BMJ Open

Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis

Shamil Haroon, Rachel Jordan, Yemisi Takwoingi, Peymane Adab

ABSTRACT

Background: Chronic obstructive pulmonary disease (COPD) is widely underdiagnosed. A number of studies have evaluated the accuracy of screening tests for COPD, but their findings have not been formally summarised. We therefore sought to determine and compare the diagnostic accuracy of such screening tests in primary care.

Methods: Systematic review and meta-analysis of the diagnostic accuracy of screening tests for COPD confirmed by spirometry in primary care. We searched MEDLINE, EMBASE and other bibliographic databases from 1997 to 2013 for diagnostic accuracy studies that evaluated 1 or more index tests in primary care among individuals aged ≥35 years with no prior diagnosis of COPD. Bivariate meta-analysis of sensitivity and specificity was performed where appropriate. Methodological quality was assessed independently by 2 reviewers using the QUADAS-2 tool.

Results: 10 studies were included. 8 assessed screening questionnaires (the COPD Diagnostic Questionnaire (CDQ) was the most evaluated, n=4), 4 assessed handheld flow meters (eg, COPD-6) and 1 assessed their combination. Among ever smokers, the CDQ (score threshold ≥19.5; n=4) had a pooled sensitivity of 64.5% (95% CI 59.9% to 68.8%) and specificity of 65.2% (52.9% to 75.8%), and handheld flow meters (n=3) had a sensitivity of 79.9% (95% CI 74.2% to 84.7%) and specificity of 84.4% (68.9% to 93.0%). Inadequate blinding between index tests and spirometry was the main risk of bias.

Conclusions: Handheld flow meters demonstrated higher test accuracy than the CDQ for COPD screening in primary care. The choice of alternative screening tests within whole screening programmes should now be fully evaluated.

PROSPERO registration number: CRD42012002074.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death, ranks ninth for lost disability adjusted life years, and is an important cause of healthcare expenditure. Despite this, as much as 50–90% of the disease burden remains undiagnosed.

Patients often under-recognise the significance of respiratory symptoms, and clinicians frequently miss opportunities to diagnose COPD at primary care consultations. Early detection may offer opportunities to reduce disease progression and improve quality of life, for example, through smoking cessation interventions and pulmonary rehabilitation. An analysis of the Health Survey for England suggested that over three-quarters of symptomatic smokers identified with COPD through targeted case finding could benefit from recommended therapies, which could potentially prevent hospitalisations.

There is now a policy drive to identify undiagnosed COPD. However, a systematic review of population-based screening with spirometry concluded that this should not be recommended, partly because it estimated that hundreds of smokers would need to be screened to prevent a single COPD exacerbation. Furthermore, without considering clinical symptoms, this approach could identify individuals with airflow obstruction, who would not meet the clinical criteria for COPD according to current guidelines.

Strengths and limitations of this study

This is the first systematic review and meta-analysis of the diagnostic accuracy of screening tests for chronic obstructive pulmonary disease (COPD) in primary care.

Robust methods were used to identify, appraise and summarise the available literature.

There were few head-to-head comparisons of screening tests.

The definition of COPD used in the majority of included studies was physiological, based on the presence of airflow limitation, rather than clinical, requiring the presence of relevant symptoms.

Methodological limitations of included studies included inadequate reporting of blinding of operators performing and interpreting screening and reference tests (spirometry) and reporting of withdrawals and indeterminate results.

To cite: Haroon S, Jordan R, Takwoingi Y, et al. Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis. BMJ Open 2015;5:e008133. doi:10.1136/bmjopen-2015-008133
Recently, efforts to identify undiagnosed COPD have focused on the use of initial screening tests to identify those at high risk, prior to diagnostic spirometry.\textsuperscript{13} 14 Several approaches for initial screening have been evaluated, but their findings have not been systematically reviewed and quantitatively synthesised, and it is not yet clear which test or combination is the most accurate. Although one narrative review\textsuperscript{15} compared existing symptom-based questionnaires, it did not include other screening tests and needs updating.

We report a systematic review and meta-analysis of published studies that summarises and compares the accuracy of screening tests for COPD in primary care.

**METHODS**

**Protocol and registration**

The protocol for this review was previously published\textsuperscript{16} and registered.\textsuperscript{17}

**Eligibility criteria**

We sought diagnostic accuracy studies of any design that evaluated one or more index tests, were conducted in primary care (including general practices and community pharmacies) and recruited individuals aged \( \geq 40 \) years with no prior diagnosis of COPD. Index tests included screening questionnaires, handheld flow meters (eg, Piko-6 or COPD-6), peak flow meters, chest radiography, and risk prediction models or decision aids, either alone or in combination. We only included studies that specified the target condition as COPD, and used the presence of airflow obstruction, based on prebronchodilator or postbronchodilator spirometry as the reference standard (although postbronchodilator spirometry was considered the ideal reference standard).

**Outcomes**

The primary outcome was identification of COPD. The main measures of test accuracy examined were sensitivity and specificity.

**Search strategy**

We searched the following databases from March/April 2012 for the previous 15 years: MEDLINE, EMBASE, CINAHL, Cochrane Central Register of Controlled Trials and the Health Technology Database. We also performed an updated search on MEDLINE and EMBASE up to December 2013. Searches limited to the first 100 articles were also performed on Google Scholar, Turning Research into Practice, HTAi VORTAL and DogPile, and selected conference abstracts for the previous 2 years. Search terms are listed in online supplementary table S1 and included Medical Subject Heading terms and free-text synonyms for COPD, screening tests and measures of test accuracy, with no language restrictions.

**Study selection and data extraction**

Titles and abstracts were screened independently by two reviewers. Relevant full-text articles were independently assessed for eligibility by two reviewers and disagreements resolved through discussion. Prespecified data were extracted from full-text articles by one reviewer and verified by a second. We extracted the number of true positives, false positives, true negatives and false negatives for construction of two-by-two contingency tables. Where these data were not provided, reported measures of test accuracy were used to derive these values.

**Risk of bias assessment**

Included studies were assessed independently by two reviewers for risk of methodological bias and applicability concerns against criteria from the QUADAS-2 tool.\textsuperscript{18} Online supplementary table S2 shows how this was adapted for the review. Disagreements were resolved through discussion.

**Statistical analysis**

Forest plots of sensitivity and specificity were constructed using Review Manager (RevMan) V.5.2 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012). These plots were used to visually explore between-study variation in the diagnostic accuracy of each test. We also explored differences in population screened, screening test, diagnostic criteria and study design.

Where there was sufficient clinical and methodological homogeneity, we used the xtmelogit command in Stata V.13.1 (Stata-Corp, College Station, Texas, USA) to fit the bivariate model\textsuperscript{19} 20 to derive summary estimates of sensitivity and specificity and their 95% CIs. If there were fewer than four studies, we simplified the bivariate model to two univariate random effects logistic regression models for sensitivity and specificity by assuming no correlation between both measures.\textsuperscript{21} We used two approaches to compare the diagnostic accuracy of the screening tests. First, we used all relevant studies that evaluated one or more tests, and second we restricted the analysis to studies that made direct (head-to-head) comparisons. Where meta-analysis was possible, tests were compared by adding a covariate for test type to the bivariate model to assess whether average sensitivity and/or specificity differed between the tests. To assess the statistical significance of differences in sensitivity and specificity between tests, we compared the fit of alternative models (effect of adding or removing covariate terms from the model) by using likelihood ratio tests.

Positive and negative predictive values (PPV and NPV) were estimated from the sensitivity and specificity of each test, assuming a prevalence of undiagnosed COPD of 5.5%\textsuperscript{9} in a hypothetical population of 1000 patients aged \( \geq 40 \) years. We estimated the number-needed-to-screen to identify one individual with COPD as the total number screened divided by the number of true
positives, and the number of diagnostic assessments needed as the reciprocal of the PPV.

RESULTS
Study selection
The stages of study selection are shown in figure 1. After excluding duplicates, our search yielded 2605 records. From these, full-text articles were retrieved for 266 studies. Ten studies met the inclusion criteria, and five were suitable for meta-analysis (since these were sufficiently similar with respect to the included population, screening tests and definition of COPD). Figure 1 lists the reasons for excluding articles, the most common of which was the inclusion of patients with previously known COPD.

Study characteristics
Characteristics of included studies are summarised in tables 1 and 2 (see online supplementary tables S3–5 for details of each study). All were cross-sectional test accuracy studies, of which two used a paired design to compare two screening tests (screening questionnaires and handheld flow meters).22 23 Nine studies were multi-centre and all were based in general practices.

Recruitment and population selection
Four studies opportunistically recruited patients routinely attending primary care, three actively recruited participants through postal invitations or local advertisements, two used a combination of both strategies and one study did not report the method of recruitment.24 All studies specified age in the inclusion criteria with most requiring participants to be over 40 years. Seven studies also required a positive smoking history, but only one required participants to report respiratory symptoms as part of the entry criteria.13 The main exclusion criterion was an established history of lung disease.

Index and reference tests
All studies first applied one or more index tests to the eligible population and then performed the reference test (spirometry) on either all (n=8 studies) or a random sample25 26 (n=2) of participants. Index tests included screening questions or questionnaires (n=8) and handheld flow meters (n=4). One study also assessed the combined accuracy of using a screening questionnaire sequentially with a handheld flow meter.22 No studies evaluating other screening tests met the inclusion criteria.

Reference standard
Prebronchodilator and postbronchodilator spirometry was the reference standard in two25 27 and eight studies, respectively (tables 1 and 2). Most studies sufficiently described spirometry and quality control procedures. Spirometry was performed by trained technicians (n=4),
general practitioners (GPs; n=1), pulmonary physicians (n=1) and nurses (n=2), while quality control was usually performed by a respiratory specialist or physiologist who reviewed spirometry results.

**Methodological quality**
Most studies gave a clear description of participants, index and reference tests, and diagnostic criteria (see online supplementary figure S1 and table S6). However, there was often under-reporting of withdrawals (n=4), participant flow diagrams (n=5) and uninterpretable spirometry tests (n=5). The main risk of bias arose from inadequate blinding between index and reference tests (n=7; figures 2 and 3). There was also potential for bias in the flow and timing domain (n=5), where the number of participants undergoing index and reference tests was unclear, and where significant numbers of participants were excluded from the analysis.

**Screening questionnaires**
Alltogether four screening questionnaires were evaluated on a total of 9472 participants in eight studies (table 1), of which the COPD Diagnostic Questionnaire (CDQ), also referred to as the International Primary Airways Group (IPAG) Questionnaire, was the most widely validated (n=4). All instruments included questions related to the presence of respiratory symptoms (usually cough, dyspnoea and wheeze). Other items included in some, but not all questionnaires related to smoking history, allergies, age, body mass index (BMI).

---

**Table 1 Characteristics of studies evaluating screening questionnaires (8 studies)**

| Characteristic | Range/number of studies |
|---------------|-------------------------|
| Study designs | Cross-sectional test accuracy 8 |
| Participants | 237–3158 |
| Mean age (years) | 52.3–65.3 |
| Male (%) | 38.1–69.0 |
| Required smoking status | Only current/ex-smokers 5 |
| | Included never-smokers 3 |
| Required respiratory symptoms | 1 |
| Setting | General practice(s) 8 |
| Number of centres | 1–36 |
| Recruitment strategy | Multicentre 7 |
| | Single centre 1 |
| Questionnaires | COPD Diagnostic Questionnaire* 4 |
| | Lung Function Questionnaire 2 |
| Common items | Not named 2 |
| Reference test—spirometry | 6 |
| | Post-BD FEV1/FVC <0.7 7 |
| Entry criteria—spirometry | Other† 1 |
| | Included symptoms in definition of COPD 1 |
| Spirometry quality control | Yes 8 |
| Range of results | Sensitivity 57–93% |
| | Specificity 24–80% |
| | Severity of new COPD cases ≥80% 11–39% |
| | (FEV1 % predicted)‡ 50–80% 43–61% |
| | <50% 10–37% |

*Also referred to as the Respiratory Health Screening Questionnaire and the IPAG questionnaire.
†Pre-BD FEV1/FVC <88.5% predicted for men and FEV1/FVC <89.3% for women.
‡Based on five studies.
and physical functioning. Overall, participants were similar in age (range 52.3–65.3 years) but varied by sex (range 38–69% male).

COPD Diagnostic Questionnaire

Four studies that evaluated the CDQ in ever smokers were included in a meta-analysis. Using a score threshold of ≥19.5, the pooled sensitivity was 64.5% (95% CI 59.9% to 68.8%) and specificity 65.2% (95% CI 52.9% to 75.8%; table 3). With a prevalence of undiagnosed COPD of 5.5%, this gave a PPV of 9.7% (95% CI 6.9% to 14.2%), NPV of 96.9% (95% CI 95.8% to 97.7%), and would require 29 individuals (95% CI 27 to 31) to complete the questionnaire and 13 (95% CI 11 to 16) to undergo a diagnostic assessment for each new diagnosis.

All other questionnaires

There was considerable between-study heterogeneity in the design of other screening questionnaires, which precluded their meta-analysis. In these four studies, sensitivities ranged from 57% to 88% and specificities from 25% to 80% (figure 4).

Handheld flow meters

The test accuracy of handheld flow meters was evaluated in 1400 participants across four studies (table 2). Participants were similar in age (range 52–65.3 years) but varied by sex (range 43–99.7% male). Only one study included never-smokers and stratified the results by smoking status.

---

Table 2 Characteristics of studies evaluating handheld flow meters (4 studies)

| Characteristic | Range/number of studies |
|---------------|-------------------------|
| Study designs | Cross-sectional test accuracy study 4 |
| Participants  | 305–2464                |
| Mean age (years) | 52.0–65.3               |
| Male (%)     | 43.3–99.7               |
| Required smoking status | Only current/ex-smokers 3 |
| Required respiratory symptoms | Included never-smokers 1 |
| Setting       | General practice(s) 4   |
| Number of centres | 4–25                   |
| Recruitment strategy | Active 1, Opportunistic 2, Active and opportunistic 1 |
| Handheld flow meter | Piko-6 3, COPD-6 1 |
| Operator      | Nurse 2, GP 1, Not reported 1 |
| Use of BD     | Pre-BD 3, Post-BD 1    |
| Test threshold | FEV1/FEV6 < 0.7, 0.70–0.75 |
| Reference test—spirometry | Post-BD FEV1/FVC <0.7 |
| Definition of airflow obstruction | Post-BD |
| Spirometry quality control | Yes 2, No 1, Unclear 1 |
| Range of results | Sensitivity 79–86%, Specificity 71–99% |
| Severity of new COPD cases (FEV1 % predicted)3 | ≥80% 35–48%, 50–80% 48–65%, <50% 0–16% |

BD, bronchodilator; COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; GP, general practitioner.

Haroon S, et al. BMJ Open 2015;5:e008133. doi:10.1136/bmjopen-2015-008133
Handheld flow meters differ from diagnostic spirometers in that they are limited to measuring the forced expiratory volume in 1 and 6 s (FEV₁ and FEV₆, respectively), are usually performed with three blows and are cheaper and quicker to administer. They were used without a bronchodilator in three studies and were supervised by either trained nurses or GPs. A narrow range of thresholds were used to denote a positive test ranging from FEV₁/FEV₆ <0.7 to 0.75. Their sensitivity ranged from 79% to 86% and specificity from 71% to 99% (figure 4). Three studies enrolling ever smokers were similar enough to be included in a meta-analysis. The pooled sensitivity was 79.9% (95% CI 74.2% to 84.7%) and specificity was 84.4% (95% CI 68.9% to 93.0%). Using the same assumptions, this would require 23 individuals (95% CI 22 to 24) to be screened and 5 (95% CI 3 to 9) to undergo a diagnostic assessment to identify 1 individual with COPD (table 3).

Combination of tests
In the single study that reported the combined accuracy of a screening questionnaire (CDQ) with a handheld flow meter, the sensitivity was 74% (95% CI 64% to 83%) and specificity was 97% (95% CI 95% to 98%). This would reduce the need for diagnostic assessment to two individuals (95% CI 2 to 3) to identify one with COPD (table 3 and figure 5).

Comparison of test accuracy
In the first comparative analysis, based on an indirect comparison in ever smokers, there was evidence from the likelihood ratio tests that the CDQ at a score threshold of ≥19.5 had a lower sensitivity (p=0.003) but no difference in specificity (p=0.09) compared with handheld flow meters. In the second analysis at the lower score threshold of ≥16.5 (or 17), there was evidence to suggest a higher sensitivity (p=0.03) but a much lower specificity (p=0.01) than handheld flow meters. Two studies directly compared handheld flow meters and the CDQ and their findings were consistent with the results of the indirect comparison. Furthermore, Frith et al also reported both higher sensitivity and specificity of handheld flow meters compared with the CDQ at the score threshold of ≥19.5.

DISCUSSION
Summary of evidence
This review incorporated evidence on the test accuracy of questionnaires and handheld flow meters for COPD screening in primary care. The CDQ developed by Price et al was the most widely validated of the four screening questionnaires included. However, use of handheld flow meters under the supervision of trained health professionals was significantly more accurate than the CDQ for discriminating between ever smokers with and without airway obstruction, and a combination of both instruments may improve the accuracy still further, potentially reducing the number of diagnostic assessments required. Studies evaluating the CDQ and handheld flow meters had generally few methodological biases, the main being insufficient clarity on blinding between index and reference tests.

Unfortunately, only one study by Kotz et al considered the accuracy of a screening test (handheld flow meter) for identifying airflow obstruction in symptomatic patients, which is closer to identifying clinical COPD. The remainder evaluated the accuracy for identifying airflow obstruction without explicitly considering the presence of symptoms. Nevertheless, the results are still likely to apply since we observed that the test accuracy reported by Kotz et al was very similar to that reported by studies that did not explicitly consider respiratory symptoms.
Table 3  Summary estimates of the accuracy of each test for diagnosis of COPD in ever smokers

| Index test* | Studies | Cases/ participants | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | NNS (95% CI) | NND (95% CI) |
|-------------|---------|---------------------|----------------------|----------------------|--------------|--------------|--------------|--------------|
| CDQ (score ≥19.5) | 3 | 495/1703 | 64.5 (59.9 to 68.8) | 65.2 (52.9 to 75.8) | 9.7 (6.9 to 14.2) | 96.9 (95.8 to 97.7) | 29 (26 to 31) | 11 (7 to 15) |
| CDQ (score ≥16.5) | 4 | 580/2322 | 87.5 (83.1 to 90.9) | 78.8 (72.7 to 85.3) | 7.7 (6.3 to 9.8) | 98.2 (96.6 to 99.0) | 21 (20 to 22) | 13 (11 to 16) |
| Handheld flow meters | 3 | 224/1133 | 79.9 (74.2 to 84.7) | 84.4 (68.9 to 93.0) | 23.0 (12.2 to 41.3) | 98.6 (97.9 to 99.1) | 23 (22 to 24) | 5 (3 to 9) |
| CDQ and handheld flow | 1 | 90/624 | 74.4 (64.2 to 83.1) | 97.0 (95.2 to 98.3) | 59.1 (43.8 to 74.0) | 98.5 (97.9 to 99.0) | 25 (22 to 29) | 2 (2 to 3) |

The PPV, NPV, NNS and NND to identify one individual with COPD have been calculated assuming a prevalence of undiagnosed COPD of 5.5% in a theoretical population of 1000 people.

*Owing to the complexity of the bivariate model and the limited number of studies, only the four CDQ studies that used a score threshold ≥16.5 were pooled using a bivariate model. We carefully examined the parameter estimates of the model, especially the variances of the random effects, to check whether the model was reliable. There were only three studies of the CDQ that used a score threshold ≥19.5 and three studies of handheld flow meters. These were pooled using univariate random effects logistic regression models.
Weaknesses result mainly from the methodological limitations of included studies, particularly with respect to inadequate reporting of withdrawals and indeterminate results and blinding of operators performing and interpreting index and reference tests. This may have resulted in overestimation of test accuracy since positive index tests could plausibly influence performance and interpretation of reference spirometry. There was also a lack of head-to-head comparisons with only two studies evaluating more than one screening test. Indirect comparisons are potentially biased because of differences in population and study characteristics.

The criteria for airflow obstruction used in the included studies is also a point of contention given that using a fixed cut-off of FEV1/FVC < 0.7 may lead to over-diagnosis of the elderly. Future studies should therefore consider using a definition that accounts for age, sex and ethnicity biases, ideally using an FEV1/FVC ratio below the lower limit of normal and using the fixed ratio for sensitivity analyses.

Finally, the included studies did not report acceptability and uptake of screening tests, which are all important for evaluating their overall effectiveness. This review can therefore only be used to comment on test accuracy and not on comparative clinical and cost-effectiveness in routine practice, which ideally should be evaluated through head-to-head trials.

**Implications for research and practice**

Our findings suggest that handheld flow meters are likely to be more accurate than questionnaires for COPD screening in primary care. However, we also highlight several key limitations of previous studies. Future studies should provide clear descriptions of withdrawals, including participant flow diagrams, ensure that spirometry is performed without prior knowledge of index tests, and that indeterminate results, particularly with respect to spirometry, are reported. Future studies should also aim to recruit participants with no prior diagnosis of COPD (thus reducing the risk of spectrum bias) and use a clinical case definition, rather than just airway obstruction, in order to increase generalisability to real-life practice. More studies are needed to evaluate the accuracy and effectiveness of combining screening tests and to assess their cost-effectiveness. Finally, it remains unclear whether early detection of COPD significantly improves clinical outcomes and quality of life. This should first be demonstrated in prospective studies before firm recommendations are made.

**CONCLUSIONS**

Handheld flow meters used under the supervision of a trained health professional are more accurate than the CDQ for detecting spirometry-confirmed COPD in...
primary care. Limited evidence suggests that combining both tests may potentially improve test accuracy. Future studies should employ a case definition of COPD that aligns with current recommendations and include head-to-head comparisons.

Acknowledgements The authors would like to thank David Moore and Clare Davenport for their advice on the review protocol, Anne Fry-Smith for advice on the search strategy, Susan Bayliss (all from the University of Birmingham) for article translation, and Richard Riley (Keele University) for his advice on the statistical analysis.

Contributors SH designed the protocol, performed the literature search, selected articles, extracted data, assessed the quality of articles, synthesised the results and wrote the manuscript. RJ and PA identified the need and conceived the idea for a systematic review. RJ advised on the protocol, selected articles, verified the extracted data, assessed the quality of articles, advised on the data synthesis and revised the manuscript. YT synthesised the results (meta-analysis), advised on the risk of bias assessment and revised the manuscript. PA advised on the protocol, verified the extracted data, assessed the quality of articles, advised on the data synthesis and revised the manuscript. PA is the guarantor.

Funding This paper presents independent research funded by the National Institute for Health Research (NIHR). SH is funded by an NIHR doctoral fellowship (DRF-2011-04-064). RJ was funded by an NIHR postdoctoral fellowship (pdf/01/2008/023). YT is funded by an NIHR doctoral fellowship (DRF-2011-04-135).

Competing interests PA and RJ are both principal investigators on an NIHR programme grant for investigating COPD in primary care (grant reference number RP-PG-0109-10061).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/

REFERENCES

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2095–128.

2. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010:
a systematic analysis for the Global Burden of Disease Study 2010. 

Lancet 2012;380:2197–223.

3. Britton M. The burden of COPD in the U.K.: results from the Confronting COPD survey. Respir Med 2003;97(Suppl C):S71–9.

4. Soriani J, Zielinski J, Bruce D. Screening for and early detection of chronic obstructive pulmonary disease. Lancet 2009;374:721–32.

5. Fromer L. Diagnosing and treating COPD: understanding the challenges and finding solutions. Int J Gen Med 2011;4:729–39.

6. Jones RCM, Price D, Ryan O, et al. Opportunities to diagnose chronic obstructive pulmonary disease in routine care in the UK: a retrospective study of a clinical cohort. Lancet Respir Med 2014;2:267–76.

7. Scanlon PD, Connett JE, Waller LA, et al. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study. Am J Respir Crit Care Med 2000;161(2 Pt 1):381–90.

8. Chronic obstructive pulmonary disease: management of chronic obstructive pulmonary disease in adults in primary and secondary care. London: National Institute for Health and Clinical Excellence, 2010.

9. Jordan R, Lam KB, Cheng K, et al. Case finding for chronic obstructive pulmonary disease: a model for optimizing a targeted approach. Thorax 2010;65:492–8.

10. An outcomes strategy for people with chronic obstructive pulmonary disease (COPD) and asthma in England. London: Department of Health, 2011.

11. Global initiative for chronic obstructive lung disease: global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Global Initiative for Chronic Obstructive Lung Disease, 2013.

12. Lin K, Watkins B, Johnson T, et al. Force USPST. Screening for chronic obstructive pulmonary disease using spirometry: summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2008;148:535–43.

13. Kotz D, Neleman P, van Schayck CP, et al. External validation of a COPD diagnostic questionnaire. Eur Respir J 2008;31:298–303.

14. Thorn J, Tilling B, Lisspers K, et al. Improved prediction of COPD in at-risk patients using lung function pre-screening in primary care: a real-life study and cost-effectiveness analysis. Prim Care Res J 2012;21:159–66.

15. van Schayck CP, Hubert RJ, Nordyke RJ, et al. Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting. Respir Med 2005:10:323–33.

16. Haroon SM, Adap PA, Jordan RE. Case finding for COPD in primary care: a systematic review. Prim Care Res J 2012;21:354–7.

17. PROSPERO. International prospective register of systematic reviews. New York: Centre for Reviews and Dissemination, University of York, 2014.

18. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004;8:ii, 1–234.

19. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–90.

20. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol 2006;59:1331–2; author reply 2–3.

21. Takwoingi Y, Guo B, Riley RD, et al. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat Methods Med Res 2015.

22. Sichletidis L, Spyridatos D, Papaioannou M, et al. A combination of the IPAG questionnaire and PiKo-6(R) flow meter is a valuable screening tool for COPD in the primary care setting. Prim Care Respir J 2011;20:184–9, 1 p following 9.

23. Firth P, Crockett A, Beilby J, et al. Simplified COPD screening: validation of the PiKo-6(R) in primary care. Prim Care Respir J 2011;20:190–8, 2 p following 8.

24. McIntyre A, Wallis GM, Mannino DM, et al. Prevalence of airway obstruction assessed by lung function questionnaire. Mayo Clin Proc 2011;86:75–81.

25. Buffels J, Degryse J, Heyrman J, et al. Office spirometry significantly improves early detection of COPD in general practice: the DIDASCO study. Chest 2004;125:194–9.

26. Duong-Quy S, Hsu-Huy T, Mai-Huu-Thanh B, et al. [Early detection of smoking related chronic obstructive pulmonary disease in Vietnam]. Rev Mal Respir 2009;26:267–74.

27. Hanania NA, Mannino DM, Yawn BP, et al. Predicting risk of airflow obstruction in primary care: validation of the lung function questionnaire (LFQ). Respir Med 2010;104:1160–70.

28. Price DB, Tinkelman DG, Nordyke RJ, et al. Scoring system and clinical application of COPD diagnostic questionnaires. Chest 2006;129:1531–9.

29. Cartwright S. An evaluation of screening for COPD against the National Screening Committee criteria, 2012. Public Health England. http://www.screening.nhs.uk/copd

30. The UK NSC policy on chronic obstructive pulmonary disease. Public Health England, UK National Screening Committee, 2013.

31. Zhang JY, Huang TC, Chi W, et al. Should fev1/fev6 replace fev1/fvc ratio to detect airflow obstruction? A metaanalysis. Chest 2009;135:991–8.

32. Stanley AJ, Hasan I, Crockett AJ, et al. Validation of the COPD Diagnostic Questionnaire in an Australian general practice cohort: a cross-sectional study. Prim Care Respir J 2014;23:92–7.

33. van den Bernt L, Wouters BCW, Grootens J, et al. Diagnostic accuracy of bronchodilator FEV1/FEV6 from microspirometry to detect airflow obstruction in primary care: a randomised cross-sectional study. NPUJ Prim Care Res Med 2014;24.

34. Hardie JA, Buist AS, Vollmer WM, et al. Risk of over-diagnosis of COPD in asymptomatic elderly never-smokers. Eur Respir J 2002;20:1117–22.

35. Miller MR, Levy ML. Chronic obstructive pulmonary disease: missed diagnosis versus missed diagnosis. BMJ 2015;351:h3021.

36. Quanjer P, Stanojevic S, Cole T, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40:1324–43.

37. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in diagnostic testing. N Engl J Med 1978;299:926–30.

38. Freeman D, Nordyke RJ, Isonaka S, et al. Questions for COPD diagnostic screening in a primary care setting. Respir Med 2005;99:1311–18.
**Supplementary tables and figures**

**Table S1 Search terms**

| Population                                              | AND     | Index test               |
|---------------------------------------------------------|---------|--------------------------|
| Chronic obstructive pulmonary disease                   | OR      | Case finding             |
| OR                                                      | OR      | Screening                |
| Chronic obstructive airways disease                     | OR      | Early detection          |
| OR                                                      | OR      | Secondary prevention     |
| Chronic obstructive lung disease                        | OR      | Spirometry               |
| OR                                                      | OR      | Questionnaire            |
| COPD                                                    | OR      | Peak flow                |
| OR                                                      | OR      | Chest X-ray              |
| COAD                                                    | OR      | Decision aid             |
| OR                                                      | OR      | Algorithm                |
| Emphysema                                               | OR      | Sensitivity              |
| OR                                                      | OR      | Specificity              |
| Chronic bronchitis                                      | OR      |                          |
| OR                                                      | OR      |                          |
| Airflow obstruction                                     | OR      |                          |
| OR                                                      | OR      |                          |
| Airflow limitation                                      | OR      |                          |
| OR                                                      | OR      |                          |
Table S2 QUADAS-2 tool for assessing methodological quality and risk of bias of included studies

| Patient selection | Signalling question                                                                 | Signalling question                                                                 | Signalling question                                                                 | Risk of bias                                                                 | Concerns about applicability |
|-------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------|-------------------------------|
| Domain 1: Patient selection | Was a consecutive or random sample of patients enrolled? | Was a case-control design avoided? | Did the study avoid inappropriate exclusions? | Could the selection of patients have introduced bias? | Are there concerns that the included patients and setting do not match the review question? |
| Yes: If all consecutive or random samples of subjects were enrolled. | Yes: If the study was not a case control design. | Yes: If there were no inappropriate exclusion criteria. | Low risk: If all signalling questions answered 'yes.' | Low concern: If selected subjects matched the review question. | High or unclear risk: If 'no or unclear' was reported for at least one signalling question. |
| No: If subjects were non-randomly selected. | No: If the study had a case control design. | No: If subjects were excluded based on inappropriate criteria such as presence of depression. | High or unclear risk: If 'no or unclear' was reported for at least one signalling question. | High concern: If selected subjects differed from those in the review question. | |
| Unclear: If sampling method was unclear. | Unclear: If the study design was unclear. | Unclear: If selection criteria were unclear. | Low risk: If all signalling questions answered 'yes.' | Low concern: If selected subjects matched the review question. | Unclear concern: If there was insufficient information on included subjects and setting. |

Domain 2: Index test

| Index test | Signalling question | Signalling question | Signalling question | Risk of bias | Concerns about applicability |
|------------|---------------------|---------------------|---------------------|--------------|-------------------------------|
| Were the index test results interpreted without knowledge of the results of the reference standard? | If a threshold was used, was it pre-specified? | Could the conduct or interpretation of the index test have introduced bias? | Are there concerns that the index test, its conduct, or interpretation differ from the review question? |
| Yes: If the index test results were interpreted without knowledge of the spirometry results. | Yes: If the threshold for a positive test result was pre-specified. | Low risk: If all signalling questions answered 'yes.' | Low concern: If the index test was performed as described in the review question. |
| No: If the index test results were interpreted with knowledge of the spirometry results. | No: If the threshold for a positive test result was not pre-specified. | High or unclear risk: If 'no' was reported for at least one signalling question. | High concern: If the index test differed from those specified in the review. |
| Unclear: If this was | | | | | |
| Domain 3: Reference standard | Domain 4: Flow and timing |
|-----------------------------|--------------------------|
| **Reference standard** | **Flow and timing** |
| Is the reference standard likely to correctly classify the target condition? | Was there an appropriate interval between the index test and reference standard? |
| Yes: If quality controlled spirometry was used. | Yes: If the time between the index and reference tests were less than six months. |
| No: If spirometry was performed without adequate quality control. | No: If the time between the index and reference tests were greater than six months. |
| Unclear: If it was unclear from the report whether spirometry quality control procedures had been implemented. | Unclear: If the time between the index and reference tests were more than six months. |

| Signalling question | Signalling question | Signalling question | Risk of bias | Concerns about applicability |
|-------------------|-------------------|-------------------|-------------|-----------------------------|
| results. Unclear: If it was unclear whether index test results were interpreted independently of spirometry results. | unclear from the report. | Could the reference standard, its conduct, or its interpretation have introduced bias? | Low risk: If all signalling questions answered 'yes.' High or unclear risk: If 'no' was reported for at least one signalling question. | Are there concerns that the target condition as defined by the reference standard does not match the review question? |
| | | | Low concern: If quality controlled spirometry was used. High concern: If quality controlled spirometry was not used. Unclear concern: If insufficient information was provided in the report on spirometry quality control. | |
| | | | | |
| | | | | |

| Yes: If spirometry results were interpreted without knowledge of the results of the index test. | Yes: If all eligible subjects recruited to the study with index test results were included in the analysis. | Low risk: If all signalling questions answered 'yes.' High or unclear risk: If 'no' was reported for at least one signalling question. | | |
| No: If spirometry results were interpreted with knowledge of the index test results. | No: If not all eligible subjects recruited to the study with index test results were included in the analysis. | | | |
| Unclear: If this was not clear from the report. | | | | |
| | | | | |
| | | | | |
| | | | | |

| Yes: If quality controlled spirometry was used. | Yes: If all eligible subjects received spirometry. | Low risk: If all signalling questions answered 'yes.' High or unclear risk: If 'no' was reported for at least one signalling question. | Low concern: If quality controlled spirometry was used. High concern: If quality controlled spirometry was not used. Unclear concern: If insufficient information was provided in the report on spirometry quality control. | |
| Signalling question                                           | Signalling question                                           | Signalling question                                           | Risk of bias                                                                 | Concerns about applicability |
|--------------------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------------------------------|------------------------------|
| index and reference tests were longer than six months.       | subjects received the reference standard.                    | No: If not all recruited subjects with index test results were included in the analysis. | ‘no’ was reported for at least one signalling question.                      |                              |
| Unclear: If this was unclear from the report.                | Unclear: If this was not clear from the report.              | Unclear: If this was unclear from the report.                |                                                                              |                              |
|                                                              |                                                              |                                                              |                                                                              |                              |
| Study          | Country   | Setting                  | Recruitment method                                                                 | Eligibility criteria                                                                 | Index and reference tests                                                                 | Definition of COPD                                                                 |
|---------------|-----------|--------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Buffels 2004  | Belgium   | 20 general practitioners | Invited patients routinely attending general practice over a 12 week period in 1999. | Inclusion criteria: Age 35-70 years Exclusion criteria: Receiving bronchodilators and/or inhaled corticosteroids | Index test: Screening questionnaire Reference test: Pre-BD spirometry in all subjects with respiratory symptoms and 10% sample of asymptomatic subjects | Pre-BD FEV₁/FVC < 88.5% predicted for men & FEV₁/FVC < 89.3% for women |
| Duong-Quy 2009 | Vietnam   | 12 primary care medical centres in one city | Broadcast an advertisement on the local television daily for one week. A recruitment company was used to help with participant recruitment (details not reported). Eligible subjects expressing an interest in participating were advised to attend one of the 12 primary care centres from January 2007 to February 2008. | Inclusion criteria: Active and former smokers with >10 pack-years and aged >40 years Exclusion criteria: Previously diagnosed respiratory disease (asthma, COPD and tuberculosis) | Index test: Pre-BD handheld flow meter (Piko-6®) Reference test: Full medical assessment including clinical examination, pulmonary radiology, ECG, and post-BD spirometry for those who had an index FEV₁/FEV₆ < 0.7 and a sample of those with FEV₁/FEV₆ ≥ 0.7 | Post-BD FEV₁/FVC < 0.7 with <200mL or 12% reversibility |
| Freeman 2005  | UK        | One general practice     | Postal invitation from October 1997 to April 2002.                                    | Inclusion criteria: Age ≥40 years & current/ex-smoker & had either received respiratory medications in the preceding 2 years or had a history of asthma Exclusion criteria: None | Index test: Screening questions Reference test: Pre-/post-BD spirometry on all subjects | Post-BD FEV₁/FVC < 0.7 and lack of reversibility (reversibility defined as increase in FEV₁ of 200mL and 15% from pre-BD FEV₁ (not clear if all were post-BD) |
| Study | Country | Setting | Recruitment method                                                                 | Eligibility criteria                                                                 | Index and reference tests                                                                 | Definition of COPD tests |
|-------|---------|---------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|--------------------------|
| Frith 2011 | Australia | 4 primary care practices | Recruited during routine practice visits, invitation to study days, and local newspaper advertisement between August and December 2006. | Inclusion criteria: Age ≥50 years & current/ex-smoker & no prior diagnosis of obstructive lung disease (COPD, emphysema, chronic bronchitis, asthma) & no treatment for obstructive lung disease in past 12 months | Index test: Pre-BD handheld flow meter (Piko-6®) & screening questionnaire (COPD Diagnostic Questionnaire) | Post-BD FEV₁/FVC<0.7 |
|        |         |         |                                                                                     | Exclusion criteria: Refusal or inability to give consent, pre-existing non-obstructive lung disease, symptoms suggestive of unstable heart disease, and spirometry contraindications | Reference test: Pre-/ post-BD spirometry on all patients |                                                                                       |
| Hanania 2010 | US | Two family physician group offices | Invited patients aged ≥40 years visiting the practices from March-May 2008 | Inclusion criteria: Age ≥40 years | Index test: Screening questionnaire (Lung Function Questionnaire) | Pre-BD FEV₁/FVC<0.7 |
|        |         |         |                                                                                     | Exclusion criteria: None                                                                 | Reference test: Pre-BD spirometry |                                                                                       |
| Study    | Country | Setting                        | Recruitment method                                                                                                                                                                                                 | Eligibility criteria                                                                                                                                                                                                 | Index and reference tests                                                                 | Definition of COPD |
|----------|---------|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|-------------------|
| Kotz 2008 | Netherlands | General population and primary care practices | Advertisements in a local newspaper, flyers, posters and mailings to households and invitation during primary care consultations from Jan 2005-Dec 2006.                                                                 | Inclusion criteria: Age 40-70 years & current smoker with ≥10 pack years & motivated to stop smoking & able to read and speak Dutch & reporting a respiratory symptom (cough, phlegm or dyspnoea) | Index test: Questionnaire (COPD Diagnostic Questionnaire) | Post-BD FEV1/FVC<0.7 |
|          |         |                                |                                                                                                                                                                                                                     | Exclusion criteria: Prior respiratory diagnosis, spirometry in previous 12 months or contraindications to smoking cessation therapy                                                                                           | Reference test: Pre-/post-BD spirometry in all participants                               |                   |
| Mintz 2011 | US | 36 primary care centres | NR                                                                                                                                                                                                               | Inclusion criteria: Age ≥30 years old & current/ex- smoker with ≥10 pack years                                                                                                                                   | Index test: Screening questionnaire (Lung Function Questionnaire)                          | LFQ≤18 & post-BD FEV1/FVC<0.7 |
|          |         |                                |                                                                                                                                                                                                                     | Exclusion criteria: Regular use of respiratory medications within 4 weeks of the study, known diagnosis of substantial lung conditions with regular use of respiratory medications. | Reference test: Pre-/ post-BD spirometry                                                   |                   |
| Study     | Country   | Setting                  | Recruitment method      | Eligibility criteria                                                                                                                                                                                                 | Index and reference tests                                                                                      | Definition of COPD tests                                                                                     |
|-----------|-----------|--------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Price 2006 | UK & US   | 2 primary care practices | Postal invitation       | Inclusion criteria: Age ≥40 years & current/ex-smoker  
Exclusion criteria: Refusal to consent, history of non-obstructive lung disease, use of respiratory medications in past year, acute symptoms of unstable heart disease | Index test: Screening questionnaire (COPD Diagnostic Questionnaire)  
Reference test: Pre-/post-BD spirometry  
Post-BD FEV₁/FVC<0.7 | Post-BD FEV₁/FVC<0.7 |
| Sichletidis 2011 | Greece | 25 general practices | Invited first 50 patients meeting the inclusion criteria who visited each participating GP from 1st March-31st May 2009. | Inclusion criteria: Age >40 years  
Exclusion criteria: Confirmed diagnosis of lung disease, thoracic surgery in previous 6 months, acute respiratory infection, uncontrolled cardiac disease, or could not perform acceptable spirometry | Index tests:  
1. Screening questionnaire (International Primary Airways Group Questionnaire, also known as the COPD Diagnostic Questionnaire)  
2. Post-BD handheld flow meter (Piko-6®) (Bronchodilator=400µg salbutamol)  
Reference test: Pre-/post-BD spirometry  
Post-BD FEV₁/FVC<0.7 | Post-BD FEV₁/FVC<0.7 |
| Thorn 2012 | Sweden    | 21 primary healthcare centres | Invited patients attending participating primary healthcare centres over a 5 month period. | Inclusion criteria: Age 45-85 years & current/ex-smoker with ≥15 pack years  
Exclusion criteria: None | Index test: Pre-BD handheld flow meter (COPD-6)  
Reference test: Pre-/post-BD spirometry  
Post-BD FEV₁/FVC<0.7 | Post-BD FEV₁/FVC<0.7 |

BD=bronchodilator, FEV₁=forced expiratory volume in one second, FEV₆=forced expiratory volume in 6 seconds, FVC=forced vital capacity, NR=not reported
| Study     | Population       | Screening questionnaire | Spirometry (reference test)                                                                 | Number screened | New COPD cases |
|----------|------------------|-------------------------|-----------------------------------------------------------------------------------------------|-----------------|----------------|
| Buffels 2004 | Eligible: 3158 Invited: 3158 Attended: 3158 | - Cough >2 weeks  
- Dyspnoea during mild exercise/at night  
- Nasal allergy/hay fever  
- Visit to doctor for wheeze or chronic cough | Device: Spirobank spirometer with Winspiro software  
Bronchodilator: None  
Operator: GPs who had received 12 hours of training  
Standard: NR  
Quality control: Technical support was provided to GPs throughout the study. Accuracy of GP-performed spirometry was compared to that from a lab technician. | Index test Total: 3158 Positive: 728 | Subjects with positive index test: 126/703 (17.9%) |
| Freeman 2005 | Eligible: 1195 Invited: 1195 Attended: 624 | - Age  
- Smoking status  
- Pack-years  
- Cough  
- Dyspnoea  
- Wheeze | Device: Micro-Med handheld spirometer with Spida software  
Bronchodilator: 5mg salbutamol for those with prior respiratory medication or history of asthma or FEV1<80% predicted  
Operator: Trained respiratory nurse  
Standard: ATS standards. Minimum of 3 tests or until reproducibility within 5%.  
Quality control: All spirometry results were reviewed by a physician to ensure compliance with ATS standards. | Index test Total: 369 Positive: 121* (multiple response questionnaire), 142* (binary response questionnaire) | Subjects with positive index test: 62/369 (16.8%) |

*NR = not reported.
| Study     | Population | Screening questionnaire | Spirometry (reference test) | Number screened | New COPD cases |
|-----------|------------|-------------------------|----------------------------|----------------|----------------|
| Frith 2011 | Eligible: 233 Invited: 237 Attended: 237 | COPD diagnostic questionnaire (CDQ) | Device: EasyOne spirometer (ndd Medical) | Index test | 57/204 (27.9%) |
|           | Data on subjects with acceptable spirometry Mean age: 61 Male: 69% Smoking status: Current: 45% Former: 55% Never: <1% | Items: See Price 2006 (below) Thresholds: Score ≥19.5, ≥16.5 | Bronchodilator: 360mcg salbutamol | Total: 233 Positive: 110* (threshold ≥19.5), 165* (threshold ≥16.5) | FEV1, % predicted |
|           | | | Operator: trained operators using ATS/ERS guidelines | Reference test | ≥80%: 19 (33.3%) |
|           | | | Standard: ATS/ERS standards. At least 3 adequate baseline and post-BD FVC manoeuvres performed. | spirometry | 50-80%: 35 (61.4%) |
|           | | | Quality control: spirometry quality monitored by a respiratory physiologist blinded to the questionnaire and Piko-6® results. | Total: NR | 30-50%: 3 (5.3%) |
|           | | | | Acceptable quality: 204 | <30%: 0 |
|           | | | | | |
| Hanania 2010 | Eligible: NR Invited: NR Attended: 937 | Lung Function Questionnaire (LFQ) | Device: EasyOne spirometer (ndd Medical) | Index test | 156/837 (18.6%) |
|           | Data on subjects with acceptable spirometry and adequate data Mean age: NR Male: 38.1% Smoking status: NR | Items: | Bronchodilator: None | | FEV1 % predicted |
|           | | • Age • Cough • Wheeze • Dyspnoea • Smoking | Operator: NR | | ≥80%: 17 (11.5%) |
|           | | Threshold: Score ≤18 | Standard: NR | | 50-80%: 76 (51.4%) |
|           | | | Quality control: Investigators rated spirometry quality based on reliability and reproducibility. Only included traces considered reliable. | | 30-50%: 44 (29.7%) |
|           | | | | | <30%: 11 (7.4%) |
|           | | | | | (NB. Reported numbers do not add up to 156) |
| Study | Population | Screeninng questionnaire | Spirometry (reference test) | Number screened | New COPD cases |
|-------|------------|--------------------------|-----------------------------|----------------|---------------|
| Kotz 2008 | Eligible: 1052  Invited: 1052  Attended: 826  Data on subjects with spirometry  Mean age:52.3  Male:58.7%  Smoking status  Current: 100% | COPD Diagnostic Questionnaire (CDQ)  Items: See Price 2006 (below)  Thresholds: Score ≥19.5, ≥16.5 | Device: Vitalograph 2120  Bronchodilator: 500 µg terbutaline  Operator: Two qualified research assistants under the supervision of a pulmonologist  Standard: ATS/ERS standards  Quality control: spirometry performed according to ATS/ERS standards. All spirometry test results were validated by a pulmonologist and specialised lung function laboratory assistant not involved in the trial-both were blinded to the questionnaire scores. | Index test  Total: 1052  Analysed: 676  Positive: 549* (threshold ≥16.5) 366* (threshold ≥19.5) | 278/676 (41.1%) |
| Mintz 2011 | Eligible: 1724  Invited: 4956  Attended: 2284  Data on subjects who completed index test  Mean age: 53.9*  Male: 51.2%  Smoking status  Current: 57.6%  Former: 42.4% | Lung Function Questionnaire (LFO)  Items: Age  Cough  Wheeze  Dyspnoea  Smoking  Activity limitation  Threshold Score ≤18 | Device: Biomedical Systems, St Louis, MO  Bronchodilator: 360µg albuterol  Operator: Trained site staff  Standard: ATS standards  Quality control: Only data collected from acceptable spirometric manoeuvres were included. Patients producing unacceptable spirometry were allowed to repeat this within 7 days of the study visit. | Index test  Total: 1575  Positive: 1216 | 162/713 (22.7%) (NB. restricted to subjects ≥40 years) | FEV1 % predicted  ≥80%: 142 (51.1%)  50-80%: 119 (42.8%)  <50%: 17 (6.1%) |
| Study      | Population | Screening questionnaire | Spirometry (reference test) | Number screened | New COPD cases |
|------------|------------|-------------------------|-----------------------------|----------------|---------------|
| Price 2006 | Eligible: NR Invited: 17,361 Attended: 898 | COPD Diagnostic Questionnaire | Device: EasyOne spirometer (ndd Medical) Bronchodilator: 2.5mg salbutamol/albuterol Operator: NR Standard: ATS standards Quality control: Principal investigators conducted blinded review of all spirometry loops. A pulmonologist not associated with the study reviewed all loops on which there was disagreement | Index test Total: 898 Positive: 267* (threshold ≥16.5) 446* (threshold ≥19.5) Reference test (spirometry) Total: 898 Acceptable quality: 818 572 (70%) used for questionnaire development, 246 (30%) used for validation | 155/818 (18.9%) |
|            | Data on subjects with acceptable spirometry Mean age: 58.2 Male: 49.3% | **Items:*** Age Pack-years Weather-affected cough Productive phlegm in absence of a cold Early morning cough Wheeze Allergies **Thresholds** Score ≥19.5, ≥16.5 | **Device:** Vitalograph **Bronchodilator:** 400µg salbutamol **Operator:** Pulmonary specialists **Standard:** ATS/ERS standards **Quality control:** Spirometry performed and interpreted by pulmonary specialists according to ATS/ERS standards | **Index test Total:** 1250 **Positive:** 409* (smokers) 594* (smokers & non-smokers) **Reference test (spirometry) Total:** NR **Acceptable quality:** 1078 | Ever smokers: 90/624 (14.4%) Ever smokers & non-smokers: 111/1078 (10.3%) **FEV₁ % predicted ≥80%:40 (36.0%) 50-80%:53 (47.7%) 30-50%:16 (14.4%) <30%:2 (1.8%) |
| Sichletidis 2011 | Eligible: 1250 Invited: 1250 Attended: 1250 | COPD Diagnostic Questionnaire (also referred to as International Primary Airways Group questionnaire) **Items:** See Price 2006 (above) **Threshold** Score ≥17 | **Device:** Vitalograph **Bronchodilator:** 400µg salbutamol **Operator:** Pulmonary specialists **Standard:** ATS/ERS standards **Quality control:** Spirometry performed and interpreted by pulmonary specialists according to ATS/ERS standards | **Index test Total:** 1250 **Positive:** 409* (smokers) 594* (smokers & non-smokers) **Reference test (spirometry) Total:** NR **Acceptable quality:** 1078 | Ever smokers: 90/624 (14.4%) Ever smokers & non-smokers: 111/1078 (10.3%) **FEV₁ % predicted ≥80%:40 (36.0%) 50-80%:53 (47.7%) 30-50%:16 (14.4%) <30%:2 (1.8%) |
|            | Data on subjects with acceptable spirometry Mean age: 65.3 Male: 57.1% | **Smoking status** Ever: 48.8% Never: 51.2% | **BD=bronchodilator, NR=not reported, FEV₁=forced expiratory volume in 1 second, FVC=forced vital capacity, FEV₆=forced expiratory volume in 6 seconds** | **Derived values (may differ from reported test performance)** | |

* Derived values (may differ from reported test performance)
| Study     | Recruited population | Handheld flow meter | Spirometry (reference test) | Number screened | New COPD cases |
|-----------|----------------------|---------------------|-----------------------------|-----------------|----------------|
| Duong-Quy 2009 | Eligible: 2464 | Pre-BD Piko-6® | Device: SpiroLab II | Index test | Subjects with positive index test: 136/144 (94.4%) |
|           | Invited: NR       | Operator: NR       | Bronchodilator: short-acting β2 agonist (unspecified) | Total: 2464 | Subjects with negative index test: 3/123 (2.4%) |
|           | Attended: 2464   | 3 manoeuvres were | Operator: NR | Reference test (spirometry) | FEV₁ % predicted |
|           |                     | taken and the best of 3 selected. | Standard: Required at least 3 measures and at least 2 within 150mL to ATS/ERS standards. | Total: 144 subjects with positive index test and 123 with negative index test. | <80%: 65 (47.8%) |
|           |                     | All measures where FEV₁/FEV₆>1 were excluded. | Quality control: NR | Acceptable quality: NR | 50-79%: 63 (46.3%) |
|           |                     | Threshold FEV₁/FEV₆<0.7 | | | 30-49%: 8 (5.9%) |
|           |                     | | | | <30%: 0 |
| Frith 2011       | Eligible: 233     | Pre-BD Piko-6® | Device: EasyOne spirometer (ndd Medical) | Index test | 57/204 (27.9%) |
|           | Invited: 237      | Operator: Study nurse or GP | Bronchodilator: 360mcg salbutamol | Total: 233 | FEV₁ % predicted >80%: 19 (33.3%) |
|           | Attended:237   | Threshold FEV₁/FEV₆<0.75 (optimal cut-point) | Operator: trained operators using ATS/ERS guidelines | Positive: 101* | 50-80%: 35 (61.4%) |
|           |                     | | Standard: ATS/ERS standards. At least 3 adequate baseline and post-BD FVC manoeuvres performed. | Reference test (spirometry) | 30-50%: 3 (5.3%) |
|           |                     | | Quality control: spirometry quality monitored by a respiratory physiologist blinded to the questionnaire and Piko-6® results. | Total: NR | <30%: 0 |
| Study      | Recruited population | Handheld flow meter | Spirometry (reference test) | Number screened | New COPD cases |
|------------|----------------------|---------------------|------------------------------|-----------------|----------------|
| Sichletidis 2011 | Eligible: 1250 Invited: 1250 Attended: 1250 | Post-BD Piko-6® Bronchodilator: 400µg salbutamol Operator: GPs with 2 hours training Threshold Post-BD FEV₁/FEV₆<0.7 | Device: Vitalograph Bronchodilator: 400µg salbutamol Operator: Pulmonary specialists Standard: ATS/ERS standards Quality control: Spirometry performed and interpreted by pulmonary specialists according to ATS/ERS standards | Index test Total: 1250 Positive #: 104* (ever smokers) 137* (ever smokers & non-smokers) | Ever smokers: 90/624 (14.4%) Ever smokers & non-smokers: 111/1078 (10.3%) |
|             | Data on subjects with acceptable spirometry Mean age: 65.3 Male: 57.1% Smoking status Ever: 48.8% Never: 51.2% | | | Reference test (spirometry) Total: NR Acceptable quality: 1078 |
| Thorn 2012 | Eligible: NR Invited: NR Attended: 305 | Pre-BD COPD 6® Operator: Nurses Threshold FEV₁/FVC<0.73 | Device: NR Bronchodilator: 0.5mg terbutaline Operator: Nurses Standard: ATS standards Quality control: Spirometry performed according to ATS standards. No other quality control measures reported. | Index test Total: 305 Positive: 106* | 77/305 (25.2%) |
|             | Data on subjects who performed the index and reference tests Mean age: 61.2 Male: 43.3% Smoking status Ever: 100% | | | Reference test (spirometry) Total:305 Acceptable quality: NR |

*Derived values (may differ from reported test performance)

83* (smokers) and 109* (smokers & non-smokers) positive index tests when using a combination of the CDQ and handheld flow meter

BD=bronchodilator, NR=not reported, FEV₁=forced expiratory volume in 1 second, FVC=forced vital capacity, FEV₆=forced expiratory volume in 6 seconds
| Quality of reporting                                                                 | Buffels 2004 | Duong-Guy 2009 | Freeman 2005 | Frith 2011 | Hanania 2010 | Kotz 2008 | Mintz 2011 | Price 2006 | Sichletidis 2011 | Thorn 2012 |
|------------------------------------------------------------------------------------|--------------|----------------|--------------|------------|--------------|-----------|------------|------------|------------------|-------------|
| Clear description of recruitment                                                  | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Clear description of participants                                                  | N            | Y              | Y            | Y          | Y            | Y         | Y          | N          | Y                | Y           |
| Clear description of withdrawals                                                  | Y            | Y              | N            | N          | U            | Y         | Y          | Y          | Y                | N           |
| Participant flow diagram                                                          | Y            | N              | N            | Y          | N            | Y         | Y          | Y          | N                | N           |
| Spirometry quality control                                                        | Y            | N              | Y            | Y          | Y            | Y         | Y          | Y          | U                | Y           |
| Standard diagnostic criteria                                                       | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Representative spectrum of patients                                                | U            | N              | Y            | Y          | U            | Y         | Y          | Y          | Y                | Y           |
| Clear description of selection criteria                                            | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Spirometry as reference standard                                                   | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Spirometry performed within six months of index test                              | Y            | U              | Y            | Y          | U            | Y         | Y          | Y          | Y                | U           |
| All or random selection of participants underwent spirometry                       | Y            | N              | Y            | Y          | Y            | Y         | N          | Y          | Y                | Y           |
| Spirometry performed and interpreted independently of screening test result       | U            | U              | U            | Y          | U            | Y         | U          | Y          | U                | U           |
| Screening test performed and interpreted independently of spirometry              | Y            | Y              | U            | Y          | U            | Y         | U          | Y          | Y                | Y           |
| Intervention described in sufficient detail to permit its replication              | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Clinical data available representative of routine practice                         | Y            | Y              | Y            | Y          | Y            | Y         | Y          | Y          | Y                | Y           |
| Uninterpretable, indeterminate or intermediate results reported                    | N            | N              | N            | Y          | N            | Y         | Y          | Y          | Y                | N           |

Y=yes, N=no, U=unclear
Figure S1 Quality of reporting