]
Lee 2014 (3)	0.405465	0.095848	1.50 [1.24, 1.81]



Footnotes
(1) 10-19 (reference 0-9)
(2) 20-29
(3) 30-40

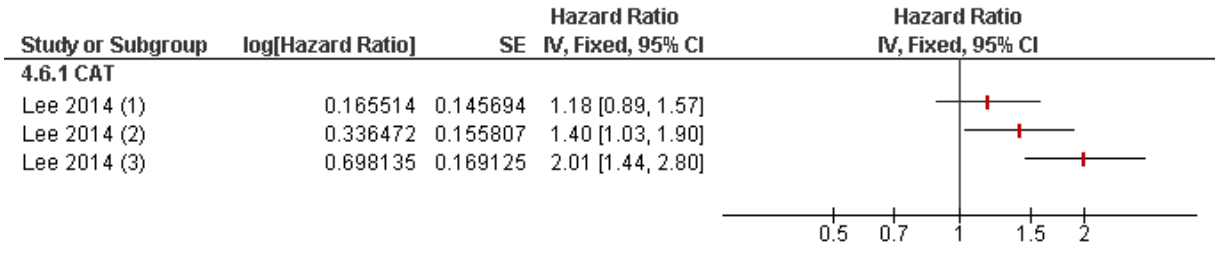
Hazard ratios by category (high risk reference)



Footnotes
(1) Quartile 1 (reference quartile 4)
(2) Quartile 2
(3) Quartile 3
(4) Stage 1 (reference stage 4)
(5) Stage 2
(6) Stage 3
(7) Group A (versus group D)
(8) Group B
(9) Group C

Severe exacerbations

Hazard ratios by category (low risk reference)



Footnotes
(1) 10-19 (reference 0-9)
(2) 20-29
(3) 30-40

Appendix G – GRADE tables

Confirming COPD or emphysema diagnosis

Computed tomography – COPD diagnosis

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
16 Multi-slice computed tomography (reference standard: full expiration average lung density)										
1 (Li 2012)[1]	SR	66	75.0 (59.5, 86.0)	92.3 (73.9, 98.1)	LR+ 9.75 (2.54, 37.36)	Serious[2]	N/A	Serious[3]	Not serious	Low
					LR- 0.27 (0.15, 0.46)	Serious[2]	N/A	Serious[3]	Not serious	Low
Low-dose computed tomography (reference standard: emphysema index in expiration)										
1 (Li 2012)[4]	SR	31	89.6 (69.9, 97.0)	94.4 (49.5, 99.7)	LR+ 16.12 (1.08, 239.25)	Serious[2]	N/A	Serious[3]	Serious[5]	Very low
					LR- 0.11 (0.03, 0.36)	Serious[2]	N/A	Serious[3]	Not serious	Low
16 Multi-slice computed tomography (reference standard: pixel index in maximum expiratory)										
1 (Li 2012)[6]	SR	66	96.3 (83.8, 99.3)	98.1 (76.4, 99.9)	LR+ 52.02 (3.33, 811.09)	Serious[2]	N/A	Serious[3]	Not serious	Low
					LR- 0.03 (0.00, 0.17)	Serious[2]	N/A	Serious[3]	Not serious	Low
16 Multi-slice computed tomography (reference standard: blood flow)										
1 (Li 2012)[7]	SR	69			LR+ 3.24	Serious[2]	N/A	Serious[3]	Serious[5]	Very low

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			81.1 (65.3, 90.7)	75.0 (57.4, 87.0)	(1.74, 6.02)					
					LR- 0.25 (0.12, 0.50)	Serious[2]	N/A	Serious[3]	Serious[5]	Very low
High resolution computed tomography (reference standard: GOLD)										
1 (Kurashima 2005)	Cross-sectional	516	81.4 (76.4, 85.6)	56.8 (50.4, 63.0)	LR+ 1.88 (1.61, 2.20)	Serious[2]	N/A	Not serious	Serious[5]	Low
					LR- 0.32 (0.25, 0.42)	Serious[2]	N/A	Not serious	Not serious	Moderate
Computed tomography (reference standard: pulmonary function tests)										
5 (Li 2012)[10]	SR	748	82.7 (78.4, 86.3)	84.6 (60.6, 95.1)	LR+ 6.45 (2.06, 17.30)	Serious[2]	Serious[8]	Serious[3]	Not serious	Very low
					LR- 0.21 (0.15, 0.32)	Serious[2]	Very serious[9]	Serious[3]	Not serious	Very low

1. Data on Li 2008 reported by Li 2012
2. Moderate risk of bias
3. Participants inclusion/exclusion criteria were not reported
4. Data on Chen 2009 reported by Li 2012
5. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)
6. Data on Long 2008 reported by Li 2012
7. Data on Miao 2010 reported by Li 2012
8. I^2 for sensitivity was 53.13%
9. I^2 for specificity was 87.83%
10. Pooled data from the Li 2012 SR

CI: confidence interval; SR: systematic review; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

Chest X-ray – emphysema diagnosis

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Computer-aided procedure to recognise emphysema on digital chest X-ray (reference standard: computed tomography)										
1 (Miniati 2011)	Cross-sectional	107[1]	90.2 (76.7, 96.3)	97.0 (88.7, 99.2)	LR+ 29.78 (7.57, 117.05)	Serious[2]	N/A	Serious[3]	Not serious	Low
					LR- 0.10 (0.04, 0.25)	Serious[2]	N/A	Serious[3]	Not serious	Low

1. Data from validation sample
2. Moderate risk of bias
3. Pulmonary arterial hypertension was suspected in 15% of the sample before CT scan was performed

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

Pulse oximetry – COPD diagnosis

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Arterial oxygen saturation <96% (reference standard: post-bronchodilator spirometry FEV1/FVC <0.70)										
1 (Garcia-Pachon)	Cross-sectional	210	50.0 (37.4, 62.6)	76.3 (68.9, 82.4)	LR+ 2.11 (1.43, 3.10)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
					LR- 0.65 (0.49, 0.86)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
Arterial oxygen saturation <97% (reference standard: post-bronchodilator spirometry FEV1/FVC <0.70)										
1 (Garcia-Pachon)	Cross-sectional	210	63.8 (50.8, 75.1)	53.3 (45.3, 61.1)	LR+ 1.36 (1.05, 1.76)	Very serious[1]	N/A	Not serious	Not serious	Low
					LR- 0.67 (0.46, 0.98)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
Arterial oxygen saturation <98% (reference standard: post-bronchodilator spirometry FEV1/FVC <0.70)										

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Garcia-Pachon)	Cross-sectional	210	79.3 (67.0, 87.9)	36.8 (29.6, 44.8)	LR+ 1.25 (1.05, 1.50)	Very serious[1]	N/A	Not serious	Not serious	Low
					LR- 0.56 (0.32, 0.96)	Very serious[1]	N/A	Not serious	Serious[2]	Very low

1. High risk of bias
2. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

Biomarker: hs-CRP – COPD diagnosis

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
hs-CRP at 2.39mg/L (reference standard: pulmonary function tests)										
1 (Tileman n 2011)	Cross-sectional	210	50.0 (34.0, 65.8)	75.3 (68.3, 81.1)	LR+ 2.02 (1.33, 3.07)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
					LR- 0.66 (0.47, 0.93)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
hs-CRP at 3.5mg/L (reference standard: pulmonary function tests)										
1 (Tileman n 2011)	Cross-sectional	210	41.7 (26.9, 58.1)	82.8 (76.4, 87.7)	LR+ 2.41 (1.45, 4.00)	Very serious[1]	N/A	Not serious	Serious[2]	Very low
					LR- 0.70 (0.53, 0.93)	Very serious[1]	N/A	Not serious	Serious[2]	Very low

1. High risk of bias
2. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

Predicting outcomes using multidimensional severity assessment indices for people with an existing diagnosis of COPD

All-cause mortality

Sensitivity, specificity and likelihood ratios

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE ≥4										
1 (Andrianopoulos 2015)	Prospective	2,010	60.0 (53.0, 67.0)	63.0 (61.0, 66.0)	LR+ 1.62 (1.42, 1.85)[1]	Not serious	N/A	Not serious	Serious[2]	Moderate
					LR- 0.63 (0.53, 0.75)[1]	Not serious	N/A	Not serious	Serious[2]	Moderate
BODE >4										
1 (Stolz 2014b)	Prospective	549	48.8 (34.4, 63.4)	78.7 (74.9, 82.0)	LR+ 2.28 (1.61, 3.24)	Not serious	N/A	Not serious	Serious[3]	Moderate
					LR- 0.65 (0.48, 0.87)	Not serious	N/A	Not serious	Serious[3]	Moderate
BODAS >5										
1 (Ansari 2016)	Prospective	458	71.4 (63.8, 78.0)[4]	60.9 (55.3, 66.2)[4]	LR+ 1.82 (1.53, 2.16)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
					LR- 0.46 (0.36, 0.61)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
BOD >2										
1 (Ansari 2016)	Prospective	458	44.8 (37.1, 52.7)[4]	77.0 (71.9, 81.4)[4]	LR+ 1.94 (1.48, 2.54)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
					LR- 0.71 (0.61, 0.83)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
BODS >4										

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Ansari 2016)	Prospective	458	64.9 (57.1, 72.1)[4]	59.2 (53.6, 64.6)[4]	LR+ 1.59 (1.33, 1.90)[4]	Very serious[5]	N/A	Not serious	Not serious	Low
					LR- 0.59 (0.46, 0.74)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
GOLD >1 (matrix [new classification A to D])										
1 (Ansari 2016)	Prospective	458	94.2 (89.2, 96.9)[4]	16.8 (13.0, 21.4)[4]	LR+ 1.13 (1.06, 1.20)[4]	Very serious[5]	N/A	Not serious	Not serious	Low
					LR- 0.34 (0.17, 0.68)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
GOLD >2 (old GOLD stages 1 to 4)										
1 (Ansari 2016)	Prospective	458	37.7 (30.4, 45.6)[4]	72.4 (67.1, 77.1)[4]	LR+ 1.36 (1.03, 1.79)[4]	Very serious[5]	N/A	Not serious	Not serious	Low
					LR- 0.86 (0.74, 0.99)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
ADO >3										
1 (Ansari 2016)	Prospective	458	64.3 (56.4, 71.5)[4]	64.8 (59.3, 70.0)[4]	LR+ 1.82 (1.50, 2.21)[4]	Very serious[5]	N/A	Not serious	Serious[3]	Very low
					LR- 0.55 (0.43, 0.69)[4]	Very serious[5]	V	Not serious	Serious[3]	Very low

1. We calculated 95% CI for positive and negative likelihood ratios because these were not reported by Andrianopoulos 2015
2. Effect size and 95% confidence interval for likelihood ratio did not reach a defined MID threshold
3. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)
4. Ansari 2016 did not report confidence intervals. Therefore, we calculated 95% CI for sensitivity and specificity as well as positive and negative likelihood ratios and their 95% CI
5. Study at high risk of bias

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

c-statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ADO index (Age, Dyspnoea/ breathlessness and Obstruction)								
1 (Ansari 2016)	Prospective cohort	458	0.70 (0.66, 0.74)	Serious[11]	Very serious[7]	Not serious	Serious[12]	Very low
1 (Esteban 2011)	Prospective cohort	348	0.74 (0.69, 0.80)					
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.65*					
1 (Marin 2013 [Pamplona cohort])	Prospective cohort	190	0.62*					
1 (Marin 2013 [Requena I cohort])	Prospective cohort	174	0.75*					
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.72*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.67*					
1 (Marin 2013 [Seville cohort])	Prospective cohort	596	0.5*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	137	0.60*					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.68*					
1 (Neo 2017)	Prospective cohort	124	0.68 (0.55, 0.82)					
1 (Ou 2014)	Prospective cohort	689	0.70*					
BOD index (BMI, Obstruction, Dyspnoea/ breathlessness)								
1 (Ansari 2016)	Prospective cohort	458	0.62 (0.57, 0.66)	Serious[2]	Serious[10]	Not serious	Not serious	Low
1 (Chan 2016)	Prospective cohort	1,110	0.72 (0.72, 0.72)					
1 (Stolz 2014b)	Prospective cohort	549	0.64*					
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Esteban 2011)	Prospective cohort	348	0.72 (0.66, 0.78)	Serious[11]	Very serious[7]	Not serious	Serious[12]	Very low
1 (Casanova 2005)	Prospective cohort	689	0.80 (0.76, 0.84)					
1 (Imfeld 2006)	Prospective cohort	186	0.74[1]					
1 (Marin 2011)	Prospective cohort	1,398	0.77 (0.74, 0.81)					
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.63*					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Marin 2013 [Pamplona cohort])	Prospective cohort	190	0.56*					
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.64*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.62*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	137	0.59*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.69*					
1 (Puhan 2009 [Swiss Cohort])	Prospective cohort	232	0.67*					
1 (Stolz 2014 [validation cohort])	Prospective cohort	243	0.62*					
1 (Neo 2017)	Prospective cohort	124	0.71 (0.58, 0.86)					
1 (Soler-Cataluna 2009)	Prospective cohort	185	0.75 (0.66, 0.84)					
1 (Pedone 2014)	Prospective cohort	468	0.63*					
BODE ≥4								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Andrianopoulos 2015)	Prospective cohort	2,010	0.67*	Not serious	N/A	Not serious	Not serious	High
BODAS index (BMI, Obstruction, Dyspnoea/ breathlessness, Age and Smoking [pack years])								
1 (Ansari 2016)	Prospective cohort	458	0.72 (0.67, 0.76)	Very serious[9]	N/A	Not serious	Serious[4]	Very low
BODS index (BMI, Obstruction, Dyspnoea/ breathlessness and Smoking [pack years])								
1 (Ansari 2016)	Prospective cohort	458	0.66 (0.61, 0.70)	Very serious[9]	N/A	Not serious	Serious[4]	Very low
BODEx index (BMI, Obstruction, Dyspnoea/ breathlessness, Exacerbation)								
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.56*	Serious[11]	Very serious[7]	Not serious	Very serious[13]	Very low
1 (Marin 2013 [Requena I cohort])	Prospective cohort	174	0.75*					
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.64*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.62*					
1 (Marin 2013 [Seville cohort])	Prospective cohort	596	0.5*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.65*					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Chen 2015b)	Prospective cohort	354	0.67 (0.6, 0.74)					
1 (Soler-Cataluna 2009)	Prospective cohort	185	0.74 (0.65, 0.83)					
BODE and COTE (Copd cO-morbidity TEst) combined								
1 (Divo 2012)	Prospective cohort	1,659	0.79*	Not serious	N/A	Not serious	Not serious	High
e- BODE index (exacerbations, BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.61*	Serious[2]	Serious[10]	Not serious	Serious[12]	Very low
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.67*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.63*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.69*					
1 (Soler-Cataluna 2009)	Prospective cohort	185	0.77 (0.67, 0.86)					
BOSA index (BMI, Obstruction, SGRQ and Age)								
1 (Chan 2017)	Prospective cohort	772	0.69 (0.64, 0.74)	Serious[3]	N/A	Not serious	Serious[4]	Low
CAT (COPD Assessment Test)								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Casanova 2015)	Prospective cohort	768	0.59*	Not serious	N/A	Not serious	Not serious	High
CCQ (Clinical COPD Questionnaire)								
1 (Casanova 2015)	Prospective cohort	768	0.59*	Not serious	N/A	Not serious	Not serious	High
Clinical basic model (age-adjusted Charlson comorbidity score, sex, FEV1 % predicted and smoking status)								
1 (Stolz 2014 [validation cohort])	Prospective cohort	243	0.72*	Not serious	N/A	Not serious	Very serious[5]	Low
CPI (COPD Prognostic Index)								
1 (Ou 2014)	Prospective cohort	594	0.72*	Not serious	N/A	Not serious	Not serious	High
DOSE index (Dyspnoea/ breathlessness, Obstruction, Smoking status and prior exacerbation history)								
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.58*	Serious[3]	Very serious[7]	Not serious	Serious[8]	Very low
1 (Marin 2013 [Requena I cohort])	Prospective cohort	174	0.75*					
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.63*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.61*					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Marin 2013 [Seville cohort])	Prospective cohort	596	0.5*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.64*					
FEV1 % predicted								
1 (Imfeld 2006)	Prospective cohort	186	0.63*	Serious[2]	Serious[10]	Not serious	Serious[13]	Very low
1 (Casanova 2005)	Prospective cohort	689	0.69 (0.64, 0.73)					
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.56*					
1 (Marin 2013 [Pamplona cohort])	Prospective cohort	190	0.61*					
1 (Marin 2013 [Requena I cohort])	Prospective cohort	174	0.67*					
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.59*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.68*					
1 (Marin 2013 [Seville cohort])	Prospective cohort	596	0.5*					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	137	0.61*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	1,150	0.62*					
GOLD stage 2-4								
1 (Goossens 2014)	Prospective cohort	5,630	0.69*	Not serious	N/A	Not serious	Not serious	High
GOLD 2007								
1 (Ansari 2016)	Prospective cohort	458	0.56 (0.52, 0.61)	Serious[2]	Very serious[7]	Not serious	Serious[1]	Very low
1 (Chan 2016)	Prospective cohort	1,110	0.70 (0.70, 0.70)					
1 (Chen 2015a)	Prospective cohort	471	0.61 (0.55, 0.68)					
1 (Johannessen 2013)	Prospective cohort[6]	912	0.81*					
GOLD stage A to D								
1 (Ansari 2016)	Prospective cohort	458	0.52 (0.47, 0.56)	Not serious	Serious[10]	Not serious	Not serious	Moderate
1 (Goossens 2014)	Prospective cohort	5, 630	0.67[1]					
GOLD 2011								
1 (Chan 2016)	Prospective cohort	1,110	0.71 (0.71, 0.71)	Serious[2]	Very serious[7]	Not serious	Serious[1]	Very low

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Chan 2017)	Prospective cohort	772	0.64 (0.59, 0.69)					
1 (de Torres 2014)	Prospective cohort	707	0.59 (0.50, 0.68)					
1 (Johannessen 2013)	Prospective cohort[6]	912	0.81[1]					
GOLD 2013								
1 (Chen 2015a)	Prospective cohort	471	0.66 (0.60, 0.72)	Not serious	N/A	Not serious	Serious[4]	Moderate
HADO index (Health status, Activity, Dyspnoea/ breathlessness and Obstruction)								
1 (Esteban 2011)	Prospective cohort	348	0.70 (0.64, 0.76)	Not serious	Serious[10]	Not serious	Not serious	Moderate
1 (Esteban 2006)	Prospective cohort	611	0.68*					
HADO-AH index (Health status, Activity, Dyspnoea/ breathlessness, Obstruction, Age and Hospitalisation)								
1 (Esteban 2011)	Prospective cohort	348	0.76 (0.71, 0.81)	Not serious	N/A	Not serious	Serious[4]	Moderate
mBODE% (BMI, Obstruction, Dyspnoea/ breathlessness, oxygen uptake measured at peak exercise (V'O2))								
1 (Cote 2008)	Prospective cohort	444	0.72 (0.66, 0.78)	Not serious	N/A	Not serious	Serious[4]	Moderate
SAFE index (quality of life by SGRQ, Obstruction and 6MWD)								
1 (Marin 2013 [Galdakao cohort])	Prospective cohort	543	0.63*	Serious[3]	Not serious	Not serious	Serious[8]	Low

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Marin 2013 [Requena II cohort])	Prospective cohort	186	0.63*					
1 (Marin 2013 [Tenerife cohort])	Prospective cohort	275	0.62*					
1 (Marin 2013 [Zaragoza I cohort])	Prospective cohort	137	0.62*					

* 95% confidence interval not provided or calculable

1. For studies without 95% CI, the mean study population > 500; in cases with a 95% CI, >33% of population from studies where the CI spans 2 categories of test effectiveness.
2. > 33% of population came from studies at moderate risk of bias
3. Individual study or studies at moderate risk of bias
4. 95% CI spans 2 categories of test effectiveness
5. Sample size < 250
6. Data extracted for model adjusted for gender, age, smoking, BMI, comorbidities [diabetes, heart attack/angina, high blood pressure]
7. The range of effect point estimate values span 3 categories of test effectiveness
8. Average study population < 500, but >250
9. Individual study or studies at high risk of bias
10. The range of effect point estimate values span 2 categories of test effectiveness
11. >33% of study population came from studies at moderate or high risk of bias
12. For studies without 95% CI, the mean study population < 500, but >250, and for studies with 95% CI this spans ≥ 3 categories of test effectiveness
13. For studies without 95% CI, the mean study population < 500, but >250, and for studies with 95% CI this spans ≥ 2 categories of test effectiveness

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BOD index (BMI, Obstruction, Dyspnoea/ breathlessness)								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Celli 2004 [model II])[1]	Prospective cohort	625	HR 1.32 (1.23, 1.40)	Not serious	Serious[16]	Not serious	Serious[21]	Low
1 (de Torres 2008[2])	Prospective cohort	203	HR 1.40 (1.22, 1.61)					
1 (Thabut 2014[8])	Prospective cohort	191	HR 1.52 (1.14, 2.00)					
1 (Soler-Cataluna 2009[18])	Prospective cohort	185	Quartile 1: Reference Quartile 2: HR 1.15 (0.48, 2.76) Quartile 3: HR 2.32 (0.98, 5.50) Quartile 4: HR 4.30 (1.72, 10.75)					
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Andrianopoulos 2015[19])	Prospective cohort	2,010	BODE < 4: Reference BODE ≥4 HR: 1.47 (0.96, 2.24)	Not serious	N/A	Not serious	Serious[21]	Moderate
eBODE index (exacerbations, BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Soler-Cataluna 2009[17])	Prospective cohort	185	Quartile 1: Reference Quartile 2: HR 1.59 (0.56, 4.4) Quartile 3: HR 3.22 (1.22, 8.48) Quartile 4: HR 9.71 (3.36, 28.1)	Not serious	N/A	Not serious	Not serious	High
BODEx (BMI, Obstruction, Dyspnoea/ breathlessness, Exacerbations)								
1 (Soler-Catluna 2009[17])	Prospective cohort	185	Quartile 1: Reference Quartile 2: HR 1.52 (0.66, 3.53) Quartile 3: HR 3.16 (1.37, 7.30) Quartile 4: HR 5.86 (2.42, 14.17)	Not serious	N/A	Not serious	Not serious	High

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BOSA (BMI, Obstruction, SGRQ and Age)								
1 (Chan 2017[7])	Prospective cohort	772	Group 1 (score 0-3): HR 0.29 (0.16, 0.51) Group 2 (score 4-5): HR 0.53 (0.53, 0.82) Group 3 (score 6): HR 0.94 (0.57, 1.55) Group 4 (score 7-12): Reference	Serious[12]	N/A	Not serious	Not serious	Moderate
i-BODE index (BODE plus incremental shuttle walking test [ISWT])								
1 (Moberg 2014[10])	Prospective cohort	674	HR 1.28 (1.20, 1.37)	Not serious	N/A	Not serious	Serious[13]	Moderate
CCQ (Clinical COPD Questionnaire)								
1 (Sundh 2012a[4])	Prospective cohort	970	<1: Reference ≥1, <2: HR 0.98 (0.57, 1.66) ≥ 2, <3: HR 1.96 (1.21, 3.18) ≥3: HR 3.13 (1.98, 4.95)	Not serious	N/A	Not serious	Not serious	High
DOSE index (Dyspnoea/ breathlessness, Obstruction, Smoking status and prior exacerbation history)								
1 (Sundh 2012b[4])	Prospective cohort	1,111	0-3: Reference 4-5: HR 3.48 (2.32, 5.22) 6-7: HR 8.00 (4.67, 13.70)	Serious[12]	N/A	Not serious	Not serious	Moderate
GOLD before 2011 (Stages 1-4)								
1 (Leivseth 2013[3])	Prospective cohort	424 883 204 29	Stage 1: Reference Stage 2: HR 1.67 (1.39, 2.01) Stage 3: HR 2.88 (2.30, 3.62) Stage 4: HR 4.85 (3.16, 7.44)	Not serious	Serious[16]	Serious[14]	Not serious	Low
1 (Chan 2016[5] [GOLD 2007])	Prospective cohort	1,110	Stage 1: HR 0.00 (0.00, 0.00)[6] Stage 2: HR 0.53 (0.35, 0.79) Stage 3: HR 0.75 (0.53, 1.07) Stage 4: HR Reference					

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Johannessen 2013[9] [GOLD 2007)	Prospective cohort	912	Stage 2: Reference Stage 3: HR 1.70 (1.30, 2.20) Stage 4: HR 3.70 (2.50, 5.00)					
1 (Mattila 2015[11])	Prospective cohort	6636	No COPD: Reference Stage 1: HR 1.27 (1.06, 1.51) Stage 2: HR 1.40 (1.21, 1.63) Stage 3: HR 1.55 (1.21, 1.97) Stage 4: HR 2.85 (1.65, 4.94)					
GOLD 2011 Groups A-D								
1 (Leivseth 2013[3])	Prospective cohort	731 216 142 115	Group A: Reference Group B: HR 0.96 (0.78, 1.19) Group C: HR 1.38 (1.10, 1.74) Group D: HR 1.92 (1.51, 2.45)	Serious[15]	Serious[16]	Not serious	Not serious	Low
1 (Chan 2016[5])	Prospective cohort	1,110	Group A: HR 0.31 (0.17, 0.56) Group B: HR 0.77 (0.52, 1.14) Group C: HR 0.65 (0.45, 0.95) Group D: Reference					
1 (Johannessen 2013[9])	Prospective cohort	912	Group A: Reference Group B: HR 1.40 (0.90, 2.20) Group C: HR 0.90 (0.40 , 2.00) Group D: HR 2.90 (1.90, 4.40)					
HADO index (Health status, Activity, Dyspnoea/ breathlessness and Obstruction)								
1 (Esteban 2006[20])	Prospective cohort	611	Mild: HR 0.25 (0.13, 0.47) Moderate: HR 0.34 (0.22, 0.53) Severe: Reference	Not serious	N/A	Not serious	Not serious	High

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality

1. Model adjusted for comorbidities using the Charlson index.
2. Model adjusted for age, gender, pack years, cardiovascular risk or disease, treatment with inhaled corticosteroids.
3. Data pooled for women and men; model adjusted for age, smoking status and educational attainment.
4. Model adjusted for age, sex, heart disease
5. Paper states that model is adjusted for confounding variables, but does not specify them.
6. No events
7. Model adjusted for race and gender
8. Model adjusted for study centre and augmentation therapy
9. Model adjusted for gender, age, smoking, BMI, comorbidities (diabetes, heart attack/angina, high blood pressure).
10. Model adjusted for age, sex, pack-years, current smoking, oxygen saturation at rest, desaturation >4% during SWT, maintenance prednisolone, LTOT.
11. Model adjusted for age, sex, smoking.
12. Study at moderate risk of bias.
13. Study or studies have 95% CI that crosses one side of a defined MID.
14. > 33% of the population came from studies with partial applicability to the research question.
15. > 33% of the population came from studies with moderate risk of bias.
16. Studies are incomparable as they use different reference standards to each other or mix HR with references and those without.
17. Model adjusted for age, co-morbidities, blood gases.
18. Model adjusted for age, co-morbidities, blood gases and history of acute exacerbations of COPD.
19. Model adjusted for age, sex, BMI, FEV1, FEV1/FVC ratio, SGRQ, emphysema, and smoking.
20. Model adjusted for age and smoking.
21. Studies with a single HR that have a 95% CI that crosses one side of a defined MID and (for the studies with a reference category) the most extreme category has a 95% CI that crosses one end of defined MD in > 33% of studies by population.

Mortality due to respiratory causes

c- statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1 (Celli 2004)	Prospective cohort	625	0.74*	Not serious	Serious[4]	Not serious	Serious[1]	Low
1 (Esteban 2010)	Prospective cohort	453	0.87 (0.82. 0.93)					
FEV1 % predicted								
1 (Celli 2004)	Prospective cohort	625	0.65*	Not serious	N/A	Not serious	Not serious	High
GOLD 2007								
1 (Chen 2015a)	Prospective cohort	471	0.65 (0.57, 0.73)	Not serious	Very serious[5]	Not serious	Serious[6]	Very low
1 (Johannessen 2013)	Prospective cohort[2]	912	0.83*					
GOLD 2011								
1 (Johannessen 2013)	Prospective cohort[2]	912	0.82*	Not serious	N/A	Not serious	Not serious	High
GOLD 2013								
1 (Chen 2015a)	Prospective cohort	471	0.71 (0.64, 0.77)	Not serious	N/A	Not serious	Serious[3]	Moderate
HADO index (Health status, Activity, Dyspnoea/ breathlessness and Obstruction)								
1 (Esteban 2010)	Prospective cohort	453	0.86 (0.81, 0.91)	Not serious	N/A	Not serious	Serious[3]	Moderate

* 95% confidence interval not provided or calculable

1. 95% CI spans 2 categories of test effectiveness.
2. Data extracted for model adjusted for gender, age, smoking, BMI, comorbidities [diabetes, heart attack/angina, high blood pressure]
3. 95% CI spans 2 categories of test effectiveness
4. The range of effect point estimate values span 2 categories of test effectiveness
5. The range of effect point estimate values span 3 categories of test effectiveness

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
6. 95% CI spans 3 categories of test effectiveness and Johannessen study population is > 500.								

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Celli 2004 [model II])[1]	Prospective cohort	625	HR 1.63 (1.48, 1.8)	Not serious	N/A	Not serious	Not serious	High
GOLD 2007								
1 (Johannessen 2013[2])	Prospective cohort	912	Stage 2: Reference Stage 3: HR 2.80 (1.70, 4.70) Stage 4: HR 9.30 (5.50, 15.70)	Not serious	Serious[5]	Serious[4]	Not serious	Low
1 (Mattila 2015[3])	Prospective cohort	6636	No COPD: Reference Stage 1: HR 1.81 (1.04, 3.16) Stage 2: HR 2.92 (1.93, 4.40) Stage 3: HR 4.95 (2.94, 8.35) Stage 4: HR 15.95 (5.77, 44.11)					
GOLD 2011								
1 (Johannessen 2013[2])	Prospective cohort	912	Group A: Reference Group B: HR 2.10 (0.70, 6.40) Group C: HR 1.20 (0.20, 6.40) Group D: HR 9.40 (3.40, 25.75)	Not serious	N/A	Not serious	Not serious	High

1. Model adjusted for comorbidities using the Charlson index.
2. Model adjusted for gender, age, smoking, BMI, comorbidities (diabetes, heart attack/angina, high blood pressure).
3. Model adjusted for age, sex, smoking.

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality

4. >33% of the population came from studies that were partially applicable to the research question.
5. Studies are incomparable as they use different reference standards to each other.

All-cause hospitalisations

Sensitivity, specificity and likelihood ratios

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE ≥3										
1 (Andrianopoulos 2015)	Prospective	2,010	57.0 (53.0, 61.0)	69.0 (67.0, 72.0)	LR+ 1.84 (1.66, 2.04)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.62 (0.56, 0.68)	Not serious	N/A	Not serious	Serious[1]	Moderate
CAT ≥10[2]										
1 (Suetomo 2014)	Prospective	123	80.0 (53.0, 93.4)	51.9 (42.5, 61.1)	LR+ 1.66 (1.20, 2.28)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.38 (0.13, 1.07)	Not serious	N/A	Not serious	Serious[1]	Moderate
GOLD stages III and IV										
1 (Suetomo 2014)	Prospective	123	60.0 (34.8, 80.8)	74.1 (65.0, 81.5)	LR+ 2.31 (1.37, 3.90)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.54 (0.28, 1.01)	Not serious	N/A	Not serious	Serious[1]	Moderate

1. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)
2. Suetomo 2014 reported the 'best' sensitivity and specificity for hospitalisations with a cut-off CAT score of 29 points with extractable data only for the cut-off of ≥10 points. Therefore, we calculated sensitivity, specificity and likelihood ratios using CAT ≥10

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable.

c- statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index ≥ 3 (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Andrianopoulos 2015)	Prospective cohort	2,010	0.69*	Not serious	N/A	Not serious	Not serious	High
GOLD 2007								
1 (Johannessen 2013)	Prospective cohort[1]	912	0.76*	Not serious	N/A	Not serious	Not serious	High
GOLD 2011								
1 (Johannessen 2013)	Prospective cohort[1]	912	0.77*	Not serious	N/A	Not serious	Not serious	High

*95% confidence interval not provided or calculable

1. Data extracted for model adjusted for gender, age, smoking, BMI, comorbidities [diabetes, heart attack/angina, high blood pressure]

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
COPD severity score (per 0.5 SD increment in COPD Severity Score)								
1 (Eisner 2010[1])	Prospective cohort	1,202	HR 1.59 (1.44, 1.75)	Very serious[4]	N/A	Not serious	Not serious	Low
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Andrianopoulos 2015[5])	Prospective cohort	2, 010	BODE< 3: Reference ≥3: HR 1.40 (1.08, 1.82)	Not serious	N/A	Not serious	Serious[6]	Moderate
i-BODE (BODE plus incremental shuttle walking test [ISWT])								
1 (Moberg 2014[3])	Prospective cohort	674	HR 1.21 (1.14, 1.28)	Not serious	N/A	Not serious	Serious[6]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
GOLD 2007								
1 (Johannessen 2013[2])	Prospective cohort	912	Stage 2: Reference Stage 3: HR 1.20 (0.90, 1.50) Stage 4: HR 2.00 (1.50, 2.60)	Not serious	N/A	Not serious	Not serious	High
GOLD 2011								
1 (Johannessen 2013[2])	Prospective cohort	912	Group A: Reference Group B: HR 1.60 (1.20, 2.00) Group C: HR 1.50 (1.00, 2.40) Group C: HR 2.00 (1.5, 2.50)	Not serious	N/A	Not serious	Not serious	High

1. Model adjusted for age, sex, race, and educational attainment.
2. Model adjusted for gender, age, smoking, BMI, comorbidities (diabetes, heart attack/angina, high blood pressure).
3. Model adjusted for age, sex, pack-years, oxygen saturation at rest, desaturation >4% during SWT, LTOT.
4. Study at high risk of bias.
5. Model adjusted for age, sex, BMI, FEV1, FEV1/FVC ratio, SGRQ, emphysema, and smoking.
6. 95% CI crosses one side of a MID (0.8, 1.25).

Respiratory specific hospitalisations

c- statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Model 1 (age, race, educational attainment, tobacco history, and medical comorbidities (heart failure, coronary artery disease, diabetes, and sleep apnoea)								
1 (Omachi 2008)	Prospective cohort	267	0.79*	Not serious	N/A	Not serious	Serious[1]	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Model 2 (model one plus COPD severity score)								
1 (Omachi 2008)	Prospective cohort	267	0.91*	Not serious	N/A	Not serious	Serious[1]	Moderate
BODEX index (BMI, Obstruction, Dyspnoea/ breathlessness, Exacerbation)								
1 (Moy 2014)	Prospective cohort	167	0.68*	Not serious	N/A	Not serious	Very serious[2]	Low

* 95% confidence interval not provided or calculable

1. Sample size < 500, but >250
2. Sample size < 250

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
i-BODE (BODE plus incremental shuttle walking test [ISWT])								
1 (Moberg 2014[1])	Prospective cohort	674	HR 1.15 (1.09, 1.20)	Not serious	N/A	Not serious	Not serious	High
GOLD 2007								
1 (Johannessen 2013[2])	Prospective cohort	912	Stage 2: Reference Stage 3: HR 3.60 (2.30, 5.80) Stage 4: HR 8.30 (5.00, 13.80)	Not serious	N/A	Not serious	Not serious	High
GOLD 2011								
1 (Johannessen 2013[2])	Prospective cohort	912	Group A: Reference Group B: HR 1.70 (0.70, 4.20) Group C: HR 2.20 (0.70, 6.90) Group C: HR 8.20 (3.70, 17.80)	Not serious	N/A	Not serious	Not serious	High

1. Model adjusted for age, sex, pack-years, current smoking, oxygen saturation at rest, desaturation >4% during SWT, maintenance prednisolone, LTOT.

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
2. Model adjusted for gender, age, smoking, BMI, comorbidities (diabetes, heart attack/angina, high blood pressure)								

Exacerbations

Sensitivity, specificity and likelihood ratios.

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index >1.9 (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)										
1 (Marin 2009)	Prospective	275	70.8 (64.8, 76.2)	77.1 (60.5, 88.1)	LR+ 3.09 (1.67, 5.72)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.37 (0.28, 0.49)	Not serious	N/A	Not serious	Not serious	High
BODE class II (stages 3 to 4) – exacerbations										
1 (Faganello 2010)	Prospective	120	48.3 (36.1, 60.8)	73.3 (60.8, 83.0)	LR+ 1.81 (1.10, 2.97)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.70 (0.52, 0.94)	Not serious	N/A	Not serious	Serious[1]	Moderate
CAT (COPD Assessment Test) cut-off 17/40 – moderate to severe exacerbations										
1 (Lee 2014)	Prospective	495	52.2 (45.7, 58.7)	68.7 (62.8, 74.0)	LR+ 1.66 (1.34, 2.07)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.69 (0.59, 0.81)	Not serious	N/A	Not serious	Serious[1]	Moderate
CAT (COPD Assessment Test) ≥10										
1 (Suetomo 2014)[2]	Prospective	123	72.4 (59.6, 82.4)	66.2 (53.9, 76.6)	LR+ 2.13 (1.47, 3.11)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.41 (0.26, 0.65)	Not serious	N/A	Not serious	Serious[1]	Moderate

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Effect size (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
GOLD stages III and IV										
1 (Suetomo 2014)	Prospective	123	48.3 (35.8, 61.0)	86.2 (75.5, 92.6)	LR+ 3.48 (1.79, 6.76)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.60 (0.46, 0.78)	Not serious	N/A	Not serious	Serious[1]	Moderate
GOLD 2003 stage III – exacerbations										
1 (Faganello 2010)	Prospective	120	58.3 (45.6, 70.1)	73.3 (60.8, 83.0)	LR+ 2.18 (1.36, 3.50)	Not serious	N/A	Not serious	Serious[1]	Moderate
					LR- 0.56 (0.40, 0.79)	Not serious	N/A	Not serious	Serious[1]	Moderate

1. 95% confidence interval for likelihood ratio crosses one end of a defined MID interval – (0.5, 2)
2. Suetomo 2014 reported the 'best' sensitivity and specificity for moderate or severe exacerbations with a cut-off CAT score of 8 points with extractable data only for the cut-off of ≥10 points. Therefore, we calculated sensitivity, specificity and likelihood ratios using CAT ≥10

CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; CAT: COPD assessment tool; BODE: dyspnoea/ breathlessness, body mass index, FEV1 and exercise;

c- statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
ADO (Age, Dyspnoea/ breathlessness, Obstruction)								
1 (Motegi 2013)	Prospective cohort	183	0.64 (0.56, 0.73)	Not serious	N/A	Not serious	Very serious[2]	Low
BOD index (BMI, Obstruction, Dyspnoea/ breathlessness)								
1 (Chan 2016)	Prospective cohort	1,110	0.61 (0.61, 0.61)	Serious[1]	N/A	Not serious	Not serious	Moderate

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise) **Stage 3-4**								
1 (Fagenello 2010)	Prospective cohort	120	0.62*	Not serious	N/A	Not serious	Very serious[3]	Low
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Marin 2009)	Prospective cohort	275	0.81 (0.75, 0.87)	Not serious	Very serious[7]	Not serious	Very serious[9]	Very low
1 (Moy 2014)	Prospective cohort	167	0.62*					
1 (Motegi 2013)	Prospective cohort	183	0.65 (0.56, 0.73)					
BODEx ((BMI, Obstruction, Dyspnoea/ breathlessness, Exacerbations)								
1 (Chen 2015b)	Prospective cohort	262	0.73 (0.67, 0.79)	Serious[1]	N/A	Not serious	Serious[4]	Low
CAT (COPD Assessment Test)								
1 (Lee 2014)	Prospective cohort	495	0.79 (0.75, 0.84)	Not serious	Serious[6]	Not serious	Serious[8]	Low
1 (Pothirat 2015)	Prospective cohort	140	0.89 (0.84, 0.94)					
1 (Varol 2014)	Prospective cohort	165	0.78 (0.71, 0.85)					
DOSE (Dyspnoea/ breathlessness, Obstruction, Smoking, Exacerbation)								
1 (Motegi 2013)	Prospective cohort	183	0.75 (0.67, 0.82)	Not serious	N/A	Not serious	Very serious[2]	Low
GOLD 2003 stage III								
1 (Fagenello 2010)	Prospective cohort	120	0.69*	Not serious	N/A	Not serious	Very serious[3]	Low

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
GOLD stage 1-4								
1 (Motegi 2013)	Prospective cohort	183	0.66 (0.57, 0.74)	Not serious	N/A	Not serious	Very serious[2]	Low
GOLD 2007								
1 (Chan 2016)	Prospective cohort	1,110	0.59 (0.59, 0.59)	Serious[5]	Serious[6]	Not serious	Not serious	Low
1 (Chen 2015a)	Prospective cohort	338	0.67 (0.61, 0.73)					
GOLD 2011								
1 (Chan 2016)	Prospective cohort	1,1110	0.62 (0.62, 0.62)	Serious[1]	N/A	Not serious	Not serious	Moderate
GOLD 2013								
1 (Chen 2015a)	Prospective cohort	338	0.78 (0.73, 0.83)	Serious[1]	N/A	Not serious	Serious[4]	Low

* 95% confidence interval not provided or calculable

1. Individual study at moderate risk of bias
2. 95% CI spans 3 categories of test effectiveness
3. Sample size < 250
4. 95% CI spans 2 categories of test effectiveness
5. >33% of study population came from studies at moderate risk of bias
6. The range of effect point estimate values span 2 categories of test effectiveness
7. The range of effect point estimate values span 3 categories of test effectiveness
8. >33% of the studies have 95% CI that cross 2 categories of test effectiveness
9. Moy study population is < 250, and Marin and Motegi 95% CI cross ≥2 categories of test effectiveness. Averaged as very serious.

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BOD index (quartiles)								
1 (Chan 2016[1])	Prospective cohort	1,110	1: HR 0.53 (0.4, 0.7) 2: HR 0.6 (0.45, 0.8) 3: HR 1.02 (0.78, 1.33) 4: Reference	Serious[3]	N/A	Not serious	Not serious	Moderate
CAT								
1 (Lee 2014[2])	Prospective cohort	495	CAT 0-9:Reference CAT 10-19: HR 1.3 (1.09, 1.56) CAT 20-29: HR 1.37 (1.14, 1.65) CAT 30-40: HR 1.5 (1.24, 1.81)	Not serious	N/A	Not serious	Serious[4]	Moderate
GOLD 2007								
1 (Chan 2016[1])	Prospective cohort	1,110	Stage 1: HR 0.21 (0.08, 0.59) Stage 2: HR 0.65 (0.48, 0.89) Stage 3: HR 0.91 (0.69, 1.21) Stage 4: HR Reference	Serious[3]	N/A	Not serious	Not serious	Moderate
GOLD 2011 (Groups A-D)								
1 (Chan 2016[1])	Prospective cohort	1,110	Group A: HR 0.39 (0.28, 0.55) Group B: HR 0.67 (0.50, 0.90) Group C: HR 0.71 (0.56, 0.92) Group D: Reference	Serious[3]	N/A	Not serious	Not serious	Moderate

1. Study states that model is adjusted for confounding variables but does not specify them.
2. Model adjusted for age, gender, body mass index (BMI), GOLD stage (1-4), number of COPD exacerbations in the previous year, duration of COPD, current smoking status, number of comorbidities, history of influenza vaccination and country.
3. Study at moderate risk of bias.
4. Study has a 95% CI that crosses one side of a defined MID.

Severe exacerbations

c- statistics

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
BODE index (BMI, Obstruction, Dyspnoea/ breathlessness, Exercise)								
1 (Marin 2009)	Prospective cohort	275	0.88 (0.83, 0.92)	Not serious	N/A	Not serious	Serious[2]	Moderate
CAT (COPD Assessment Test)								
1 (Lee 2014)	Prospective cohort	495	0.72 (0.68, 0.77)	Not serious	N/A	Not serious	Serious[2]	Moderate
GOLD 2007								
1 (Chen 2015a)	Prospective cohort	338	0.66 (0.60, 0.72)	Serious[1]	N/A	Not serious	Serious[2]	Low
GOLD 2013								
1 (Chen 2015a)	Prospective cohort	338	0.78 (0.73, 0.83)	Serious[1]	N/A	Not serious	Serious[2]	Low

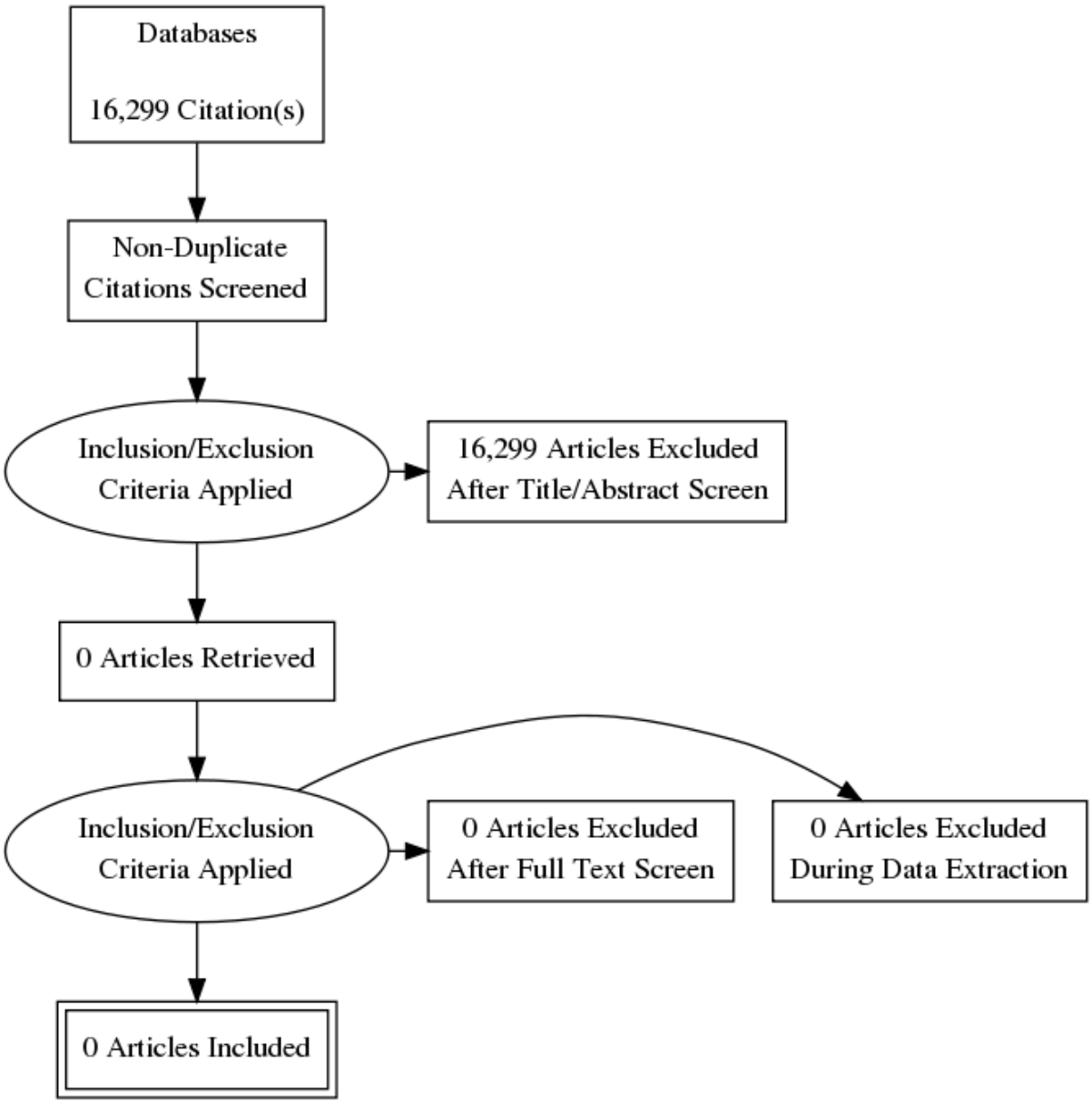
1. Individual study at moderate risk of bias
2. 95% CI spans 2 categories of test effectiveness

Hazard ratios

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
CAT (COPD Assessment Test)								
1 (Lee 2014[1])	Prospective cohort	495	CAT 0-9:Reference CAT 10-19: HR 1.18 (0.89, 1.57) CAT 20-29: HR 1.4 (1.03, 1.9) CAT 30-40: HR 2.01 (1.45, 2.80)	Not serious	N/A	Not serious	Not serious	High

No. of studies	Study design	Sample size	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1. Model adjusted for age, gender, body mass index (BMI), GOLD stage (1-4), number of COPD exacerbations in the previous year, duration of COPD, current smoking status, number of comorbidities, history of influenza vaccination and country.								

Appendix H – Economic evidence study selection



Appendix I – Excluded studies

Clinical studies

Author (year)	Title	Reasons for exclusion
Abascal-Bolado (2015)	Forecasting COPD hospitalization in the clinic: optimizing the chronic respiratory questionnaire	• Not a relevant study design *Not a prospective cohort study*
Almagro (2014)	Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study	• Study population is mixed and data for people with stable COPD cannot be extracted *Study contains people with stable COPD and those recruited following an exacerbation.*
Al-Mutairi (2007)	Impulse oscillometry: an alternative modality to the conventional pulmonary function test to categorise obstructive pulmonary disorders	• Reference standard in study does not match that specified in protocol
Arghir (2011)	The use of FEV1-body mass index corellation in evaluating the severity and prognosis of COPD in active workers patients	• Conference abstract
Baughman (2012)	Combined effect of lung function level and decline increases morbidity and mortality risks	• Study does not include any relevant prognostic variables *Prognostic variable was not multidimensional*
Bhatt (2013)	Comparison of spirometric thresholds in diagnosing smoking-related airflow obstruction	• Does not contain a population of people with suspected COPD
Bhatt (2014)	FEV(1)/FEV(6) to diagnose airflow obstruction. Comparisons with computed tomography and morbidity indices	• Study does not include any relevant prognostic variables
Blumenthal (2016)	Biobehavioral Prognostic Factors in Chronic Obstructive Pulmonary Disease: Results From the INSPIRE-II Trial	• Study does not contain any relevant index tests
Boeck (2015)	Prognostic assessment in chronic obstructive pulmonary disease using copeptin: A simplified risk index	• Conference abstract
Boeck (2016)	Prognostic assessment in COPD without lung function: the B-AE-D indices	• Study population is mixed and data for people with stable COPD cannot be extracted *The multiple cohorts included in this study recruited stable and*

		exacerbated patients.
Boutou (2013)	Lung function indices for predicting mortality in COPD	• Retrospective prognostic cohort study
Boutou (2014)	A combined pulmonary function and emphysema score prognostic index for staging in Chronic Obstructive Pulmonary Disease	• Retrospective prognostic cohort study
Broekhuizen (2011)	Does a decision aid help physicians to detect chronic obstructive pulmonary disease?	• Study does not contain any relevant index tests *Study does not include a diagnostic test that could be used to confirm diagnosis of COPD.*
Brusse-Keizer (2015)	Adrenomedullin optimises mortality prediction in COPD patients	• Study population is mixed and data for people with stable COPD cannot be extracted *Mixed population of people with stable COPD and those recruited following an exacerbation.*
Brusse-Keizer (2017)	Comparing the 2007 and 2011 GOLD Classifications as Predictors of all-Cause Mortality and Morbidity in COPD	• Study population is mixed and data for people with stable COPD cannot be extracted *The COMIC COPD cohort consists of people with stable COPD and those recruited during an exacerbation. Data is presented for the groups together.*
Celli (2012)	Inflammatory biomarkers improve clinical prediction of mortality in chronic obstructive pulmonary disease	• Study investigates the prognostic value of biomarkers
Chan (2015)	Prognostication in patients with chronic obstructive pulmonary disease using the BOS index	• Conference abstract
Chang (2014)	Utility of the combination of serum highly-sensitive C-reactive protein level at discharge and a risk index in predicting readmission for acute exacerbation of COPD	• Does not contain a population of people with stable COPD *Participants are recruited from hospital following an acute exacerbation of COPD.*
Chen (2012)	Using post-bronchodilator FEV1 is better than pre-bronchodilator FEV1 in evaluation of COPD severity	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)

Chen (2016)	Importance of fractional exhaled nitric oxide in the differentiation of asthma-COPD overlap syndrome, asthma, and COPD	• Not a relevant study design *Retrospective study*
Dal Negro (2014)	Sensitivity of the COPD assessment test (CAT questionnaire) investigated in a population of 681 consecutive patients referring to a lung clinic: the first Italian specific study	• Study does not contain any relevant outcomes (mortality, hospitalisations, exacerbations, change in FEV1)
Dal Negro (2017)	Patient Related Outcomes-BODE (PRO-BODE): A composite index incorporating health utilization resources predicts mortality and economic cost of COPD in real life.	•Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Dawkins (2003)	Predictors of mortality in alpha1-antitrypsin deficiency	• Does not contain a population of people with stable COPD *Participants have patients with a1-antitrypsin deficiency, but are not recruited based on a diagnosis of COPD.*
Dawkins (2009)	Mortality in alpha-1-antitrypsin deficiency in the United Kingdom	• Does not contain a population of people with stable COPD *Participants have patients with a1-antitrypsin deficiency, but are not recruited based on a diagnosis of COPD.*
DeVries (2016)	Validation of the breathlessness, cough and sputum scale to predict COPD exacerbation	• Does not contain a population of people with stable COPD *Participants were recruited during contact with medical services for an exacerbation.*
Dijk (2011)	Multidimensional prognostic indices for use in COPD patient care. A systematic review	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Eisner (2005)	Development and validation of a survey-based COPD severity score	• Study does not include any relevant measures (e.g. HR, c-statistic etc.) *Paper present OR and uses univariate models*
Eriksson (2011)	BMI and risk for death in COPD	• Conference abstract

Esteban (2008)	Predictors of mortality in patients with stable COPD	• Study does not include any relevant prognostic variables
Feliz-Rodriguez (2013)	Evolution of the COPD Assessment Test score during chronic obstructive pulmonary disease exacerbations: determinants and prognostic value	• Does not contain a population of people with stable COPD
Fragoso (2011)	Staging the severity of chronic obstructive pulmonary disease in older persons based on spirometric Z-scores	• Does not contain a population of people with stable COPD *Participants do not all have a diagnosis of COPD at baseline.* • Study does not contain any relevant index tests *Study is examining a different method of using spirometry measures to predict outcomes.*
Gedebjerg (2018)	Prediction of mortality in patients with chronic obstructive pulmonary disease with the new Global Initiative for Chronic Obstructive Lung Disease 2017 classification: a cohort study.	• Not a relevant study design *Retrospective cohort study*
Ghobadi (2012)	The Relationship between COPD Assessment Test (CAT) Scores and Severity of Airflow Obstruction in Stable COPD Patients	• Study does not contain any relevant outcomes (mortality, hospitalisations, exacerbations, change in FEV1)
Guirguis-Blake (2016)	Screening for Chronic Obstructive Pulmonary Disease: Evidence Report and Systematic Review for the US Preventive Services Task Force	• Systematic review does not contain any included studies *Review was focused on assessing the benefit of screening asymptomatic adults for COPD.*
Gupta (2014)	The COPD assessment test: a systematic review	• Systematic review used as a source of RCTs, but not for data extraction
Haroon (2014)	Screening test accuracy of case finding interventions for chronic obstructive pulmonary disease in primary care: A systematic review	• Conference abstract
Haroon (2015)	Effectiveness of case finding strategies for COPD in primary care: A systematic review and meta-analysis	• Systematic review does not contain any included studies *Included studies look at COPD case finding rather than diagnosis of COPD in people presenting with relevant symptoms.*

Haroon (2015)	Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis	• Systematic review does not contain any included studies *Focus on screening for COPD rather than diagnosis in people with suspected COPD.*
Hernandez (2012)	Prognostic factors in COPD patients controlled in two outpatient clinics	• Conference abstract
Horita (2016)	Chronic obstructive pulmonary disease prognostic score: A new index	• Retrospective analysis missing key variables *mMRC dyspnoea/ breathlessness was not measured in the original study, but was estimated from other available data.*
Jing (2009)	Should FEV1/FEV6 replace FEV1/FVC ratio to detect airway obstruction? A metaanalysis	• Systematic review does not contain any included studies
Jones (2016)	Multi-component assessment of chronic obstructive pulmonary disease: an evaluation of the ADO and DOSE indices and the global obstructive lung disease categories in international primary care data sets	• Retrospective prognostic cohort study
Jung (2014)	Clinical features and prognostic factors of patients hospitalized due to acute exacerbation of chronic obstructive pulmonary disease	• Conference abstract • Does not contain a population of people with stable COPD
Kelly (2012)	Health status assessment in routine clinical practice: the chronic obstructive pulmonary disease assessment test score in outpatients.	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Ko (2011)	A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD	• Does not contain a population of people with stable COPD
Kostianev (2008)	Multidimensional system for assessment of COPD patients. Comparison with BODE index	• Full text paper not available
Kwon (2010)	Prognosis of heart failure patients with reduced and preserved ejection fraction and coexistent chronic obstructive pulmonary disease	• Retrospective prognostic cohort study
Labor (2015)	Exhaled breath temperature as a possible early marker in smokers at risk for COPD (MARKO study)	• Conference abstract
Lamprecht (2007)	Six-second spirometry for detection of airway obstruction: a population-based study in Austria	• Does not contain a population of people with stable COPD • Does not contain a population of

		people with suspected COPD
Lange (2012)	Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population	• Retrospective prognostic cohort study
Lederer (2007)	Lung-volume reduction surgery for pulmonary emphysema: Improvement in body mass index, airflow obstruction, dyspnea, and exercise capacity index after 1 year	• Study does not contain any relevant outcomes (mortality, hospitalisations, exacerbations, change in FEV1)
Lee (2015)	Pharmacological treatment response according to the severity of symptoms in patients with chronic obstructive pulmonary disease	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Liu (2011)	High value of combined serum C-reactive protein and BODE score for mortality prediction in patients with stable COPD	• Study does not include any relevant measures (e.g. HR, c-statistic etc.) *Data is reported as OR.*
Lou (2016)	Interaction of Depression and Nicotine Addiction on the Severity of Chronic Obstructive Pulmonary Disease: A Prospective Cohort Study	• Study does not contain any relevant outcomes (mortality, hospitalisations, exacerbations, change in FEV1) • Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Manninen (1988)	Large-image intensifier photofluorography and conventional radiography in pulmonary emphysema. Correlation with computed tomography	• Does not contain a population of people with suspected COPD
Mannino (2006)	Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study	• Does not contain a population of people with suspected COPD
Marsh (2007)	Utility of lung density measurements in the diagnosis of emphysema	• Does not contain a population of people with suspected COPD
Martinez (2017)	Respiratory Symptoms Items from the COPD Assessment Test Identify Ever-Smokers with Preserved Lung Function at Higher Risk for Poor Respiratory Outcomes. An Analysis of the Subpopulations and Intermediate	• Does not contain a population of people with stable COPD *Participants are part of a COPD study cohort, but do not have COPD at baseline or suspected*

	Outcome Measures in COPD Study Cohort	*COPD.*
Mets (2011)	Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans	• Does not contain a population of people with suspected COPD
Miravitlles (2009)	Validation of the COPD severity score for use in primary care: the NEREA study	• Retrospective prognostic cohort study
Mohamed (2012)	Spirometric thresholds for diagnosing COPD: 0.70 or LLN, pre- or post-dilator values?	• Study does not contain any relevant index tests *Spirometry optimisation is out of the scope of this review.*
Morris (2012)	The diagnostic importance of a reduced FEV1/FEV6	• Study does not contain any relevant index tests *Spirometry optimisation is out of the scope of this review.*
Motegi (2011)	Efficacy of prognostic assessment using ado index for Japanese patients with COPD	• Conference abstract
Moya (2014)	Airflow reversibility on long term outcomes of patients with COPD	• Conference abstract
Navarro (2015)	Prognostic assessment of mortality and hospitalizations of outpatients with advanced chronic obstructive pulmonary disease. Usefulness of the CODEX index	• Study not reported in English
Neo (2016)	A pilot study examining predictors of short term mortality in advanced COPD-importance of nutritional, systemic inflammatory and physical performance indices	• Conference abstract
Nishimura (2002)	Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Nizet (2005)	Survival of chronic hypercapnic COPD patients is predicted by smoking habits, comorbidity, and hypoxemia	• Study does not contain any relevant index tests *Study analyses prognostic factors, but does not develop a multidimensional prognostic index.*
Oga (2002)	Health status measured with the CRQ does not predict mortality in COPD	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)

Oga (2011)	Predictive properties of different multidimensional staging systems in patients with chronic obstructive pulmonary disease	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Ong (2005)	A multidimensional grading system (BODE index) as predictor of hospitalization for COPD	• Univariate analysis
Ozgur (2012)	An integrated index combined by dynamic hyperinflation and exercise capacity in the prediction of morbidity and mortality in COPD	• Study does not contain any relevant index tests *Only multidimensional because the index includes a biomarker*
Panjabi (2015)	Usefulness of the COPD assessment test (CAT) in patients with (A) stable disease, and (B) exacerbations	• Conference abstract
Papaioannou (2013)	The impact of depressive symptoms on recovery and outcome of hospitalised COPD exacerbations	• Does not contain a population of people with stable COPD
Papaioannou (2014)	COPD assessment test: a simple tool to evaluate disease severity and response to treatment	• Study does not include any relevant measures (e.g. HR, c-statistic etc.) *Data is presented as OR and correlations (r).*
Pavasini (2015)	Cardiac troponin elevation predicts all-cause mortality in patients with acute exacerbation of chronic obstructive pulmonary disease: Systematic review and meta-analysis	• Systematic review does not contain any included studies *Study participants have acute COPD exacerbations at baseline.*
Percival (2014)	Utility of the copd assessment test (CAT) in evaluating copd severity	• Conference abstract
Pla (2014)	Short and medium term prognosis in patients hospitalized for acute exacerbation of COPD (AECOPD): The codex index	• Conference abstract
Pudney (2016)	Plain chest x-ray (CXR) in the diagnosis of chronic obstructive pulmonary disease (COPD)	• Conference abstract
Puhan (2012)	Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts	• Study population is mixed and data for people with stable COPD cannot be extracted *Participants came from a number of clinical and population-based cohorts, but latter did not have a confirmed diagnosis of COPD at baseline.*

Qu (2017)	Sagittal-lung CT measurements in the evaluation of asthma-COPD overlap syndrome: a distinctive phenotype from COPD alone	• Does not contain a population of people with suspected COPD
Rahman (2014)	Prognostic evaluation of COPD patients using bode index	• Conference abstract
Rozenberg (2012)	Non-spirometric pulmonary function parameters for differentiating COPD and asthma	• Conference abstract
Sanders (1988)	Detection of emphysema with computed tomography. Correlation with pulmonary function tests and chest radiography	• Not a relevant study design *Retrospective study*
Schapira (1993)	The value of the forced expiratory time in the physical diagnosis of obstructive airways disease	• Study population is mixed and data for people with stable COPD cannot be extracted
Schonenberger (2012)	Prediction of mortality in the Swiss chronic obstructive pulmonary disease (COPD) cohort using the age dyspnoe and airflow obstruction index (ADO)	• Conference abstract
Singer (2005)	Mortality in a recent study of 625 patients with chronic obstructive pulmonary disease compared with results of 3 older studies	• Study does not include any relevant measures (e.g. HR, c-statistic etc.) *Data is presented as mortality ratios.*
Smith (2017)	Prognostic variables and scores identifying the end of life in COPD: a systematic review.	• Systematic review used as a source of individual studies, but not for data extraction
Soriano (2013)	Distribution and prognostic validity of the new Global Initiative for Chronic Obstructive Lung Disease grading classification	• Study population is mixed and data for people with stable COPD cannot be extracted *The study included participants from multiple cohorts, including those recruited from hospital following an acute COPD exacerbation.*
Soriano (2015)	Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data	• Study population is mixed and data for people with stable COPD cannot be extracted *The study includes a number of cohorts consisting of people with stable COPD, those recruited during an exacerbation and people*

		taking part in population studies.
Stolz (2012)	Proadrenomedullin improves the prognostic property of the BODE index	• Conference abstract
Su (2014)	Alternative options to indentify at-risk COPD patients more easily: The utility of peak expiratory flow and COPD assessment test	• Conference abstract
Suzuki (2015)	Influence of pulmonary emphysema on COPD assessment test-oriented categorization in GOLD document	• Study does not contain any relevant outcomes (mortality, hospitalisations, exacerbations, change in FEV1)
Svoboda (2016)	Risk scores for predicting death in COPD patients	• Conference abstract
Swanney (2000)	FEV(6) is an acceptable surrogate for FVC in the spirometric diagnosis of airway obstruction and restriction	• Study does not contain any relevant index tests
Tejero (2017)	Classification of Airflow Limitation Based on z-Score Underestimates Mortality in Patients with Chronic Obstructive Pulmonary Disease	• Study does not include any relevant prognostic variables
Topalovic (2013)	Computer quantification of airway collapse on forced expiration to predict the presence of emphysema	• Does not contain a population of people with suspected COPD
Topalovic (2017)	Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Tsoumakidou (2004)	Is there any correlation between the ATS, BTS, ERS and GOLD COPD's severity scales and the frequency of hospital admissions?	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Tsushima (2010)	Identification of occult parechymal disease such as emphysema or airway disease using screening computed tomography.	• Does not contain a population of people with suspected COPD
Wei (2015)	Association between serum interleukin-6 concentrations and chronic obstructive pulmonary disease: a systematic review and meta-analysis.	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)
Williams	Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index	• Retrospective prognostic cohort study
Yamamoto (2013)	Mini nutritional assessment short form (MNA-SF) can predict exacerbation in	• Conference abstract

	COPD independently of COPD assessment test (CAT)	
Zhang (2014)	Comparison of symptom and risk assessment methods among patients with chronic obstructive pulmonary disease	• Retrospective prognostic cohort study
Zhu (2014)	Sputum myeloperoxidase in chronic obstructive pulmonary disease.	• Study does not include any relevant measures (e.g. HR, c-statistic etc.)

Appendix J – Research recommendations

Diagnosing COPD

Question	What are the characteristics of people diagnosed with COPD as a result of an incidental finding of emphysema on a CT scan, compared with those diagnosed with symptoms?
Population	People diagnosed with COPD as a result of an incidental finding of emphysema on a CT scan
Characteristics of interest	• FEV1 • FVC • TLCO • Exercise capacity • Smoking history and status • BMI • Comorbidities
Study design	Prospective cohort study

Potential criterion	Explanation
Importance to patients, service users or the population	Currently, most people with COPD are diagnosed once they become symptomatic and attend their general practice. However, the ability to trigger diagnosis of COPD using an incidental CT scan is likely to result in an earlier diagnosis, which may predate COPD symptoms. It is important to assess whether the characteristics of people identified in this manner are sufficiently similar to those diagnosed in the usual way to ensure that they will benefit from the same treatment pathways.
Relevance to NICE guidance	Low priority: a recommendation was made on the use of incidental CT scans in the diagnosis of COPD, but the additional information provided by a study looking the characteristics of these people could change other recommendations in the treatment pathway.
Current evidence base	There is limited evidence on the use of incidental CT scans to diagnose COPD and the people diagnosed in this manner have yet to be well characterised.
Equality	This study could improve equality of access to a diagnosis as it provides a route to diagnosis that does not rely on people self-presenting with symptoms of COPD.
Feasibility	As the use of CT scans increases, there should be a large enough group of people diagnosed with COPD following incidental scans to carry out this study.

Predicting outcomes

Question	How can the individual factors associated with COPD prognosis (collected from a range of sources including primary care, imaging and pulmonary rehabilitation results) be combined into a multidimensional analysis that provides accurate and useful information on prognosis?
Population	People diagnosed with COPD
Assessment tools	Novel multidimensional assessment indices with components including: • Breathlessness • Chronic hypoxia and/or cor pulmonale • Long Term Oxygen Therapy (LTOT) and/or domiciliary Non-Invasive Ventilation (NIV) • Smoking status • Severity and frequency of exacerbations • Hospital admissions • Multimorbidity • Symptom burden (for example CAT score) • Frailty • BMI • FEV1 • Exercise capacity (for example, 6-minute walk test) • Transfer factor for carbon monoxide (TLCO) • Imaging results (for example, CT scan, X-Ray) • Additional pulmonary rehabilitation data
Outcomes	• Mortality • Hospitalisations • Exacerbations • Change in FEV1
Measures	• Sensitivity/specificity (preferred outcomes) • c-statistic, • Hazard ratios • Model fit (e.g. r-squared)
Study design	Prospective cohort study

Potential criterion	Explanation
Importance to patients, service users or the population	People with COPD can experience anxiety concerning their disease prognosis and the availability of suitable prognostic tools could help alleviate this stress and allow them to make plans for the future. In addition, accurate disease prognosis could help physicians tailor the appropriate level of monitoring and treatment for a person with COPD, with the goal of achieving improved outcomes.
Relevance to NICE guidance	Moderate priority: a negative recommendation was made due to the lack of a suitable prognostic index. The development of an effective test that is applicable for primary care usage could change this recommendation.
Current evidence base	There were multiple studies available looking at many indices with a range of components. However, the indices were either poor at classifying people

Potential criterion	Explanation
	into risk groups, had poor predictive ability for key outcomes or were time consuming and consisted of components that were not readily available in primary care.
Equality	No specific equality concerns are relevant to this research recommendation.
Feasibility	There is a large enough population of people with COPD that prognostic accuracy studies in this area should be feasible.

Appendix K – References

Included references

Confirming diagnosis of COPD

Systematic review

Li Jian-sheng, Zhang Hai-long, Bai Yun-ping, Wang Yan-fang, Wang Hai-feng, Wang Ming-hang, Li Su-yun, and Yu Xue-qing (2012) Diagnostic value of computed tomography in chronic obstructive pulmonary disease: a systematic review and meta-analysis. COPD 9, 563-70

Observational studies

Garcia-Pachon Eduardo (2004) Can pulse oximetry select patients for screening spirometry?.Primary care respiratory journal: journal of the General Practice Airways Group 13, 155-8

Kurashima Kazuyoshi, Takayanagi Noboru, Sato Nagato, Kanauchi Tetsu, Hoshi Toshiko, Tokunaga Daido, Ubukata Mikio, Yanagisawa Tsutomu, Sugita Yutaka, and Kanazawa Minoru (2005) High resolution CT and bronchial reversibility test for diagnosing COPD. Respirology (Carlton, and Vic.) 10, 316-22

Miniati Massimo, Coppini Giuseppe, Monti Simonetta, Bottai Matteo, Paterni Marco, and Ferdeghini Ezio Maria (2011) Computer-aided recognition of emphysema on digital chest radiography. European journal of radiology 80, e169-75

Tilemann Lisa, Gindner Lena, Meyer Franz, Szecsenyi Joachim, and Schneider Antonius (2011) Differences in local and systemic inflammatory markers in patients with obstructive airways disease. Primary care respiratory journal : journal of the General Practice Airways Group 20, 407-14

Predicting outcomes using multidimensional severity assessment indices for people with an existing diagnosis of COPD

Andrianopoulos Vasileios, Wouters Emiel F. M, Pinto-Plata Victor M, Vanfleteren Lowie E. G. W, Bakke Per S, Franssen Frits M. E, Agusti Alvar, MacNee William, Rennard Stephen I, Tal-Singer Ruth, Vogiatzis Ioannis, Vestbo Jorgen, Celli Bartolome R, and Spruit Martijn A (2015) Prognostic value of variables derived from the six-minute walk test in patients with COPD: Results from the ECLIPSE study. Respiratory medicine 109, 1138-46

Ansari Khalid, Keaney Niall, Kay Andrea, Price Monica, Munby Joan, Billett Andrew, Haggerty Sharon, Taylor Ian K, Al Otaibi, and Hajed (2016) Body mass index, airflow obstruction and dyspnea and body mass index, airflow obstruction, dyspnea scores, age and pack years-predictive properties of new multidimensional prognostic indices of chronic obstructive pulmonary disease in primary care. Annals of thoracic medicine 11, 261-268

Casanova Ciro, Cote Claudia, de Torres , Juan P, Aguirre-Jaime Armando, Marin Jose M, Pinto-Plata Victor, and Celli Bartolome R (2005) Inspiratory-to-total lung capacity ratio predicts mortality in patients with chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine 171, 591-7

Casanova Ciro, Marin Jose M, Martinez-Gonzalez Cristina, de Lucas-Ramos , Pilar , Mir-Viladrich Isabel, Cosio Borja, Peces-Barba German, Solanes-Garcia Ingrid, Aguero Ramon, Feu-Collado Nuria, Calle-Rubio Miryam, Alfageme Inmaculada, de Diego-Damia , Alfredo , Irigaray Rosa, Marin Margarita, Balcells Eva, Llunell Antonia, Galdiz Juan Bautista, Golpe Rafael, Lacarcel Celia, Cabrera Carlos, Marin Alicia, Soriano Joan B, Lopez-Campos Jose Luis, Soler-Cataluna Juan Jose, de-Torres Juan P, and Cohort Copd History Assessment in Spain (2015) Differential Effect of Modified Medical Research Council Dyspnea, COPD Assessment Test, and Clinical COPD Questionnaire for Symptoms Evaluation Within the New GOLD Staging and Mortality in COPD. Chest 148, 159-68

Celli Bartolome R, Cote Claudia G, Marin Jose M, Casanova Ciro, Montes de Oca, Maria , Mendez Reina A, Pinto Plata, Victor , and Cabral Howard J (2004) The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. The New England journal of medicine 350, 1005-12

Chan Hiang Ping, Mukhopadhyay Amartya, Chong Pauline Lee Poh, Chin Sally, Wong Xue Yun, Ong Venetia, Chan Yiong Huak, Lim Tow Keang, and Phua Jason (2016) Prognostic utility of the 2011 GOLD classification and other multidimensional tools in Asian COPD patients: a prospective cohort study. International journal of chronic obstructive pulmonary disease 11, 823-9

Chan Hiang Ping, Mukhopadhyay Amartya, Chong Pauline Lee Poh, Chin Sally, Wong Xue Yun, Ong Venetia, Chan Yiong Huak, Lim Tow Keang, and Phua Jason (2017) Role of BMI, airflow obstruction, St George's Respiratory Questionnaire and age index in prognostication of Asian COPD. Respirology (Carlton, and Vic.) 22, 114-119

Chen Chiung-Zuei, Ou Chih-Ying, Hsu Chih-Hui, and Hsiue Tzuen-Ren (2015a) Validation of the GOLD 2013 classification in predicting exacerbations and mortality in Taiwanese patients with chronic obstructive pulmonary disease. Journal of the Formosan Medical Association = Taiwan yi zhi 114, 1258-66

Chen Chiung-Zuei, Ou Chih-Ying, Yu Chun-Hsiang, Yang Szu-Chun, Chang Han-Yu, and Hsiue Tzuen-Ren (2015b) Comparison of global initiative for chronic obstructive pulmonary disease 2013 classification and body mass index, airflow obstruction, dyspnea, and exacerbations index in predicting mortality and exacerbations in elderly adults with chronic obstructive pulmonary disease. Journal of the American Geriatrics Society 63, 244-50

Cote C G, Pinto-Plata V M, Marin J M, Nekach H, Dordelly L J, and Celli B R (2008) The modified BODE index: validation with mortality in COPD. The European respiratory journal 32, 1269-74

de Torres , Juan P, Casanova Ciro, Marin Jose M, Pinto-Plata Victor, Divo Miguel, Zulueta Javier J, Berto Juan, Zagaceta Jorge, Sanchez-Salcedo Pablo, Cabrera Carlos, Carrizo Santiago, Cote Claudia, and Celli Bartolome R (2014) Prognostic evaluation of COPD patients: GOLD 2011 versus BODE and the COPD comorbidity index COTE. Thorax 69, 799-804

de Torres , Juan P, Pinto-Plata Victor, Casanova Ciro, Mullerova Hanna, Cordoba-Lanus Elizabeth, de Fuentes , Mercedes Muros, Aguirre-Jaime Armando, and Celli Bartolome R (2008) C-reactive protein levels and survival in patients with moderate to very severe COPD. Chest 133, 1336-1343

Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Zulueta J, Cabrera C, Zagaceta J, Hunninghake G, and Celli B (2012) Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.. American journal of respiratory and critical care medicine 186(2), 155-61

Eisner Mark D, Omachi Theodore A, Katz Patricia P, Yelin Edward H, Iribarren Carlos, and Blanc Paul D (2010) Measurement of COPD severity using a survey-based score: validation in a clinically and physiologically characterized cohort. Chest 137, 846-51

Esteban C, Quintana J M, Aburto M, Moraza J, and Capelastegui A (2006) A simple score for assessing stable chronic obstructive pulmonary disease. QJM 99, 751-759

Esteban Cristobal, Quintana Jose M, Moraza Javier, Aburto Myriam, Aguirre Urko, Aguirregomoscorta Jose I, Aizpiri Susana, Basualdo Luis V, and Capelastegui Alberto (2010) BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice?. BMC medicine 8, 28

Esteban Cristobal, Quintana Jose Maria, Aburto Myriam, Moraza Javier, Arostegui Inmaculada, Espana P P, Aizpiri Susana, and Capelastegui Alberto (2011) The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. Respiratory medicine 105, 1662-70

Faganello Marcia Maria, Tanni Suzana Erico, Sanchez Fernanda Figueiroa, Pelegrino Nilva Regina Gelamo, Lucheta Paulo Adolfo, and Godoy Irma (2010) BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. The American journal of the medical sciences 339, 10-4

Goossens Lucas M. A, Leimer Inge, Metzdorf Norbert, Becker Karin, Rutten-van Molken, and Maureen P M. H (2014) Does the 2013 GOLD classification improve the ability to predict lung function decline, exacerbations and mortality: a post-hoc analysis of the 4-year UPLIFT trial. BMC pulmonary medicine 14, 163

Imfeld Stephan, Bloch Konrad E, Weder Walter, and Russi Erich W (2006) The BODE index after lung volume reduction surgery correlates with survival. Chest 129, 873-8

Johannessen Ane, Nilsen Roy M, Storebo Michael, Gulsvik Amund, Eagan Tomas, and Bakke Per (2013) Comparison of 2011 and 2007 Global Initiative for Chronic Obstructive Lung Disease guidelines for predicting mortality and hospitalization. American journal of respiratory and critical care medicine 188, 51-9

Lee Sang-Do, Huang Ming-Shyan, Kang Jian, Lin Ching-Hsiung, Park Myung Jae, Oh Yeon-Mok, Kwon Namhee, Jones Paul W, Sajkov Dimitar, Investigators of the Predictive Ability of, and C A T. in Acute Exacerbations of Copd Study (2014) The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. Respiratory medicine 108, 600-8

Leivseth L, Brumpton BM, Nilsen TI, Mai XM, Johnsen R, and Langhammer A (2013) GOLD classifications and mortality in chronic obstructive pulmonary disease: the HUNT Study, Norway.. Thorax 68(10), 914-21

Marin J M, Alfageme I, Almagro P, Casanova C, Esteban C, Soler-Cataluna J J, De Torres , J P, Martinez-Camblor P, Miravitlles M, Celli B R, and Soriano J B (2013) Multicomponent indices to predict survival in COPD: The COCOMICS study. European Respiratory Journal 42, 323-332

Marin Jose M, Carrizo Santiago J, Casanova Ciro, Martinez-Camblor Pablo, Soriano Joan B, Agusti Alvar G. N, and Celli Bartolome R (2009) Prediction of risk of COPD exacerbations by the BODE index. Respiratory medicine 103, 373-8

Marin Jose M, Cote Claudia G, Diaz Orlando, Lisboa Carmen, Casanova Ciro, Lopez Maria V, Carrizo Santiago J, Pinto-Plata Victor, Dordelly Luis J, Nekach Hafida, and Celli Bartolome R (2011) Prognostic assessment in COPD: health related quality of life and the BODE index. Respiratory medicine 105, 916-21

Mattila T, Vasankari T, Kanervisto M, Laitinen T, Impivaara O, Rissanen H, Knekt P, Jousilahti P, Saarelainen S, Puukka P, and Heliovaara M (2015) Association between all-cause and cause-specific mortality and the GOLD stages 1-4: A 30-year follow-up among Finnish adults. Respiratory Medicine 109, 1012-1018

Moberg Mia, Vestbo Jorgen, Martinez Gerd, Williams Johanna E. A, Ladelund Steen, Lange Peter, and Ringbaek Thomas (2014) Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD participating in pulmonary rehabilitation. COPD 11, 381-7

Motegi Takashi, Jones Rupert C, Ishii Takeo, Hattori Kumiko, Kusunoki Yuji, Furutate Ryuko, Yamada Kouich, Gemma Akihiko, and Kida Kozui (2013) A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. International journal of chronic obstructive pulmonary disease 8, 259-71

Moy Marilyn L, Teylan Merilee, Danilack Valery A, Gagnon David R, and Garshick Eric (2014) An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. Annals of the American Thoracic Society 11, 149-57

Neo Han-Yee, Xu Hui-Ying, Wu Huei-Yaw, and Hum Allyn (2017) Prediction of Poor Short-Term Prognosis and Unmet Needs in Advanced Chronic Obstructive Pulmonary Disease: Use of the Two-Minute Walking Distance Extracted from a Six-Minute Walk Test. Journal of palliative medicine 20, 821-828

Omachi Theodore A, Yelin Edward H, Katz Patricia P, Blanc Paul D, and Eisner Mark D (2008) The COPD severity score: a dynamic prediction tool for health-care utilization. COPD 5, 339-46

Ou Chih-Ying, Chen Chiung-Zuei, Yu Chun-Hsiang, Shiu Chih-Hui, and Hsiue Tzuen-Ren (2014) Discriminative and predictive properties of multidimensional prognostic indices of chronic obstructive pulmonary disease: a validation study in Taiwanese patients. Respirology (Carlton, and Vic.) 19, 694-9

Pedone Claudio, Scarlata Simone, Forastiere Francesco, Bellia Vincenzo, Antonelli Incalzi, and Raffaele (2014) BODE index or geriatric multidimensional assessment for the prediction of very-long-term mortality in elderly patients with chronic obstructive pulmonary disease? a prospective cohort study. Age and ageing 43, 553-8

Pothirat Chaicharn, Chaiwong Warawut, Limsukon Atikun, Deesomchok Athavudh, Liwsrisakun Chalerm, Bumroongkit Chaiwat, Theerakittikul Theerakorn, and Phetsuk Nittaya (2015) Detection of acute deterioration in health status visit among COPD patients by monitoring COPD assessment test score. International journal of chronic obstructive pulmonary disease 10, 277-82

Puhan Milo A, Garcia-Aymerich Judith, Frey Martin, ter Riet, Gerben , Anto Josep M, Agusti Alvar G, Gomez Federico P, Rodriguez-Roisin Roberto, Moons Karel G. M, Kessels Alphons G, and Held Ulrike (2009) Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. Lancet (London, and England) 374, 704-11

Soler-Cataluna Juan Jose, Martinez-Garcia Miguel Angel, Sanchez Lourdes Sanchez, Tordera Miguel Perpina, and Sanchez Pilar Roman (2009) Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. Respiratory medicine 103, 692-9

Stolz Daiana, Meyer Anja, Rakic Janko, Boeck Lucas, Scherr Andreas, and Tamm Michael (2014a) Mortality risk prediction in COPD by a prognostic biomarker panel. The European respiratory journal 44, 1557-70

Stolz Daiana, Kostikas Kostantinos, Blasi Francesco, Boersma Wim, Milenkovic Branislava, Lacoma Alicia, Louis Renaud, Aerts Joachim G, Welte Tobias, Torres Antoni, Rohde Gernot G. U, Boeck Lucas, Rakic Janko, Scherr Andreas, Hertel Sabine, Giersdorf Sven, and Tamm Michael (2014b) Adrenomedullin refines mortality prediction by the BODE index in COPD: the "BODE-A" index. The European respiratory journal 43, 397-408

Suetomo Masashi, Kawayama Tomotaka, Kinoshita Takashi, Takenaka Shinichi, Matsuoka Masanobu, Matsunaga Kazuko, and Hoshino Tomoaki (2014) COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. Respiratory investigation 52, 288-95

Sundh Josefin, Janson Christer, Lisspers Karin, Montgomery Scott, and Stallberg Bjorn (2012a) Clinical COPD Questionnaire score (CCQ) and mortality. International journal of chronic obstructive pulmonary disease 7, 833-42

Sundh Josefin, Janson Christer, Lisspers Karin, Stallberg Bjorn, and Montgomery Scott (2012b) The Dyspnoea, Obstruction, Smoking, Exacerbation (DOSE) index is predictive of mortality in COPD. Primary care respiratory journal: journal of the General Practice Airways Group 21, 295-301

Thabut Gabriel, Mornex Jean-Francois, Pison Christophe, Cuvelier Antoine, Balduyck Malika, Pujazon Marie-Christine, Fournier Michel, AitIlalne Brahim, and Porcher Raphael (2014) Performance of the BODE index in patients with alpha1-antitrypsin deficiency-related COPD. The European respiratory journal 44, 78-86

Varol Yelda, Ozacar Rifat, Balci Gunseli, Usta Levent, and Taymaz Zuhre (2014) Assessing the effectiveness of the COPD Assessment Test (CAT) to evaluate COPD severity and exacerbation rates. COPD 11, 221-5

Excluded references for all reviews in this evidence review

Abascal-Bolado Beatriz, Novotny Paul J, Sloan Jeff A, Karpman Craig, Dulohery Megan M, and Benzo Roberto P (2015) Forecasting COPD hospitalization in the clinic: optimizing the chronic respiratory questionnaire. International journal of chronic obstructive pulmonary disease 10, 2295-301

Almagro Pere, Martinez-Camblor Pablo, Soriano Joan B, Marin Jose M, Alfageme Inmaculada, Casanova Ciro, Esteban Cristobal, Soler-Cataluna Juan J, de-Torres Juan P,

Celli Bartolome R, and Miravitlles Marc (2014) Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study. PloS one 9, e89866

Al-Mutairi S S, Sharma P N, Al-Alawi A, and Al-Deen J S (2007) Impulse oscillometry: an alternative modality to the conventional pulmonary function test to categorise obstructive pulmonary disorders. Clinical and experimental medicine 7, 56-64

Arghir O, Manescu V, Suta M, and Postolache P (2011) The use of FEV1-body mass index corellation in evaluating the severity and prognosis of COPD in active workers patients. Chest 140,

Baughman Penelope, Marott Jacob L, Lange Peter, Martin Christopher J, Shankar Anoop, Petsonk Edward L, and Hnizdo Eva (2012) Combined effect of lung function level and decline increases morbidity and mortality risks. European journal of epidemiology 27, 933-43

Bhatt S P, Sieren J C, Dransfield M T, Washko G R, Newell Jr, J D, Stinson D A, Zamba G K, and Hoffman E A (2013) Comparison of spirometric thresholds in diagnosing smoking-related airflow obstruction. COPD: Journal of Chronic Obstructive Pulmonary Disease 10, 400

Bhatt Surya P, Kim Young-Il, Wells James M, Bailey William C, Ramsdell Joe W, Foreman Marilyn G, Jensen Robert L, Stinson Douglas S, Wilson Carla G, Lynch David A, Make Barry J, and Dransfield Mark T (2014) FEV(1)/FEV(6) to diagnose airflow obstruction. Comparisons with computed tomography and morbidity indices. Annals of the American Thoracic Society 11, 335-41

Blumenthal James A, Smith Patrick J, Durheim Michael, Mabe Stephanie, Emery Charles F, Martinu Tereza, Diaz Philip T, Babyak Michael, Welty-Wolf Karen, and Palmer Scott (2016) Biobehavioral Prognostic Factors in Chronic Obstructive Pulmonary Disease: Results From the INSPIRE-II Trial. Psychosomatic medicine 78, 153-62

Boeck L, Kostikas K, Blasi F, Boersma W, Milenkovic B, Lacoma A, Louis R, Aerts J, Welte T, Torres A, Rohde G, Rakic J, Scherr A, Koller M, Tamm M, and Stolz D (2015) Prognostic assessment in chronic obstructive pulmonary disease using copeptin: A simplified risk index. European Respiratory Journal 46,

Boeck Lucas, Soriano Joan B, Brusse-Keizer Marjolein, Blasi Francesco, Kostikas Konstantinos, Boersma Wim, Milenkovic Branislava, Louis Renaud, Lacoma Alicia, Djamin Remco, Aerts Joachim, Torres Antoni, Rohde Gernot, Welte Tobias, Martinez-Camblor Pablo, Rakic Janko, Scherr Andreas, Koller Michael, van der Palen , Job , Marin Jose M, Alfageme Inmaculada, Almagro Pere, Casanova Ciro, Esteban Cristobal, Soler-Cataluna Juan J, de-Torres Juan P, Miravitlles Marc, Celli Bartolome R, Tamm Michael, and Stolz Daiana (2016) Prognostic assessment in COPD without lung function: the B-AE-D indices. The European respiratory journal 47, 1635-44

Boutou Afroditi K, Shrikrishna Dinesh, Tanner Rebecca J, Smith Cayley, Kelly Julia L, Ward Simon P, Polkey Michael I, and Hopkinson Nicholas S (2013) Lung function indices for predicting mortality in COPD. The European respiratory journal 42, 616-25

Boutou Afroditi K, Nair Arjun, Douraghi-Zadeh Dariush, Sandhu Ranbir, Hansell David M, Wells Athol U, Polkey Michael I, and Hopkinson Nicholas S (2014) A combined pulmonary

function and emphysema score prognostic index for staging in Chronic Obstructive Pulmonary Disease. PloS one 9, e111109

Broekhuizen B D. L, Sachs A, Janssen K, Geersing G J, Moons K, Hoes A, and Verheij T (2011) Does a decision aid help physicians to detect chronic obstructive pulmonary disease?. British Journal of General Practice 61, e674-e679

Brusse-Keizer Marjolein, Zuur-Telgen Maaike, van der Palen , Job , VanderValk Paul, Kerstjens Huib, Boersma Wim, Blasi Francesco, Kostikas Konstantinos, Milenkovic Branislava, Tamm Michael, and Stolz Daiana (2015) Adrenomedullin optimises mortality prediction in COPD patients. Respiratory medicine 109, 734-42

Brusse-Keizer M, Klatte M, Zuur-Telgen M, Koehorst-Ter Huurne, K , van der Palen , J , and VanderValk P (2017) Comparing the 2007 and 2011 GOLD Classifications as Predictors of all-Cause Mortality and Morbidity in COPD. COPD 14, 7-14

Celli Bartolome R, Locantore Nicholas, Yates Julie, Tal-Singer Ruth, Miller Bruce E, Bakke Per, Calverley Peter, Coxson Harvey, Crim Courtney, Edwards Lisa D, Lomas David A, Duvoix Annelyse, MacNee William, Rennard Stephen, Silverman Edwin, Vestbo Jorgen, Wouters Emiel, Agusti Alvar, and Investigators Eclipse (2012) Inflammatory biomarkers improve clinical prediction of mortality in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine 185, 1065-72

Chan H P, Mukhopadhyay A, Chong P, Chin S, Wong X Y, Ong V, Chan Y H, Lim T K, and Phua J (2015) Prognostication in patients with chronic obstructive pulmonary disease using the BOS index. Chest 148,

Chang Chun, Zhu Hong, Shen Ning, Han Xiang, Chen Yahong, and He Bei (2014) Utility of the combination of serum highly-sensitive C-reactive protein level at discharge and a risk index in predicting readmission for acute exacerbation of COPD. Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia 40, 495-503

Chen Chiung-Zuei, Ou Chih-Ying, Wang Wen-Ling, Lee Cheng-Hung, Lin Chien-Chung, Chang Han-Yu, and Hsiue Tzuen-Ren (2012) Using post-bronchodilator FEV1 is better than pre-bronchodilator FEV1 in evaluation of COPD severity. COPD 9, 276-80

Chen Feng-Jia, Huang Xin-Yan, Liu Yang-Li, Lin Geng-Peng, and Xie Can-Mao (2016) Importance of fractional exhaled nitric oxide in the differentiation of asthma-COPD overlap syndrome, asthma, and COPD. International journal of chronic obstructive pulmonary disease 11, 2385-2390

Dal Negro, Roberto W, Bonadiman Luca, and Turco Paola (2014) Sensitivity of the COPD assessment test (CAT questionnaire) investigated in a population of 681 consecutive patients referring to a lung clinic: the first Italian specific study. Multidisciplinary respiratory medicine 9, 15

Dal Negro, R W, and Celli B R (2017) Patient Related Outcomes-BODE (PRO-BODE): A composite index incorporating health utilization resources predicts mortality and economic cost of COPD in real life. Respiratory Medicine 131, 175-178

Dawkins P A, Dowson L J, Guest P J, and Stockley R A (2003) Predictors of mortality in alpha1-antitrypsin deficiency. Thorax 58, 1020-6

Dawkins Paul, Wood Alice, Nightingale Peter, and Stockley Robert (2009) Mortality in alpha-1-antitrypsin deficiency in the United Kingdom. Respiratory medicine 103, 1540-7

DeVries Rebecca, Kriebel David, and Sama Susan (2016) Validation of the breathlessness, cough and sputum scale to predict COPD exacerbation. NPJ primary care respiratory medicine 26, 16083

Dijk Wouter D. van, Bemt Lisette van den, Haak-Rongen Saskia van den, Bischoff Erik, Weel Chris van, Veen Johannes C. C. M. in 't, and Schermer Tjard R. J (2011) Multidimensional prognostic indices for use in COPD patient care. A systematic review. Respiratory research 12, 151

Eisner Mark D, Trupin Laura, Katz Patricia P, Yelin Edward H, Earnest Gillian, Balmes John, and Blanc Paul D (2005) Development and validation of a survey-based COPD severity score. Chest 127, 1890-7

Eriksson B, Ronmark E, Lindberg A, Mullerova H, and Lundback B (2011) BMI and risk for death in COPD. European Respiratory Journal 38,

Esteban Cristobal, Quintana Jose M, Aburto Myriam, Moraza Javier, Egurrola Mikel, Espana Pedro Pablo, Perez-Izquierdo Julio, and Capelastegui Alberto (2008) Predictors of mortality in patients with stable COPD. Journal of general internal medicine 23, 1829-34

Feliz-Rodriguez Darwin, Zudaire Santiago, Carpio Carlos, Martinez Elizabet, Gomez-Mendieta Antonia, Santiago Ana, Alvarez-Sala Rodolfo, and Garcia-Rio Francisco (2013) Evolution of the COPD Assessment Test score during chronic obstructive pulmonary disease exacerbations: determinants and prognostic value. Canadian respiratory journal 20, e92-7

Fragoso Carlos A. Vaz, Concato John, McAvay Gail, Yaggi H Klar, Van Ness , Peter H, and Gill Thomas M (2011) Staging the severity of chronic obstructive pulmonary disease in older persons based on spirometric Z-scores. Journal of the American Geriatrics Society 59, 1847-54

Gedebjerg Anne, Szepligeti Szimonetta Komjathine, Wackerhausen Laura-Maria Holm, Horvath-Puho Erzsebet, Dahl Ronald, Hansen Jens Georg, Sorensen Henrik Toft, Norgaard Mette, Lange Peter, and Thomsen Reimar Wernich (2018) Prediction of mortality in patients with chronic obstructive pulmonary disease with the new Global Initiative for Chronic Obstructive Lung Disease 2017 classification: a cohort study. The Lancet. Respiratory medicine 6, 204-212.

Ghobadi Hassan, Ahari Saeid Sadeghieh, Kameli Azadeh, and Lari Sharzad M (2012) The Relationship between COPD Assessment Test (CAT) Scores and Severity of Airflow Obstruction in Stable COPD Patients. Tanaffos 11, 22-6

Guirguis-Blake Janelle M, Senger Caitlyn A, Webber Elizabeth M, Mularski Richard A, and Whitlock Evelyn P (2016) Screening for Chronic Obstructive Pulmonary Disease: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 315, 1378-93

Gupta N, Pinto L M, Morogan A, and Bourbeau J (2014) The COPD assessment test: a systematic review. The European respiratory journal 44(4), 873-884

Haroon S, Jordan R, and Adab P (2014) Screening test accuracy of case finding interventions for chronic obstructive pulmonary disease in primary care: A systematic review. American Journal of Respiratory and Critical Care Medicine 189,

Haroon Shamil, Jordan Rachel, Takwoingi Yemisi, and Adab Peymane (2015) Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis. BMJ open 5, e008133

Haroon S M, Jordan R E, O'Beirne-Elliman J, and Adab P (2015) Effectiveness of case finding strategies for COPD in primary care: A systematic review and meta-analysis. npj Primary Care Respiratory Medicine 25, 15056

Hernandez V, Sobradillo P, Tomas L, Cancelo L, Ribas F J, Azpiazu M, Zorrilla V, Lahidalga B, and Lobo J L (2012) Prognostic factors in COPD patients controlled in two outpatient clinics. European Respiratory Journal 40,

Horita Nobuyuki, Koblizek Vladimir, Plutinsky Marek, Novotna Barbora, Hejduk Karel, and Kaneko Takeshi (2016) Chronic obstructive pulmonary disease prognostic score: A new index. Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, and Czechoslovakia 160, 211-8

Jing Ji-Yong, Huang Tian-Cha, Cui Wei, Xu Feng, and Shen Hua-Hao (2009) Should FEV1/FEV6 replace FEV1/FVC ratio to detect airway obstruction? A metaanalysis. Chest 135, 991-998

Jones Rupert C, Price David, Chavannes Niels H, Lee Amanda J, Hyland Michael E, Stallberg Bjorn, Lisspers Karin, Sundh Josefin, van der Molen , Thys , Tsiligianni Ioanna, and IPCRG Unlock Group of the (2016) Multi-component assessment of chronic obstructive pulmonary disease: an evaluation of the ADO and DOSE indices and the global obstructive lung disease categories in international primary care data sets. NPJ primary care respiratory medicine 26, 16010

Jung Y R, Kim San, H , Yong S J, Shin K C, Lee Wo, Y , and Lee M K (2014) Clinical features and prognostic factors of patients hospitalized due to acute exacerbation of chronic obstructive pulmonary disease. Respirology 19, 102

Kelly JL, Bamsey O, Smith C, Lord VM, Shrikrishna D, Jones PW, Polkey MI, and Hopkinson NS (2012) Health status assessment in routine clinical practice: the chronic obstructive pulmonary disease assessment test score in outpatients.. Respiration, and international review of thoracic diseases 84(3), 193-9

Ko Fanny W. S, Tam Wilson, Tung Alvin H. M, Ngai Jenny, Ng Susanna S. S, Lai Kenneth, Au Ka-Fai, and Hui David S. C (2011) A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD. Respiratory medicine 105, 266-73

Kostianev Stefan S, Hodgev Vladimir A, and Iluchev Dimitar H (2008) Multidimensional system for assessment of COPD patients. Comparison with BODE index. Folia medica 50, 29-38

Kwon Beom-June, Kim Dong-Bin, Jang Sung-Won, Yoo Ki-Dong, Moon Keun-Woong, Shim Byung Ju, Ahn Seo-Hee, Cho Eun-Ju, Rho Tae-Ho, and Kim Jae-Hyung (2010) Prognosis of heart failure patients with reduced and preserved ejection fraction and coexistent chronic obstructive pulmonary disease. European journal of heart failure 12, 1339-44

Labor M, Vrbica Z, Gudelj I, Labor S, Juric I, Rosandic-Piasevoli R, Tudja K, Mestrovic Z, Martinkovic D, and Plavec D (2015) Exhaled breath temperature as a possible early marker in smokers at risk for COPD (MARKO study). European Respiratory Journal 46,

Lamprecht Bernd, Schirnhofer Lea, Tiefenbacher Falko, Kaiser Bernhard, Buist Sonia A, Studnicka Michael, and Enright Paul (2007) Six-second spirometry for detection of airway obstruction: a population-based study in Austria. American journal of respiratory and critical care medicine 176, 460-4

Lange Peter, Marott Jacob Louis, Vestbo Jorgen, Olsen Kim Rose, Ingebrigtsen Truls Sylvan, Dahl Morten, and Nordestgaard Borge Gronne (2012) Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. American journal of respiratory and critical care medicine 186, 975-81

Lederer David J, Thomashow Byron M, Ginsburg Mark E, Austin John H. M, Bartels Matthew N, Yip Chun K, Jellen Patricia A, Brogan Frances L, Kawut Steven M, Maxfield Roger A, DiMango Angela M, Simonelli Paul F, Gorenstein Lyall A, Pearson Gregory D. N, and Sonett Joshua R (2007) Lung-volume reduction surgery for pulmonary emphysema: Improvement in body mass index, airflow obstruction, dyspnea, and exercise capacity index after 1 year. The Journal of thoracic and cardiovascular surgery 133, 1434-8

Lee Jae Seung, Seo Joon Beom, Lee Sang Min, Park Tai Sun, Lee Sei Won, Oh Yeon-Mok, Lee Ji-Hyun, Kim Eun-Kyung, Kim Tae-Hyung, Park Joo Hun, Sheen Seung Soo, Lim Seong Yong, Jung Ina, and Lee Sang-Do (2015) Pharmacological treatment response according to the severity of symptoms in patients with chronic obstructive pulmonary disease. Journal of thoracic disease 7, 1765-73

Liu Shih-Feng, Wang Chin-Chou, Chin Chien-Hung, Chen Yung-Che, and Lin Meng-Chih (2011) High value of combined serum C-reactive protein and BODE score for mortality prediction in patients with stable COPD. Archivos de bronconeumologia 47, 427-32

Lou Peian, Chen Peipei, Zhang Pan, Yu Jiaxi, Wang Yong, Chen Na, Zhang Lei, Wu Hongmin, and Zhao Jing (2016) Interaction of Depression and Nicotine Addiction on the Severity of Chronic Obstructive Pulmonary Disease: A Prospective Cohort Study. Iranian journal of public health 45, 146-57

Manninen H, Partanen K, Soimakallio S, and Rytkonen H (1988) Large-image intensifier photofluorography and conventional radiography in pulmonary emphysema. Correlation with computed tomography. Acta radiologica (Stockholm, and Sweden : 1987) 29, 293-7

Mannino David M, Doherty Dennis E, Sonia Buist, and A (2006) Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study. Respiratory medicine 100, 115-22

Marsh Suzanne, Aldington Sarah, Williams Mathew V, Nowitz Michael R, Kingzett-Taylor Andrew, Weatherall Mark, Shirtcliffe Philippa M, McNaughton Amanda A, Pritchard Alison, and Beasley Richard (2007) Utility of lung density measurements in the diagnosis of emphysema. Respiratory medicine 101, 1512-20

Martinez Carlos H, Murray Susan, Barr R Graham, Bleecker Eugene, Bowler Russell P, Christenson Stephanie A, Comellas Alejandro P, Cooper Christopher B, Couper David, Criner Gerard J, Curtis Jeffrey L, Dransfield Mark T, Hansel Nadia N, Hoffman Eric A,

Kanner Richard E, Kleerup Eric, Krishnan Jerry A, Lazarus Stephen C, Leidy Nancy K, O'Neal Wanda, Martinez Fernando J, Paine Robert 3rd, Rennard Stephen I, Tashkin Donald P, Woodruff Prescott G, Han MeiLan K, Subpopulations , Intermediate Outcome Measures in, and Copd Study Investigators (2017) Respiratory Symptoms Items from the COPD Assessment Test Identify Ever-Smokers with Preserved Lung Function at Higher Risk for Poor Respiratory Outcomes. An Analysis of the Subpopulations and Intermediate Outcome Measures in COPD Study Cohort. Annals of the American Thoracic Society 14, 636-642

Mets Onno M, Buckens Constantinus F. M, Zanen Pieter, Isgum Ivana, van Ginneken , Bram , Prokop Mathias, Gietema Hester A, Lammers Jan-Willem J, Vliegenthart Rozemarijn, Oudkerk Matthijs, van Klaveren , Rob J, de Koning , Harry J, Mali Willem P. Th M, de Jong , and Pim A (2011) Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. JAMA 306, 1775-81

Miravitlles M, Llor C, de Castellar , R , Izquierdo I, Baro E, and Donado E (2009) Validation of the COPD severity score for use in primary care: the NEREA study. The European respiratory journal 33, 519-27

Mohamed Hoesein, Firdaus A A, Zanen Pieter, Sachs Alfred P. E, Verheij Theo J. M, Lammers Jan-Willem J, and Broekhuizen Berna D. L (2012) Spirometric thresholds for diagnosing COPD: 0.70 or LLN, pre- or post-dilator values?. COPD 9, 338-43

Morris Zachary Q, Huda Najia, and Burke Robert R (2012) The diagnostic importance of a reduced FEV1/FEV6. COPD 9, 22-8

Motegi T, Wakabayashi R, Hattori M, Ishii T, Kamio K, Yamada K, Morii K, Kurahara M, Gemma A, and Kida K (2011) Efficacy of prognostic assessment using ado index for Japanese patients with COPD. COPD: Journal of Chronic Obstructive Pulmonary Disease 8, 43-44

Moya V, Boldova A, Minchole E, Carrizo S, Navarro C, Sam A, Miravitlles M, and Marin J M (2014) Airflow reversibility on long term outcomes of patients with COPD. European Respiratory Journal 44,

Navarro A, Costa R, Rodriguez-Carballeira M, Yun S, Lapuente A, Barrera A, Acosta E, Vinas C, Heredia J L, and Almagro P (2015) Prognostic assessment of mortality and hospitalizations of outpatients with advanced chronic obstructive pulmonary disease. Usefulness of the CODEX index. Valoracion pronostica de mortalidad y hospitalizaciones en pacientes ambulatorios con EPOC avanzada. Utilidad del indice CODEX. 215, 431-8

Neo H Y, Xu H Y, Hum A, and Wu H Y (2016) A pilot study examining predictors of short term mortality in advanced COPD-importance of nutritional, systemic inflammatory and physical performance indices. Palliative Medicine 30, NP238

Nishimura K, Izumi T, Tsukino M, and Oga T (2002) Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. Chest 121, 1434-1440

Nizet Tessa A. C, van den Elshout , Frank J J, Heijdra Yvonne F, van de Ven , Marjo J T, Mulder Paul G. H, and Folgering Hans Th M (2005) Survival of chronic hypercapnic COPD patients is predicted by smoking habits, comorbidity, and hypoxemia. Chest 127, 1904-10

Oga T, Nishimura K, Tsukino M, Sato S, Hajiro T, Ikeda A, and Mishima M (2002) Health status measured with the CRQ does not predict mortality in COPD. European Respiratory Journal 20, 1147-1151

Oga Toru, Tsukino Mitsuhiro, Hajiro Takashi, Ikeda Akihiko, and Nishimura Koichi (2011) Predictive properties of different multidimensional staging systems in patients with chronic obstructive pulmonary disease. International journal of chronic obstructive pulmonary disease 6, 521-6

Ong Kian-Chung, Earnest Arul, and Lu Suat-Jin (2005) A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. Chest 128, 3810-6

Ozgur Eylem Sercan, Nayci Sibel Atis, Ozge Cengiz, and Tasdelen Bahar (2012) An integrated index combined by dynamic hyperinflation and exercise capacity in the prediction of morbidity and mortality in COPD. Respiratory care 57, 1452-9

Panjabi C, Khake S, Isser V S, and Kochhar G S (2015) Usefulness of the COPD assessment test (CAT) in patients with (A) stable disease, and (B) exacerbations. Chest 148,

Papaioannou A I, Bartziokas K, Tsikrika S, Karakontaki F, Kastanakis E, Banya W, Haniotou A, Papiris S, Loukides S, Polychronopoulos V, and Kostikas K (2013) The impact of depressive symptoms on recovery and outcome of hospitalised COPD exacerbations. European Respiratory Journal 41, 815-823

Papaioannou Maria, Pitsiou Georgia, Manika Katerina, Kontou Paschalina, Zarogoulidis Pavlos, Sichletidis Lazaros, and Kioumis Ioannis P (2014) COPD assessment test: a simple tool to evaluate disease severity and response to treatment. COPD 11, 489-95

Pavasini R, D'Ascenzo F, Campo G, Biscaglia S, Ferri A, Contoli M, Papi A, Ceconi C, and Ferrari R (2015) Cardiac troponin elevation predicts all-cause mortality in patients with acute exacerbation of chronic obstructive pulmonary disease: Systematic review and meta-analysis. International Journal of Cardiology 191, 187-193

Percival M, Sriram K, Chapman L, Bishop P, Drysdale A, Allen P, Murrells J, and Sivakumaran P (2014) Utility of the copd assessment test (CAT) in evaluating copd severity. Respirology 19, 98

Pla J J, Mena P A, Soriano J B, Cabrera F, Boixeda R, Ortiz M B. A, Murio C, Diez J, Barreiro B, and Heredia J L (2014) Short and medium term prognosis in patients hospitalized for acute exacerbation of COPD (AECOPD): The codex index. Chest 145,

Pudney E, and Doherty M (2016) Plain chest x-ray (CXR) in the diagnosis of chronic obstructive pulmonary disease (COPD). European Respiratory Journal 48,

Puhan Milo A, Hansel Nadia N, Sobradillo Patricia, Enright Paul, Lange Peter, Hickson Demarc, Menezes Ana M, ter Riet, Gerben , Held Ulrike, Domingo-Salvany Antonia, Mosenifar Zab, Anto Josep M, Moons Karel G. M, Kessels Alphons, Garcia-Aymerich Judith, and International Copd Cohorts Collaboration Working Group (2012) Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. BMJ open 2,

Qu Yanjuan, Cao Yiyuan, Liao Meiyan, and Lu Zhiyan (2017) Sagittal-lung CT measurements in the evaluation of asthma-COPD overlap syndrome: a distinctive phenotype from COPD alone. La Radiologia medica 122, 487-494

Rahman S (2014) Prognostic evaluation of COPD patients using bode index. Chest 146,

Rozenberg D, and Stanbrook M B (2012) Non-spirometric pulmonary function parameters for differentiating COPD and asthma. American Journal of Respiratory and Critical Care Medicine 185,

Sanders C, Nath P H, and Bailey W C (1988) Detection of emphysema with computed tomography. Correlation with pulmonary function tests and chest radiography. Investigative radiology 23, 262-6

Schapira R M, Schapira M M, Funahashi A, McAuliffe T L, and Varkey B (1993) The value of the forced expiratory time in the physical diagnosis of obstructive airways disease. JAMA 270, 731-6

Schonenberger L, Jochmann A, Bridevaux P O, Chhajed P N, Fitting J W, Geiser T, Kohler M, Maier S, Miedinger D, Torok S S, Scherr A, Thurnheer R, Tamm M, and Leuppi J D (2012) Prediction of mortality in the Swiss chronic obstructive pulmonary disease (COPD) cohort using the age dyspnoe and airflow obstruction index (ADO). European Respiratory Journal 40,

Singer Richard B (2005) Mortality in a recent study of 625 patients with chronic obstructive pulmonary disease compared with results of 3 older studies. Journal of insurance medicine (New York, and N.Y.) 37, 179-84

Smith LE, Moore E, Ali I, Smeeth L, Stone P, and Quint JK (2017) Prognostic variables and scores identifying the end of life in COPD: a systematic review.. International journal of chronic obstructive pulmonary disease 12, 2239-2256

Soriano Joan B, Alfageme Inmaculada, Almagro Pere, Casanova Ciro, Esteban Cristobal, Soler-Cataluna Juan J, de Torres , Juan P, Martinez-Camblor Pablo, Miravitlles Marc, Celli Bartolome R, and Marin Jose M (2013) Distribution and prognostic validity of the new Global Initiative for Chronic Obstructive Lung Disease grading classification. Chest 143, 694-702

Soriano Joan B, Lamprecht Bernd, Ramirez Ana S, Martinez-Camblor Pablo, Kaiser Bernhard, Alfageme Inmaculada, Almagro Pere, Casanova Ciro, Esteban Cristobal, Soler-Cataluna Juan J, de-Torres Juan P, Miravitlles Marc, Celli Bartolome R, Marin Jose M, Puhan Milo A, Sobradillo Patricia, Lange Peter, Sternberg Alice L, Garcia-Aymerich Judith, Turner Alice M, Han MeiLan K, Langhammer Arnulf, Leivseth Linda, Bakke Per, Johannessen Ane, Roche Nicolas, and Sin Don D (2015) Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. The Lancet. Respiratory medicine 3, 443-50

Stolz D, Boersma W, Milenkovic B, Kostikas K, Blasi F, Aerts J, Louis R, Rohde G, Lacoma A, Boeck L, Rakic J, Scherr A, Torres A, Welte T, Hartmann O, Giersdorf S, and Tamm M (2012) Proadrenomedullin improves the prognostic property of the BODE index. European Respiratory Journal 40,

Su K C, and Perng D W (2014) Alternative options to indentify at-risk COPD patients more easily: The utility of peak expiratory flow and COPD assessment test. Respirology 19, 128

Suzuki Toshio, Tada Yuji, Kawata Naoko, Ikari Jun, Kasahara Yasunori, Sakurai Yoriko, Iesato Ken, Nishimura Rintaro, West James, and Tatsumi Koichiro (2015) Influence of pulmonary emphysema on COPD assessment test-oriented categorization in GOLD document. International journal of chronic obstructive pulmonary disease 10, 1199-205

Svoboda M, Svancara J, Jarkovsky J, Novotna B, Popelkova P, Plutinsky M, Brat K, Zatloukal J, Volakova E, Hejduk K, Kocova E, Kopecky M, Neumannova K, Fecaninova M, Safranek P, Dvorak T, Sobotik O, Liptakova Z, Heribanova L, Snelerova B, Majerciakova M, Musilova P, and Koblizek V (2016) Risk scores for predicting death in COPD patients. European Respiratory Journal 48,

Swanney M P, Jensen R L, Crichton D A, Beckert L E, Cardno L A, and Crapo R O (2000) FEV(6) is an acceptable surrogate for FVC in the spirometric diagnosis of airway obstruction and restriction. American journal of respiratory and critical care medicine 162, 917-9

Tejero Elena, Prats Eva, Casitas Raquel, Galera Raul, Pardo Paloma, Gavilan Adelaida, Martinez-Ceron Elisabet, Cubillos-Zapata Carolina, Del Peso , Luis , and Garcia-Rio Francisco (2017) Classification of Airflow Limitation Based on z-Score Underestimates Mortality in Patients with Chronic Obstructive Pulmonary Disease. American journal of respiratory and critical care medicine 196, 298-305

Topalovic Marko, Exadaktylos Vasileios, Peeters Anneleen, Coolen Johan, Dewever Walter, Hemeryck Martijn, Slagmolen Pieter, Janssens Karl, Berckmans Daniel, Decramer Marc, and Janssens Wim (2013) Computer quantification of airway collapse on forced expiration to predict the presence of emphysema. Respiratory research 14, 131

Topalovic Marko, Laval Stefan, Aerts Jean-Marie, Troosters Thierry, Decramer Marc, Janssens Wim, Belgian Pulmonary Function Study, and investigators (2017) Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints. Respiration, and international review of thoracic diseases 93, 170-178

Tsoumakidou Maria, Tzanakis Nikolaos, Voulgaraki Olga, Mitrouska Ioanna, Chrysofakis Georgios, Samiou Maria, and Siafakas Nikolaos M (2004) Is there any correlation between the ATS, BTS, ERS and GOLD COPD's severity scales and the frequency of hospital admissions?. Respiratory medicine 98, 178-83

Tsushima K, Sone S, Fujimoto K, Kubo K, Morita S, Takegami M, and Fukuhara S (2010) Identification of occult parechymal disease such as emphysema or airway disease using screening computed tomography.. COPD 7(2), 117-25

Wei J, Xiong XF, Lin YH, Zheng BX, and Cheng DY (2015) Association between serum interleukin-6 concentrations and chronic obstructive pulmonary disease: a systematic review and meta-analysis.. PeerJ 3, e1199

Williams J E A, Green R H, Warrington V, Steiner M C, Morgan M D. L, and Singh S J ( ) Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index. Respiratory Medicine ,

Yamamoto Y, Yoshikawa M, Fuzita Y, Tomoda K, Yamauchi M, Koyama N, Fukuoka A, and Kimura H (2013) Mini nutritional assessment short form (MNA-SF) can predict exacerbation in COPD independently of COPD assessment test (CAT). Respirology 18, 183

Zhang Rongbao, Tan Xingyu, He Quanying, Chen Qing, Gai Jun, Wei Jing'an, and Wang Yan (2014) Comparison of symptom and risk assessment methods among patients with chronic obstructive pulmonary disease. Chinese medical journal 127, 2594-8

Zhu A, Ge D, Zhang J, Teng Y, Yuan C, Huang M, Adcock IM, Barnes PJ, and Yao X (2014) Sputum myeloperoxidase in chronic obstructive pulmonary disease.. European journal of medical research 19, 12