BMJ Open

Accuracy of different screening tests and their combinations for undiagnosed COPD among primary care patients in China: a screening test accuracy study. Findings from the Breathe Well group

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-051811
Article Type:	Original research
Date Submitted by the Author:	02-Apr-2021
Complete List of Authors:	Pan, Zihan; Peking University First Hospital; Peking University Third Hospital, Pulmonary and Critical Care Medicine Dickens, Andrew; University of Birmingham, Institute of Applied Health Research Chi, Chunhua; Peking University First Hospital, General Practice Department, Peking University First Hospital, Beijing, 100034, China Kong, Xia; Peking University First Hospital, General Practice Enocson, Alexandra; University of Birmingham, Applied Health Research Cooper, Brendan; Lung Investigation Unit, University Hospitals Birmingham NHS Foundation Trust Adab, Peymane; University of Birmingham, Public Health Cheng, Kar Keung; University of Birmingham, Department of Public Health and Epidemiology Sitch, Alice; University of Birmingham, Public Health Jowett, Sue; University of Birmingham, Health Economics Unit Adams, Rachel; University of Birmingham, Institute of Applied Health Research Correia-de-Sousa, Jaime; Life and Health Sciences Research Institute (ICVS)/3B's — PT Government Associate Laboratory, University of Minho; Horizonte Family Health Unit, Farley, Amanda; University of Birmingham, Gale, Nicola K.; Univ Birmingham, Health Services Management Centre Jolly, Kate; University of Birmingham Maglakelidze, Mariam; Georgian Respiratory Association Maglakelidze, Tamaz; Georgian Respiratory Association Martins, Sonia; ABC Medical School, Family Medicine Stavrikj, Katarina; Medical Faculty, Centre for Family Medicine Stelmach, Rafael ; University of São Paulo, (5) Pulmonary Department, Heart Institute (InCor), School of Medicine Turner, Alice; University of Birmingham, Institute of Inflammation & Aging; Heart of England NHS Foundation Trust Williams, Sian; National Heart and Lung Institute Jordan, Rachel; University of Birmingham, Public Health and Epidemiology
Keywords:	GENERAL MEDICINE (see Internal Medicine), RESPIRATORY MEDICINE (see Thoracic Medicine), Chronic airways disease < THORACIC MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page

Accuracy of different screening tests and their combinations for undiagnosed COPD among primary care patients in China: a screening test accuracy study. Findings from the Breathe Well group

†Zihan Pan[1,2]; †Andrew P Dickens[3]; *Chunhua Chi[1]; Xia Kong[1]; Alexandra Enocson[3]; Brendan G Cooper[3,4]; *Peymané Adab[3]; Kar Keung Cheng[3,5]; Alice Sitch[3,6]; Sue Jowett[3], Rachel Adams[3], Jaime Correia-de-Sousa[7,8], Amanda Farley[3], Nicola Gale[9], Kate Jolly[3], Mariam Maglakelidze[10,11], Tamaz Maglakelidze[10], Sonia M Martins[12], Katarina Stavrikj[13], Rafael Stelmach[14], Alice M Turner[15], Siân Williams[7], Rachel E Jordan[3].

†Joint first authors

*Joint corresponding authors:

Name: Chunhua Chi

Postal address: Peking University First Hospital, No.8 XiShiKu Street, Xicheng District, Beijing, 100034, China

E-mail: chichunhua2012@qq.com

Telephone: +86 13910987530,

Fax numbers: +86 010-66158996

Name: Peymane Adab

Postal address: Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK B15 2TT

Email: p.adab@bham.ac.uk

Telephone: 0121 4143777

Co-authors details:

1. Department of General Practice, Peking University First Hospital, Beijing, China
2. Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China
3. Institute of Applied Health Research, University of Birmingham, Birmingham, UK.

4. University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital Birmingham, Birmingham, UK
5. General Practice Development and Research Centre, Peking University Health Science Centre, Beijing, China
6. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, UK
7. International Primary Care Respiratory Group
8. Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga Portugal. ICVS/3B's, PT Government Associate Laboratory, Braga/Guimarães, Portugal
9. Health Services Management Centre, School of Social Policy, College of Social Sciences, University of Birmingham, Birmingham, UK
10. Georgian Respiratory Association, Georgia
11. Petre Shotadze Tbilisi Medical Academy, Georgia
12. Family Medicine, ABC Medical School, Sao Paolo, Brazil
13. Center for Family Medicine, Faculty of Medicine, Ss.Cyril and Methodius University in Skopje, North Macedonia
14. Pulmonary Division, Heart Institute (InCor), Hospital das Clinicas da Faculdade de Medicina da Uinversidade de Sao Paulo, Brazil
15. Institute of Inflammation and Ageing, University of Birmingham, UK

Key word: COPD; screening test accuracy; screening strategies, health economics; primary care; multicentre study

Word count: 3200 words

Abstract

Objectives: To examine the accuracy and cost-effectiveness of various screening tests and combinations within a Chinese primary care population.

Design Screening test accuracy study

Setting: Urban and rural community health centres in four municipalities of China: Beijing (north), Chengdu (southwest), Guangzhou (south) and Shenyang (northeast).

Participants: Community dwelling residents aged 40 years and above who attended community health centres for any reason were invited to participate. 2445 participants (mean age 59.8 [SD 9.6] years, 39.1% [n=956] male) completed the study (February-December 2019), 68.9% (n=1684) were never-smokers and 3.6% (n=88) had an existing COPD diagnosis. 13.7% (n=333) of participants had spirometry-confirmed airflow obstruction.

Interventions: Participants completed six index tests (screening questionnaires [CDQ, CAPTURE, Chinese Symptom-based questionnaire or C-SBQ, COPD-SQ], microspirometry [COPD-6], peak flow [USPE]) and the reference test (ndd Easy On-PC).

Primary and secondary outcomes: Cases were defined as those with FEV_1/FVC below the lower limit of normal (LLN-GLI) on the reference test. Performance of individual screening tests and their combinations was evaluated, with cost-effectiveness analyses providing cost per additional true case detected.

Results: Airflow measurement devices (sensitivities 64.9% and 67.3%, specificities 89.7% and 82.6% for microspirometry and peak flow respectively) generally performed better than questionnaires, the most accurate of which was C-SBQ (sensitivity 63.1% [95% CI 57.6%, 68.3%], specificity 74.2% [95% CI 72.3%, 76.1%]). The combination of C-SBQ and microspirometry used in parallel maximised sensitivity (81.4%) and had specificity of 68%, with an incremental cost-effectiveness ratio of £64.20 (CNY385) per additional case detected compared with peak flow.

Conclusions: Simple screening tests to identify undiagnosed COPD within the primary care setting in China is possible, and a combination of C-SBQ and microspirometry is the most sensitive. Further work is required to explore optimal cut-points and effectiveness of programme implementation.

Trial registration: ISRCTN13357135

Article summary

Strengths and limitations of this study

- This is the first study assessing the accuracy of individual screening tools and their combinations to identify undiagnosed COPD within Chinese community populations.
- Defining airflow obstruction according to the lower limit of normal increased the likelihood that identified cases were true COPD.
- Recruiting participants from both urban and rural community hospitals maximised the generalisability of our findings to primary care patients.
- This study did not explore optimal cut-points for index tests, thus further work is required.

Introduction

Chronic obstructive pulmonary disease (COPD) is a common long-term condition characterized by persistent respiratory symptoms and airflow limitation[1]. Nearly one-third of the 3.2 million annual global deaths from COPD are from China[2, 3] where COPD ranks among the top three

leading causes of death with associated direct medical costs of 118% of local average annual income[4]. COPD develops slowly, resulting in delays in symptom recognition and high rates of underdiagnosis. 90% of the estimated 100 million people with COPD in China are undiagnosed; slightly higher than the 60-80% underdiagnosis rate worldwide[5-9]. Symptom reporting and recognition are lower in China, with 60% of diagnosed patients not reporting symptoms such as cough, expectoration and wheeze[10].

While COPD screening programmes are not currently endorsed in the United States and UK[11-13], considering the high proportion and heavy burden of undiagnosed disease[4], early identification is being prioritised in China. National policies recommend screening for undiagnosed COPD[14], but do not specify which screening tests to use. Furthermore, though spirometry is required for clinical diagnosis[1], it is not widely available in primary care settings in China. Therefore screening could reduce the numbers needing spirometry referral.

Globally, various COPD screening tools have been developed, including questionnaires and airflow measurement devices[15-17]. However, accuracy studies were mainly conducted in Western countries and have not been validated in a Chinese population where the distribution and underlying causes of undiagnosed COPD may differ. Furthermore, the majority of Chinese studies have used secondary or tertiary care COPD populations rather than people from community settings[18, 19]. Finally, the cost-effectiveness of different screening tests have not been previously estimated in China; a crucial consideration given the high prevalence of COPD in this middle-income country.

We examined the accuracy and cost-effectiveness of various screening tests and combinations within a Chinese primary care population.

Methods

Study design and participants

We conducted a cross-sectional, multicentre study to evaluate the accuracy and cost-effectiveness of various COPD screening tests and test combinations in primary care in China. Full details of participant recruitment and study assessments are described in the published protocol[20].

Participants were recruited from one urban and one rural community health centre (CHC) in

each of four municipalities: Beijing (North China), Chengdu (southwest China), Guangzhou (south China) and Shenyang (northeast China) (Figure 1). Between February-December 2019, community dwelling residents aged 40 years and above who attended CHCs for any reason were invited to participate, either directly by the attending clinician, or through poster or social media (WeChat) advertisements. Participants who were unable to give informed consent, had contraindications for spirometry or unable to perform the test for other reasons were excluded.

Eligible participants provided informed consent at the start of the assessment visit, prior to height and weight measurement and completion of all index and reference tests. Participants also completed a study questionnaire concerning demographics, smoking status, exposures, medical diagnoses, respiratory symptoms[21] and quality of life[22]. Data were entered into a secure online REDCap database[23, 24].

Participants with airflow obstruction on the reference test were offered health education, smoking cessation advice, influenza vaccination and inhalers if relevant, or referred to tertiary hospitals for further treatment including pharmacotherapy or pulmonary rehabilitation.

Study assessment

Index tests

The six index tests included four screening questionnaires: COPD Diagnostic Questionnaire (CDQ, cut-point ≥20)[16, 25], CAPTURE (cut-point ≥2)[26], COPD Screening Questionnaire (COPD-SQ, cut-point ≥16)[19] and, the Chinese symptom-based questionnaire (C-SBQ, cut-point ≥17)[18] and two airflow measurement devices: microspirometry (Vitalograph COPD-6, cut-point for positive test FEV_1/FEV_6 <0.78)[27, 28], peak flow (USPE, cut-point <350 l/min men, <250 l/min women)[26]. Questionnaires were selected to maximize symptom capture and minimize item duplication, whilst allowing comparison of the most relevant questionnaires (Appendix 1). Previously defined cut-points were used to identify participants at risk of COPD.

Trained researchers provided instructions before participants performed 3 pre-bronchodilator manoeuvres with each airflow measurement device. The order of administering peak flow or microspirometry alternated by participant, and the best FEV_1 and FEV_6 measure for each device were used for analyses, irrespective of which attempt they came from.

Participants completed the four screening questionnaires immediately after administration of a bronchodilator (400ug, Salbutamol). Questionnaires were intended to be self-completed, although researchers were available to assist if needed.

Reference test

The reference test was quality diagnostic spirometry (ndd Easy On-PC), performed 20-60 minutes after bronchodilation. Spirometry was administered by a second researcher not involved in the index tests and blind to their results. Participants performed a minimum of 3 blows, and a maximum of 6, until repeatability within 100mls or 5% [29]. Flow volume curves were classified according to the ATS/ERS[29] criteria. Tests with at least 3 curves meeting these criteria, were "Good." "Acceptable" tests contained at least one curve which concurred with the criteria, allowing accurate assessment of FEV_1. If accurate assessment was not possible the curves were classified as "unacceptable", and the test was excluded from analysis. All traces were over-read for quality by one of three independent respiratory experts and graded according to standard criteria[29], without knowledge of the index test results.

Airflow obstruction was defined as post-bronchodilator FEV_1/FVC ratio below the lower limit of normal (LLN) using the GLI equations.

Sample size

The Alonzo method[30] for paired test accuracy studies was used to calculate the sample size. Assuming independence of tests and prevalence of 12%, we required 1622 participants to detect a difference in sensitivity of 10% (95% vs 85%[16, 26, 31, 32]) with 90% power. With lower test sensitivity (90% vs 80%) 2279 participants are needed to detect this difference with 90% power.

Statistical analysis

The diagnostic performance of each index test was investigated by presenting 2x2 tables and calculating the sensitivity, specificity, positive predictive value and negative predictive value with 95% confidence intervals. Comparative test accuracy was assessed by calculating the difference in sensitivity and specificity, presenting 95% confidence intervals and using McNemar's test.

The primary analysis compared the sensitivity and specificity between the CAPTURE screening questionnaire and the peak flow meter. The comparison was specified a priori as CAPTURE was

rigorously developed, accounted for exposures other than smoking and was intended for use in conjunction with peak flow. Secondary analyses evaluated the comparative performance of all other individual index tests, as well as plausible combination test strategies. Test strategies were formed using two pre-determined combinations for appropriate pairs of individual index tests (questionnaires and lung function tests); firstly, to maximise sensitivity, where a participant with a positive result for either index test would be positive for the strategy (parallel testing strategy) and secondly, to maximise specificity, where a participant would need a positive result on both index tests to be positive for the strategy (serial testing strategy).

All analyses were conducted in Stata v15.

Economic analysis

We conducted a cost-effectiveness analysis to calculate the cost per additional case detected for all tests and combination strategies. The strategies were ordered by the number of true cases detected, from least to greatest, and the principle of dominance was applied to eliminate redundant strategies (where they were more costly and less effective). Each test was then compared with the next best alternative. For the purpose of this paper, the individual index tests and the combination strategy with the highest sensitivity were compared.

The unit costs and quantity of any equipment, medication and consumables required, staff time (and salary costs) to deliver each individual test and use of facilities were determined to calculate the health care costs of delivering each screening test/strategy. Each individual test was timed at a sample of assessment clinics to estimate an overall mean time and range for each test. Equipment costs were depreciated (at 3.5% a year) over the estimated lifespan of the equipment (ranging from 1 to 6 years). Cost per patient visit was calculated assuming the equipment would be used for 12,000 patients per clinic per year. It was also assumed that positive cases would be confirmed with quality diagnostic spirometry (assuming 4000 patients/year). Costs were calculated in UK£ for a price year of 2019, and converted to Chinese Yuan (¥) using Purchasing Power Parities (PPP[33]) with a conversion rate of 6.0 (Appendix 2).

The paper follows the STARD guidance[34] for reporting studies of diagnostic accuracy.

Results

Sample

6198 eligible people were invited to the study. A total of 2560 (41.3%) consented, of whom 25 withdrew and 90 were excluded from analysis (86 because of incorrect inhaler technique, four had unusable spirometry data). 2445 participants with complete data on all index and reference test were included in the final analysis (Figure 2). Approximately two thirds (68.0%) were recruited through their attending clinician, 24.5% via advertisements and 7.5% through word of mouth.

The mean age of participants was 59.8 (SD 9.6), 39.1% (n=956) were male, two thirds (n=1684, 68.9%) were never smokers and over half lived in an urban area (1338, 54.7%). 46.7% had no diagnosed conditions (n=1142); the most common diagnosed condition was hypertension (n=842, 34.4%), 3.6% (n=88) had an existing COPD diagnosis and 8.4% (n=205) had an existing chronic bronchitis/emphysema diagnosis (Table 1). 99.8% of participants had an acceptable usable spirometry (with 63.3% (n=1547) defined as good). 13.6% (n=333) of participants had spirometry-confirmed airflow obstruction using the LLN criteria, of whom 175 (52.5%) had moderate to severe obstruction i.e. GOLD stage II or above [1]. Respiratory symptoms of wheeze, productive cough or breathlessness (mMRC≥2) were reported by 52.9% of those with airflow obstruction (66.3% of those who were GOLD stage II or above), and 25.1% of those without. Amongst participants with no previously reported COPD diagnosis, the prevalence of obstruction was 9.9% (n=218), of whom 89 (40.8%) were GOLD stage II or above. Using the $FEV_1/FVC<0.7$ criteria[1], 17.4% (n=425) of all participants had airflow obstruction.

Performance of individual tests and screening strategies

Among the screening questionnaires, the C-SBQ had the highest sensitivity in detecting airflow obstruction at 63.1% (57.6%, 68.3%), CAPTURE the lowest sensitivity (51.7% [46.1, 57.1]), with CDQ (55.0% [49.4%, 60.4%]) similar to COPD-SQ [55.3% (49.7%, 60.7%)]. The CDQ had the highest specificity (78.6% [76.8%, 80.4%]) (Table 2). CAPTURE compared to CDQ had the most obvious difference in specificity of 8.4% (-10.7, -6.0; p<0.001) (Table 4).

Both peak flow and microspirometry devices had higher sensitivity and specificity compared to all questionnaires (Table 3). Peak flow had the highest sensitivity (67.3%) and microspirometry the

highest specificity (89.7%) (Table 2).

Of the combined screening strategies, C-SBQ combined with airflow measurement devices in parallel (i.e. recorded as screen-positive if either test was positive) had the best performance, with sensitivities of 80.5%-81.4%, and specificities of 65.5%-68%. Parallel strategies (requiring either test to be positive) optimised sensitivity and serial strategies (requiring both tests to be positive) optimised specificity. Taking CAPTURE and peak flow as an example, a parallel combination had sensitivity of 77.2% compared to 41.7% in serial combination, while the specificity significantly increased from 59.1% to 93.7% (Table 2).

Overall, test performance was slightly higher when screening questionnaires were combined with microspirometry rather than peak flow. Strategies including CAPTURE performed less well compared to those based on other questionnaires. Parallel strategies including the C-SBQ had the highest sensitivities, whereas those based on the CDQ had the highest specificity (Table 2, Table 3). Full comparisons of serial and parallel strategies are described in Appendix 3.

Cost-effectiveness of preferred screening tests

Analysis of the C-SBQ parallel strategies revealed that the most costly strategy was the combination of C-SBQ and microspirometry, but this also detected the most true cases (Table 5). The C-SBQ alone was dominated by microspirometry (more costly, less effective). The incremental cost-effectiveness ratio (ICER) for C-SBQ and microspirometry (versus peak flow) was greatest at £64.20 (CNY 385.20), but could be considered cost-effective if the threshold willingness to pay for an additional true case detected in China is at least CNY 385.

Discussion

This is the first study assessing the accuracy of individual screening tools and their combinations to identify undiagnosed COPD within Chinese community populations. We showed that the combination of a simple questionnaire and airflow measurement device could adequately identify adults requiring diagnostic spirometry. Our overall findings were consistent with a meta-analysis of studies from other countries[35], that airflow measurement devices were more accurate than questionnaires, and that combinations of screening tests improved ability to detect COPD in

primary care. Within single test strategies, microspirometry had the best performance (sensitivity 64.9%, specificity 89.7%). For combination strategies, the C-SBQ and microspirometry used in parallel, maximised sensitivity (81.4%) with reasonable specificity (68%) and would be deemed cost-effective if the Chinese health service was willing to pay ≥CNY 385 per additional case detected.

C-SBQ had the highest sensitivity of all screening questionnaires in our study, with comparable specificity. However, accuracy of the C-SBQ was worse than reported in the validation paper of the Chinese tool, with lower sensitivity (63.1% vs 82.5%) but slightly higher specificity (74.2% vs 72.9%). The observed discrepancy may be due to differences in the spectrum of clinical characteristics[36] (community sample rather than tertiary care population in previous study) and airflow obstruction criteria used (we used the lower limit of normal rather than the GOLD criteria).

Inclusion of the C-SBQ and the CDQ from which it was derived allowed direct comparison of the two measures, confirming that C-SBQ was more accurate for use in Chinese community populations when prioritising sensitivity (sensitivity 63.1% vs 55.0% with slightly lower specificity 74.2% vs 78.6%).

Direct comparison between our findings and those of previous studies was limited by differences in populations and pre-test probabilities. COPD among never smokers is more common in China than in western countries and we included never smokers in this study to maximise the range of potential cases. Inevitably this contributed to the lower test performance observed. Furthermore, the CAPTURE questionnaire was originally designed to detect more severe COPD. The different case definition in our study therefore precludes direct comparison with previous studies (we plan to report accuracy for detecting more severe clinically significant COPD in a future publication).

Our test accuracy study has highlighted the strengths of different screening tests, which can be used to evaluate future screening programmes. We recruited a large number of participants from urban and rural settings in four geographically diverse municipalities in China, and the proportion of never smokers in our sample (68.9%) was comparable to that found in a recent nationally representative cross-sectional study in China (71.4%)[10]. We demonstrated that lung function tests and diagnosis of COPD can be implemented by GPs and nurses after a structured training course with regular quality over reading and feedback, as evidenced by 99% usable spirometry and

consistently good quality spirometry in most GP sites. The fully paired study design enabled us to compare the accuracy of multiple index tests and strategies. Alternating the order of peak flow and microspirometry tests during assessments decreased the potential training effect that could have been introduced when conducting consecutive lung function tests in a research context.

We defined the reference test as airflow obstruction regardless of clinical symptoms, to reflect the methods of previous studies and also account for the differing symptom profile reported among Chinese populations, where chronic respiratory symptoms are less recognised. In our study, just over half of those with obstruction were likely to benefit from some treatment due to reported symptoms, and a further quarter of those obstructed would benefit from smoking cessation advice as they had a positive smoking history but no respiratory symptoms.

Accuracy might have differed if the GOLD criteria were used, though unlikely to substantially change the comparative performance of the tests. Defining airflow obstruction according to the LLN criteria increased the likelihood that participants testing positive on study spirometry were true COPD cases, rather than detecting comorbidities with similar clinical presentations such as cardiovascular disease[37]. As pre-bronchodilator spirometry was omitted from the study assessment to minimise participant burden and increase uptake in this large community-based study, we could not assess airflow reversibility.

Chinese community health centres do not have COPD registers and it was therefore not possible to exclude diagnosed COPD patients from this study. However, as the aim of our study was to determine accuracy of different screening tests by comparing all tests against a reference standard, rather than to evaluate the implementation of a screening programme, inclusion of COPD patients was justified. By including some people with known COPD, we maximised the number of test positives in the study sample.

Although China has recently introduced a national policy of COPD screening, there is no current guidance regarding the tests to use or which test characteristics (i.e. sensitivity / specificity) to prioritise. Considering the estimated high prevalence of undiagnosed COPD in China, highly sensitive strategies may be preferred to maximise the number of detected cases, although this would result in large numbers being referred for diagnostic spirometry, many of whom would be false positives. However, the potential inefficiency may be offset by a recent policy to include spirometry in routine primary care health consultations; avoiding the need to refer patients to

hospital for diagnostic assessment.

If the strategy of C-SBQ and microspirometry were used in practice and had the same accuracy as reported here, it is likely that true COPD cases who were not detected (false negatives) would have mild disease and would re-attend with recurring symptoms, offering further opportunities for referral to diagnostic spirometry.

While our analyses used recommended cut-points for the index tests, it is important to explore their optimal cut-points when applied in this context, as many tests were developed with alternate purposes and/or populations in mind. Thresholds used to indicate airflow obstruction (either in the screening tests or reference test) may not be valid in the whole Chinese population as adequate reference values for lung function are currently unreliable.

Although we have determined the accuracy of different tests when used for screening Chinese community populations for undiagnosed COPD, we did not evaluate the implementation of a screening programme. It is important to undertake a trial to compare the effectiveness and cost-effectiveness of the most efficient screening strategy identified in this study (maximising yield with acceptable false positive rate) against usual care on yield and clinical outcomes. Such a trial would need to assess uptake of screening and incorporate pathways for clinical assessment and subsequent treatment for test positive cases. In our study sample >75% had potential to benefit; >half with obstruction had treatable symptoms and a further quarter with obstruction and no symptoms would benefit from smoking cessation advice. We presented cost per additional true case detected, however no country has, to date, stated a willingness to pay threshold for this outcome. The quality-adjusted life year (QALY) is a more common metric in health economic analyses, with established cost per QALY thresholds. Although outside the remit of our test accuracy study, future work should attempt to extrapolate cases detected to the management of patients with COPD, to assess the impact on quality of life and survival to allow the calculation of QALYs.

In conclusion, we have demonstrated that within the primary care setting in China, the most efficient screening test strategy was a combination of the C-SBQ and microspirometry where a positive test in either would result in a referral for diagnostic spirometry. Further work is required to explore optimal cut-points and there is a need for a clinical trial to evaluate whether a screening programme using this test combination is clinically and cost-effective.

Contributors

Zihan Pan and Andrew P Dickens wrote the manuscript with input from all other authors. Rachel E Jordan led the design of the trial, with contributions and advice from all other authors. Chunhua Chi, Xia Kong, Peymane Adab, KK Cheng contributed to decisions on outcome measures. Chunhua Chi and KK Cheng advised on involving GP practices, Rachel E Jordan, Peymane Adab, Alexandra Enocson, Brendan Cooper and Andrew P Dickens advised on lung function testing. Andrew P Dickens and Rachel E Jordan designed the intervention. Alice Sitch and Sue Jowett designed the analysis plan and economic evaluation. Zihan Pan did the statistical analysis, supported by Alice Sitch, Sue Jowett and Andrew P Dickens. All authors contributed to acquisition, analysis, or interpretation of data. Chunhua Chi was the local PI. All authors revised the manuscript and approved the final version before submission.

We would also like to acknowledge Radmila Ristovska (1955-2020), also involved in the initiation of this study.

Acknowledgement

The investigator and collaborative team include: R Jordan, P Adab (CIs), Z Pan, AP Dickens, C Chi, X Kong, A Enocson, BG Cooper, KK Cheng, A Sitch, S Jowett, R Adams, J Correia de Sousa, A Farley, N Gale, K Jolly, M Maglakelidze, T Maglakelidze, SM Martins, K Stavrikj, R Stelmach, A Turner and S Williams. We thank the dedicated team of researchers at Department of General Practice, Peking University First Hospital for managing and co-ordinating the project. We are also grateful for support from our Trial Steering Committee (Dr. Semira Manaseki-Holland [Chair], Prof. David Mannino, Mr. Zhiwei Zhang [patient], Dr. Xueying Zhou [clinician], Prof. Chunhua Chi, Dr. Rachel Jordan), and) and The International Scientific Advisory Committee (Prof. Debbie Jarvis, Dr. Semira Manaseki-Holland, Prof. David Mannino, Prof. Niels Chevannes). We especially want to thank the patients included in the study and the clinicians at all community hospitals for their assistance, without whom this project would not have been possible.

The clinicians involved in this study included: Wenyan Duan, Mengmeng Guo, Zhi Jin (Caiyu Community Hospital, Beijing); Yuling Li, Shuo Chen, Kang Rong, Yi Cai, Jing Bai (Xinjiekou Community Hospital, Beijing); Wei Xiong, Honglin Luo, Juan Chen, Na Li (Hezuo Community Hospital, Chengdu); Min Chen, Yaling Luo, Xiaolong Zhou (Nanxin Community Hospital, Chengdu),

Shixing Liu, Guoxiong Han, Shuhua Sun, Jian Long Li, Jianlong Cai, Haimei Chen, Yuanping Chen (Shayuan Community Hospital, Guangzhou); Bin Hu, Huilian Xuan, Jieya Zhang, Huanping Chen, Fengqi Liu, Haixia Yin (Jiulong Community Hospital, Guangzhou); Ying Chen, Xinzheng Cui, Yao Chen, Hui Peng, Weihan Jin, Xiaoguang Xu (Beishi Community Hospital, Shenyang); Zhenguo Gao, Zhongning Kan, Lina Song, Yongsheng Tong, Yuxuan Zhang, Shuai Mu, Yang Gao (Taoxian Community Hospital, Shenyang); Zhennan Qi, Hui Pang, Anny Gao, Xue Jin (Peking University First Hospital).

We thank the patient advisory group for their useful comments, including Mr. Zhiwei Zhang, Mr. Chen Wang, Mr Baoyin Fang and Mr. Deqian Ma.

We thank the support of the Health Committees in Xicheng District, Beijing and in Guangzhou for identifying potential research sites.

We gratefully acknowledge International Primary Care Respiratory Group (IPCRG) for introducing us to the primary care networks involved in this study and for its continued facilitation of clinical engagement.

We obtained appropriate permissions to use the Symptom Based Questionnaire, COPD Screening Questionnaire, COPD Diagnostic Questionnaire and CAPTURE.

Ethics approval

The study has been approved by Peking University First Hospital (2018-R-141, PUFH) and University of Birmingham (ERN_18-1177, UoB).

Patient and public involvement

The research team conducted a research prioritization exercise with patients, clinicians and policy makers, and the need to identify effective screening strategies for undiagnosed COPD was one of the research areas prioritized. The patient advisory group advised on the format of study material prior to recruitment commencing. All stakeholders involved in this exercise received study updates twice a year, and were kept informed of findings and consulted at the end of the study regarding implications for practice and policy decisions, as well as advice on appropriate dissemination of study findings.

In addition, the study has a Trial Steering Committee (TSC) that meets regularly and comprises

various independent members, including a patient and a clinician representative as well as international experts in respiratory research and several members of the study research team.

Serious adverse events (SAE)

No SAE from performing the index tests or the reference test in the study.

Registration number and name of registry

The protocol for this study was previously published and registered on ISRCTN registry. The number was ISRCTN13357135 and the full study protocol can be accessed at *http://www.isrctn.com* (ISRCTN13357135).

Funding

This research was funded by the National Institute for Health Research (NIHR) NIHR global group on global COPD in primary care, University of Birmingham, (project reference: 16/137/95) using UK aid from the UK Government to support global health research. This paper presents independent research supported by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. The funder of the study had no role in study design, data collection, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

The authors declare no conflicts of interest.

Additional file

Appendix 1. Screening questionnaires

Appendix 2. Costs, timings and assumptions for case-finding strategies

Appendix 3. Comparisons of serial and parallel strategies

Figure and table legends

Figure 1 Map of Breathe Well-China research sites

Figure 2 Study flow chart

Table 1 Characteristics of study participants

Table 2 Accuracy of Index tests and strategies

Table 3 Comparative sensitivity for individual tests

Table 4 Comparative Specificity for individual tests

Table 5 Per patient cost, effectiveness and cost-effectiveness of selected screening strategies

TABLE 1 Characteristics of study participants

Characteristic	Total sample (n=2445)	Reference test positive (n=333)	Reference test negative (n=2112)
Male sex, n (%)	956 (39.1%)	199 (59.8%)	757 (35.8%)
Age in yrs ; mean(SD)	59.8 (9.6)	63.5 (8.9)	59.2 (9.6)
BMI; mean (SD)	24.9 (3.5)	24.3 (3.4)	25.0 (3.4)
Education, n (%)			
High school or below	1879 (76.9)	277 (83.2%)	1602 (75.9%)
Above High school	566 (23.1)	56 (16.8%)	510 (24.1%)
Employment status, n(%)			
Employed	674 (27.6%)	54 (16.2%)	620 (29.4%)
Unemployed	665 (27.2%)	98 (29.4%)	567 (26.9%)
Retired	1106 (45.2%)	181 (54.4%)	925 (43.8%)
Living area, n(%)			
Urban	1338 (54.7%)	174 (52.3%)	1164 (55.1%)
Smoking status, n(%)			
Current smoker	472 (19.3%)	113 (33.9%)	359 (17.0%)
Ex-smoker	289 (11.8%)	72 (21.6%)	217 (10.3%)
Never smoker	1684 (68.9%)	148 (44.5%)	1536 (72.7%)
Male	--	27 (18.2%)	--
Female	--	121 (81.8%)	--
Pack y.rs mean (SD)	9.0 (17.8)	18.0 (21.0)	7.6 (16.8)
Health in general, n(%)			
Very Good-good	1255 (51.3%)	127 (38.1%)	1128 (53.4%)
Fair-very bad	1190 (48.7%)	206 (61.9%)	984 (46.6%)
Diagnosed conditions, n(%)			
COPD	88 (3.6%)	64 (19.2%)	24 (1.1%)
Chronic bronchitis/emphysema	205 (8.4%)	93 (27.9%)	112 (5.3%)
Asthma	105 (4.3%)	48 (14.4%)	57 (2.7%)
Tuberculosis	41 (1.7%)	12 (3.6%)	29 (1.4%)
Hypertension	842 (34.4%)	119 (35.7%)	723 (34.2%)
Diabetes Mellitus	330 (13.5%)	43 (12.9%)	287 (13.6%)
Heart disease	274 (11.2%)	43 (12.9%)	231 (10.9%)
Other	269 (11.0%)	31 (9.3%)	238 (11.3%)
None of the above	1142 (46.7%)	106 (31.8%)	1036 (49.1%)
Symptoms, n(%)			
At least occasional wheeze	322 (13.2)	110 (33.0)	212 (10.0)
Productive cough	457 (18.7)	117 (35.1)	340 (16.1)
mMRC, n(%)			

Grade 0-1	2222 (90.9%)	257 (77.2%)	1965 (93.0%)
Grade 2-4	223 (9.1%)	76 (22.8%)	147 (7.0%)
CAT, mean(SD)	6.1 (5.4%)	8.9 (6.9%)	5.6 (4.9%)
Bronchitis, pneumonia or severe whooping cough in childhood	169 (6.9%)	38 (11.4%)	131 (6.2%)
Tuberculosis in childhood	45 (1.8%)	11 (3.3%)	34 (1.6%)
Exposure to pollutants*, n (%)			
Current/past exposure	2256 (92.3%)	307 (92.2%)	1949 (92.3%)
Never	189 (7.7%)	26 (7.8%)	163 (7.7%)
Year(s) of exposure, mean (SD)	8.9 (6.4)	9.1 (6.6)	8.8 (6.4)
GOLD stage if <LLN†, n (%)			
I (FEV_1 ≥80% predicted)	--	158 (47. 5%)	--
II (FEV_1 50-79% predicted)	--	137 (41.1%)	--
III (FEV_1 30-49% predicted)	--	33 (9.9%)	--
IV (FEV_1 <30% predicted)	--	5 (1.5%)	--

* cooking fumes, biomass smoking, gas, steams, dust

† LLN = lower limit of normal

TABLE 2 Accuracy of Index tests and strategies

Part 1	Part 2	Strategy type	TP*	FP*	TN*	FN*	Sensitivity% (95% CI)	Specificity% (95% CI)	PPV%* (95% CI)	NPV%* (95% CI)
CAPTURE	n/a	Individual	172	628	1484	161	51.7 (46.1, 57.1)	70.3 (68.3, 72.2)	21.5 (18.7, 24.5)	90.2 (88.7, 91.6)
CDQ	n/a	Individual	183	451	1661	150	55.0 (49.4, 60.4)	78.6 (76.8, 80.4)	28.9 (25.4, 32.6)	91.7 (90.4, 92.9)
C-SBQ	n/a	Individual	210	545	1567	123	63.1 (57.6, 68.3)	74.2 (72.3, 76.1)	27.8 (24.6, 31.2)	92.7 (91.4, 3.9)
COPD-SQ	n/a	Individual	184	479	1633	149	55.3 (49.7, 60.7)	77.3 (75.5, 79.1)	27.8 (24.4, 31.3)	91.6 (90.3, 92.9)
Peak flow	n/a	Individual	224	368	1744	109	67.3 (61.9, 72.3)	82.6 (80.9, 84.2)	37.8 (33.9, 41.9)	94.1 (92.9, 95.1)
Microspirometry	n/a	Individual	216	217	1895	117	64.9 (59.5, 70.0)	89.7 (88.4, 91.0)	49.9 (45.1, 54.7)	94.2 (93.1, 95.2)
CAPTURE	Peak flow	Parallel (OR)	257	863	1249	76	77.2 (72.3,81.6)	59.1 (57.0, 61.2)	22.9 (20.5,25.5)	94.3 (92.9,95.5)
CDQ	Peak flow	Parallel (OR)	259	663	1449	74	77.8 (72.9, 82.1)	68.6 (66.6, 70.6)	28.1 (25.2, 31.1)	95.1 (93.9, 96.2)
C-SBQ	Peak flow	Parallel (OR)	268	729	1383	65	80.5 (75.8, 84.6)	65.5 (63.4, 67.5)	26.9 (24.2,29.7)	95.5 (94.3,96.5)
COPD-SQ	Peak flow	Parallel (OR)	259	687	1425	74	77.8 (72.9, 82.1)	67.5 (65.4, 69.5)	27.4 (24.6, 30.3)	95.1 (93.8, 96.1)
CAPTURE	Microspirometry	Parallel (OR)	262	764	1348	71	78.7 (73.9, 83.0)	63.8 (61.7, 65.9)	25.5 (22.9,28.3)	95.0 (93.7,96.1)
CDQ	Microspirometry	Parallel (OR)	261	585	1527	72	78.4 (73.6, 82.7)	72.3 (70.3, 74.2)	30.9 (2.8, 34.1)	95.5 (94.4, 96.5)
C-SBQ	Microspirometry	Parallel (OR)	271	675	1437	62	81.4 (76.8, 85.4)	68.0 (66.0, 70.0)	28.6 (25.8,31.6)	95.9 (94.7,96.8)
COPD-SQ	Microspirometry	Parallel (OR)	262	620	1492	71	78.7 (73.9, 83.0)	70.6 (68.7, 72.6)	29.7 (26.7, 32.8)	95.5 (94.3, 96.4)
CAPTURE	Peak flow	Serial (AND)	139	133	1979	194	41.7 (36.4, 47.2)	93.7 (92.6, 94.7)	51.1 (45, 57.2)	91.1 (89.8, 92.2)
CDQ	Peak flow	Serial (AND)	148	156	1956	185	44.4 (39.0, 50.0)	92.6 (91.4, 93.7)	48.7 (42.9, 54.5)	91.4 (90.1, 92.5)
C-SBQ	Peak flow	Serial (AND)	166	184	1928	167	49.8 (44.4, 55.4)	91.3 (90.0, 92.5)	47.4 (42.1, 52.8)	92 (90.8, 93.2)
COPD-SQ	Peak flow	Serial (AND)	149	160	1952	184	44.7 (39.3, 50.3)	92.4 (91.2, 93.5)	48.2 (42.5, 53.9)	91.4 (90.1, 92.5)
CAPTURE	Microspirometry	Serial (AND)	126	81	2031	207	37.8 (32.6, 43.3)	96.2 (95.3, 96.9)	60.9 (53.9, 67.6)	90.8 (89.5, 91.9)
CDQ	Microspirometry	Serial (AND)	138	83	2029	195	41.4 (36.1, 46.9)	96.1 (95.2, 96.9)	62.4 (55.7, 68.8)	91.2 (90.0, 92.4)
C-SBQ	Microspirometry	Serial	155	87	2025	178	46.5	95.9	64.0	91.9

		(AND)					(41.1, 52.1)	(94.9, 96.7)	(57.7, 70.1)	(90.7, 93)
COPD-SQ	Microspirometry	Serial (AND)	138	76	2036	195	41.4 (36.1, 46.9)	96.4 (95.5, 97.2)	64.5 (57.7, 70.9)	91.3 (90.0, 92.4)

*TP: True Positive

*FP: False Positive

*TN: True Negative

*FN: False Negative

*PPV: Positive Predictive Value

*NPV: Negative Predictive Value

Serial = positive on BOTH tests required for screen positivity

Parallel = positive on EITHER test required for screen positivity

TABLE 3: Comparative sensitivity for individual tests

Individual test	CAPTURE (95%CI,*P*)	CDQ (95%CI,*P*)	C-SBQ (95%CI,*P*)	COPD-SQ (95%CI,*P*)	Peak flow (95%CI,*P*)	Microspirometry (95%CI,*P*)
CAPTURE		-3.3(-9.6, 2.9; 0.3245)	-11.4(-16.9, 5.9; <0.0001)	-3.6(-9.6, 2.5; 0.2615)	-15.6(-22.1,-9.1; <0.0001)	-13.2(-20.2,-6.2; 0.0002)
CDQ			-8.1(-12.6,-3.6; 0.0003)	-0.3(-5.3, 4.7; 1.0000)	-12.3(-18.7, -6.0; 0.0001)	-9.9(-16.7,-3.2; 0.0037)
C-SBQ				7.8(3.2, 12.4; 0.0007)	-4.2(-10.4, 2.0; 0.1978)	-1.8(-8.4, 4.8; 0.6427)
COPD-SQ					-12.0(-18.3,-5.7; 0.0002)	-9.6(-16.4, -2.8; 0.0052)
Peak flow						2.4(-4.1, 8.9; 0.5047)
Microspirometry						

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing index tests in the column against index tests in the row. For example, sensitivity for CAPTURE is 3.3% lower than for CDQ (95%CI -9.6, 2.9; 0.3245).

TABLE 4: Comparative Specificity for individual tests

Individual test	CAPTURE (95%CI,*P*)	CDQ (95%CI,*P*)	C-SBQ (95%CI,*P*)	COPD-SQ (95%CI,*P*)	Peak flow (95%CI,*P*)	Microspirometry (95%CI,*P*)
CAPTURE		-8.4 (-10.7, -6.0; <0.0001)	-3.9 (-6.2, -1.6; 0.0008)	-7.1 (-9.3, -4.8; <0.0001)	-12.3 (-14.8, -9.8; <0.0001)	-19.5 (-21.8, -17.1; <0.0001)
CDQ			4.5 (3.0, 5.9; <0.0001)	1.3 (-0.4, 3.0; 0.1335)	-3.9 (-6.1, -1.8; 0.0003)	-11.1 (-13.2, -9.0; <0.0001)
C–SBQ				-3.1 (-4.8, -1.5; 0.0002)	-8.4 (-10.6, -.6.2; <0.0001)	-15.5 (-17.7, -13.3; <0.0001)
COPD-SQ					-5.3 (-7.4, -3.1; <0.0001)	-12.4 (-14.6, -10.3; <0.0001)
Peak flow						-7.1 (-9.1, -5.2; <0.0001)
Microspirometry						

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing index tests in the column against index tests in the row. For example, specificity for CAPTURE is 8.4% lower than for CDQ (95%CI -10.7, -6.0; <0.0001).

TABLE 5 Per patient cost, effectiveness and cost-effectiveness of selected screening strategies

Strategy	Cost per test UK£ (CNY)	Difference in cost UK£ (CNY)	True cases detected	Difference in true cases detected	ICER* UK£ (CNY) per additional true case detected
C-SBQ	2.22 (13.30)	-	0.0858	-	Dominated by microspirometry
Microspirometry	1.60 (9.60)	-0.62 (-3.70)	0.0883	0.0025	18.13 (108.78) vs no screening**
Peak flow	1.71 (10.25)	0.11 (0.64)	0.0915	0.0057	32.89 (197.36) vs microspirometry
C-SBQ and microspirometry	3.43 (20.59)	1.72 (10.35)	0.1184	0.0269	64.20 (385.20) vs peak flow

* ICER: Incremental cost-effectiveness ratio

**Due to the symptom-based question being excluded from the analysis, the next option is compared with no screening

References

[1] Global Initiative for Chronic Obstructive Lung Disease (GOLD).Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2018 report).https://goldcopd.org/wp-content/uploads/2019/12/GOLD-2020-FINAL-ver1.2-03Dec19_WMV.pdf.

[2] GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017; 5:691–706.

[3] Yin P, Wang H, Vos T, et al. A Subnational Analysis of Mortality and Prevalence of COPD in China From 1990 to 2013: Findings From the Global Burden of Disease Study 2013. Chest. 2016 Dec; 150(6):1269-1280. doi: 10.1016/j.chest.2016.08.1474. Epub 2016 Sep 29.

[4] Zhu B, Wang Y, Ming J, Chen W, Zhang L. Disease burden of COPD in China: a systematic review. Int J Chron Obstruct Pulmon Dis. 2018. 13: 1353-1364.

[5] Bednarek M, Maciejewski J, Wozniak M, Kuca P, Zielinski J. Prevalence, severity and underdiagnosis of COPD in the primary care setting. Thorax. 2008. 63(5): 402-7.

[6] Casas HA, de Oca M M, López VMV, Aguirre C, Schiavi E, Jardim JR. COPD Underdiagnosis and Misdiagnosis in a High-Risk Primary Care Population in Four Latin American Countries. A Key to Enhance Disease Diagnosis: The PUMA Study. PLoS One. 2016. 11(4): e0152266.

[7] Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. Lancet Respir Med. 2017. 5(5): 426-434.

[8] Determinants of Underdiagnosis of COPD in National and International Surveys . Lamprecht B, Soriano JB, Studnicka M, Kaiser B, Vanfleteren LE, Gnatiuc L, et al. Determinants of underdiagnosis of COPD in national and international surveys. Chest. 2015;148:971–85.

[9] Zhong N, Wang C, Yao W, et al. Prevalence of chronic obstructive pulmonary disease in China: a large, population-based survey.Am J Respir Crit Care Med 2007; 176: 753–60.

[10] Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonarydisease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet. 2018. 391(10131): 1706-1717. DOI: 10.1016/S0140-6736(18)30841-9.

[11] Screening for Chronic Obstructive Pulmonary Disease. US Preventive Services Task Force. Screening for Chronic Obstructive Pulmonary Disease: US Preventive Services Task Force Recommendation Statement. *JAMA* **315**, 1372-1377 (2016).

[12] UK National Screening Committee. UK National Screening Committee. An evaluation of screening for COPD against the National Screening Committee criteria. (2013).

[13] Screening for chronic obstructive pulmonary disease (COPD) in the general adult population. External review against programme appraisal criteria for the UK National Screening Committee. (2018).

[14] National Health and Family Planning Commission of the People. National Health and Family Planning Commission of the People's Republic of China. The 13th Five-Year Plan for Healthcare.Dec

27, 2016. http://www.gov.cn/zhengce/content/

[15] A. P. Dickens, D. A. Fitzmaurice, P. Adab, et al. Accuracy of Vitalograph lung monitor as a screening test for COPD in primary care. npj Primary Care Respiratory Medicine (2020) 30:2 ; https://doi.org/10.1038/s41533-019-0158-2

[16] Stanley AJ, Hasan I, Crockett AJ, van Schayck OC, Zwar NA. COPD Diagnostic Questionnaire (CDQ) for selecting at-risk patients for spirometry: a cross-sectional study in Australian general practice. NPJ Prim Care Respir Med. 2014. 24: 14024.

[17] Jackson H, Hubbard R. Detecting chronic obstructive pulmonary disease using peak flow rate: cross sectional survey. BMJ. 2003. 327(7416): 653-4.

[18] Zhang Q, Wang M, Li X, Wang H, Wang J. Do symptom-based questions help screen COPD among Chinese populations. Sci Rep. 2016. 6: 30419.

[19] Zhou YM, Chen SY, Tian J,et al. Development and validation of a chronic obstructive pulmonary disease screening questionnaire in China. Int J Tuberc Lung Dis. 2013 Dec;17(12):1645-51.

[20] Pan Z, Dickens AP, Chi C, et al. Study to evaluate the effectiveness and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among residents (⩾40 years) in four cities in China: protocol for a multicentre cross-sectional study on behalf of the Breathe Well group. BMJ Open 2020;10:e035738. doi:10.1136/bmjopen-2019-035738.

[21] Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 1978;118:1–120.

[22] Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. Eur Respir J 2009;34:648–654.

[23] Harris, PA, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377-381

[24] Harris, PA, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208.

[25] Price DB, Tinkelman DG, Halbert RJ, et al. Symptom-based questionnaire for identifying COPD in smokers. Respiration. 2006. 73(3): 285-95.

[26] Martinez FJ, Mannino D, Leidy NK, et al. A New Approach for Identifying Patients with Undiagnosed Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2017.195(6): 748-756.

[27] Frith P, Crockett A, Beilby J, et al. Simplified COPD screening: validation of the PiKo-6® in primary care. Prim Care Respir J. 2011. 20(2): 190-8, 2 p following 198.

[28] Labor M, Vrbica Ž, Gudelj I, Labor S, Plavec D. Diagnostic accuracy of a pocket screening spirometer in diagnosing chronic obstructive pulmonary disease in general practice: a cross sectional validation study using tertiary care as a reference. BMC Fam Pract. 2016. 17(1): 112.

[29] ATS/ERS task force: standardisation of lung function testing No 1 ERJ 2005: 26:153-161

[30] Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. Stat Med. 2002. 21(6): 835-52.

[31] Represas-Represas C, Fernández-Villar A, Ruano-Raviña A, Priegue-Carrera A, Botana-Rial M. Screening for Chronic Obstructive Pulmonary Disease: Validity and Reliability of a Portable Device in Non-Specialized Healthcare Settings. PLoS One. 2016. 11(1): e0145571.

[32] Van den Bemt L, Wouters BC, Grootens J,et al. Diagnostic accuracy of pre-bronchodilator

FEV1/FEV6 from microspirometry to detect airflow obstruction in primary care: a randomised cross-sectional study. NPJ Prim Care Respir Med. 2014 Aug 14;24:14033.

[33] https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm (accessed October 5th 2020)

[34] http://www.equator-network.org/reporting-guidelines/stard/(accessed June 1st 2020)

[35] Haroon S, Jordan R, Takwoingi Y, Adab P. Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis. BMJ Open. 2015. 5(10): e008133.

[36] Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 2013 Aug 06;18511 (11).

[37] van Dijk W, Tan W, Li P, et al. Clinical relevance of fixed ratio vs lower limit of normal of FEV1/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. Ann Fam Med. 2015;13(1):41-48. doi:10.1370/afm.1714

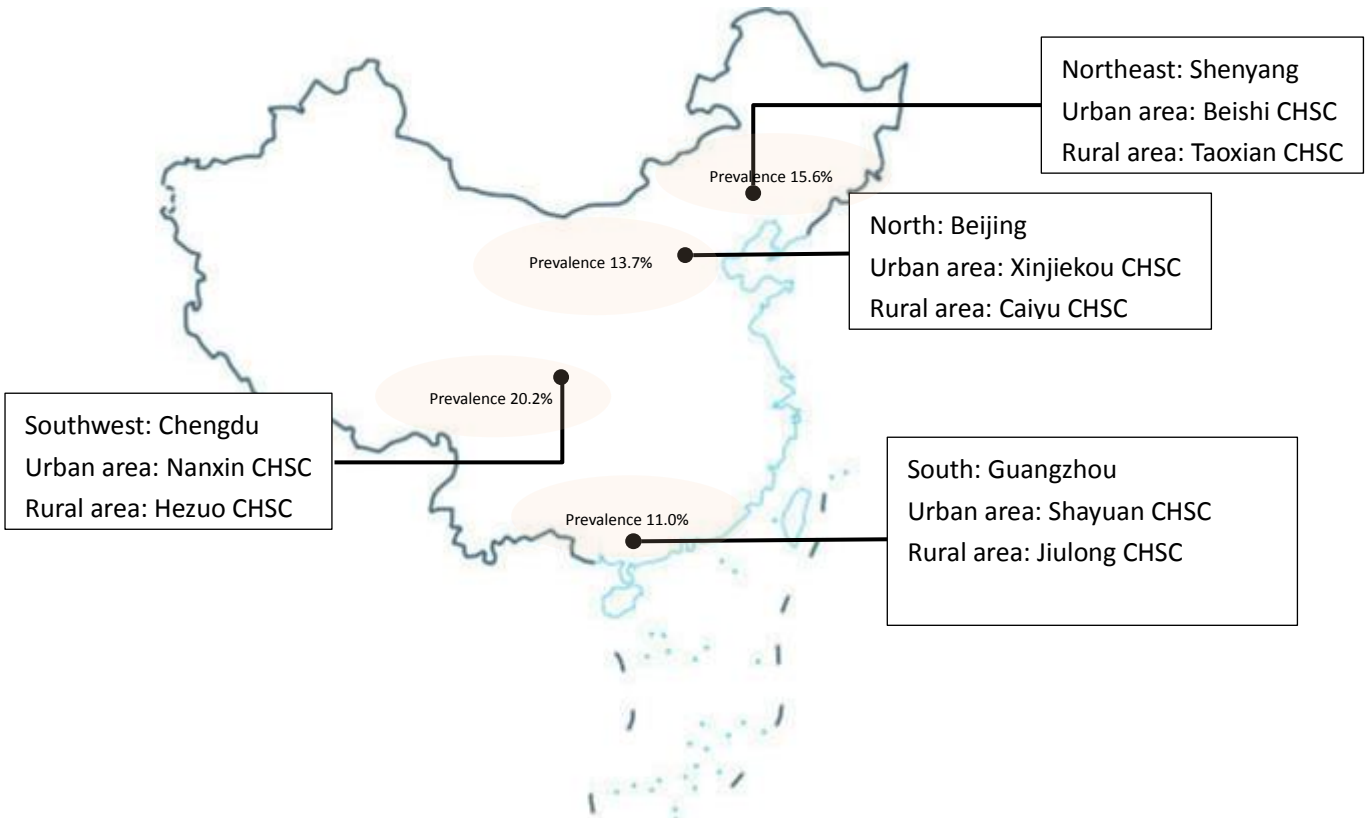


Figure 1 the map of Breathe Well-China research sites

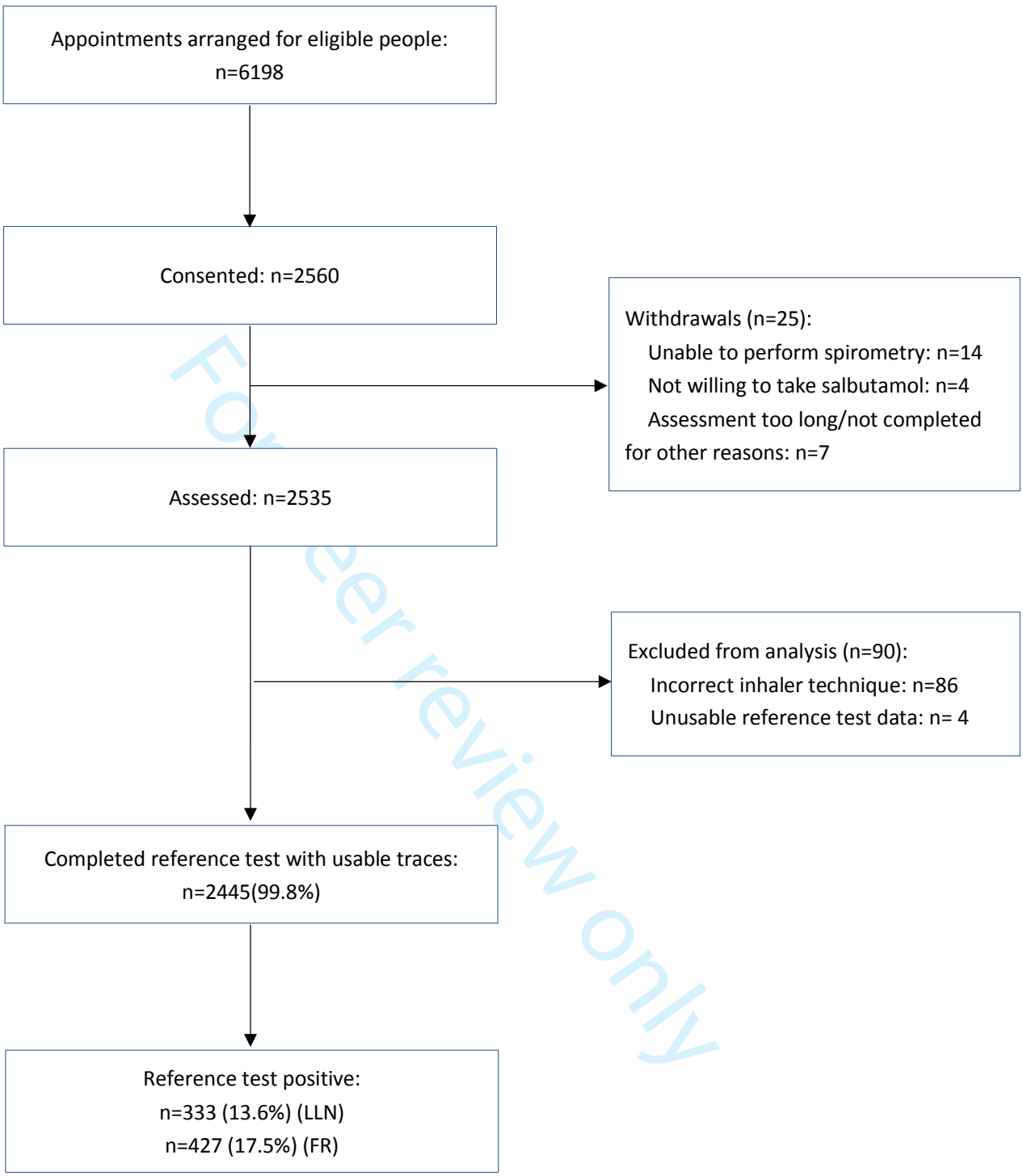


Figure 2 Study flow chart

Evaluating screening strategies for identifying undiagnosed COPD in China: a Breathe Well project

中国慢阻肺筛查策略评估：健康呼吸 Breathe Well 研究项目

Lung health questionnaire

肺部健康问卷

Participant Initials 研究对象编号	
Study ID 问卷编号	
Date 填写日期	
Interviewer ID 研究人员编号	

Some questions in the following booklets may appear similar. However, it is important that we ask these questions in slightly different ways so please complete all questions, answering them as accurately as possible.

一些问题可能相似，但是我们以稍微不同的方式提出这些问题很重要。

因此，请您完成所有的问题，并尽可能准确地作答。

CDQ

1. Age group, years
   年龄

40–49 ☐ 50-59 ☐ 60-69 ☐ 70+ ☐

2. What is your weight in kilograms?
   您的体重（公斤）？

   _____kilograms
   _____ 公斤

   What is your height in meters?
   您的身高（米）？

   _____ metres
   _____ 米

3. Smoking
   吸烟强度，包年

   What is the total number of years you have smoked?
   您一共吸烟多少年？

   _____ years

   _____ 年

   How many cigarettes do you currently smoke each day (or 'did smoke each day' if ex-smoker)?
   目前您每天吸多少支烟？（或，如果是既往吸烟者，过去您每天吸多少支烟？）

   _____ cigarettes

   _____ 支

4. Does the weather affect your cough?
   您的咳嗽是否受天气影响？

Yes ☐ No ☐

是 ☐ 否 ☐

5. Do you ever cough up phlegm (sputum) from your chest when you don't have a cold?
您不感冒的时候，会从胸腔里咳出痰吗？（区别于从嗓子中咳痰）

Yes ☐ No ☐
是 ☐ 否 ☐

6. Do you usually cough up phlegm (sputum) from your chest first thing in the morning?
清晨您的第一件事是从胸腔里咳出痰吗？

Yes ☐ No ☐
是 ☐ 否 ☐

7. How frequently do you wheeze?
您喘息的次数是多少？

Occasionally or more often ☐ Never ☐
有时候或更频繁 ☐ 从不 ☐

8. Do you have or have you had any allergies?
目前或既往您有过敏物吗？

Yes ☐ No ☐
是 ☐ 否 ☐

CAPTURE

1. Have you ever lived or worked in a place with dirty or polluted water or air, smoke or second-hand smoke or dust?
您是否曾经在有脏的或受到污染的水或空气，烟雾或二手烟雾或灰尘的地方生活或工作？

Yes ☐ No ☐
是 ☐ 否 ☐

2. Does your breathing change with seasons, weather or air quality?
您的呼吸是否随着季节、天气或空气质量而变化？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Does your breathing make it difficult to do things such as carry heavy loads, shovel dirt or snow, jog, play tennis or swim?
您的呼吸是否会使您难以进行一些工作，比如提重物，铲土或积雪，慢跑，打网球或游泳等？

Yes ☐ No ☐
是 ☐ 否 ☐

4. Compared to others your age, do you tire easily?
和您的同龄人相比，您是否容易感到疲劳？

Yes ☐ No ☐
是 ☐ 否 ☐

5. In the past 12 months, how many times did you miss work, school, or other activities due to a cold, bronchitis, or pneumonia?
在过去的 12 个月里，您有多少次因感冒、支气管炎或肺炎而错过了工作、学校或其他活动？

0 ☐ 1 ☐ 2 or more ☐
0 ☐ 1 ☐ 2 或以上 ☐

Symptom-based questionnaire

1. How frequently are you exposed to second-hand smoking?
您接触二手烟的频率是多少？

<7hrs per week ☐ ≥7hrs per week ☐
< 7小时/周 ☐ > 7 小时/周 ☐

2. Do you often cough when you do not have a cold?
您是否在不感冒的时候经常咳嗽？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Do you have more signs of shortness of breath compared with others of the same age?
和同龄人相比，您是否有更多的呼吸急促的症状？

Yes ☐ No ☐
是 ☐ 否 ☐

4. Have you had long-term exposure to dust or chemical particles?
您是否长期地接触粉尘或化学颗粒？

Yes ☐ No ☐

是 ☐ 否 ☐

5. Did you have a history of chronic respiratory diseases when you were a child?
在您孩童时期，您是否有慢性呼吸疾病的病史？

Yes ☐是 No ☐
是 ☐ 否 ☐

COPD-SQ

1. Do you often cough?
您是否经常咳嗽？

Yes ☐ No ☐
是 ☐ 否 ☐

2. Family history of respiratory disease
是否有呼吸疾病家族史？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Exposure to biomass smoke from cooking fires
是否接触烹饪产生的生物烟雾？

Yes ☐ No ☐
是 ☐ 否 ☐

Appendix 2: Costs, timings and assumptions for case-finding strategies

Assessment timings	**Minutes per patient**
Symptom questionnaire (completion and processing)	6
Peak flow	2
Microspirometry	4
Confirmatory NDD spirometry	30
Staff	**Hourly costs (UK £)**
Clinic staff	6.25
Additional unit costs	**(UK £)**
Symptom questionnaire	0.10
Peak flow	
Mouthpiece cost per patient	0.10
Overall equipment cost	8.00
Other consumable costs per patient	0.21
Microspirometry (COPD-6)	
Mouthpiece cost per patient	0.10
Overall equipment cost	75.00
Battery cost per year	5.00
Other consumable costs per patient	0.21
Confirmatory NDD spirometry	
Mouthpiece cost per patient	1.30
Overall equipment cost	1,095
Salbutamol cost per patient	0.70
Other consumable and equipment costs per patient	0.25
Assumptions	
Number of visits per year per case finding clinic (assuming 48 tests per day, 5 days a week, 50 weeks a year)	12,000
Number of visits per year per NDD spirometry clinic (assuming 16 tests per day, 5 days a week, 50 weeks a year)	4,000
Lifetime of peak flow meter	1 year
Lifetime of microspirometry	6 years
Lifetime of NDD spirometry	6 years
Proportion of patients requiring staff assistance with questionnaire	95%
Cost of case finding method per patient	**(UK £)**
Symptom questionnaire	0.70
Peak flow	0.52
Microspirometry	0.73
Confirmatory NDD spirometry	4.90

Appendix 3-TABLE 1: SERIAL (AND) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

Strategies	Peak flow	Microspirometry
CAPTURE + peak flow	-25.5 (-30.5,-20.5; <0.0001)	
CDQ + peak flow	-22.8 (-27.6,-18.0; <0.0001)	
C-SBQ + peak flow	-17.4 (-21.8,-13.0; <0.0001)	
COPD-SQ + peak flow	-22.5 (-27.3,-17.7; <0.0001)	
CAPTURE + microspirometry		-27.0 (-32.1,-22.0; <0.0001)
CDQ + microspirometry		-23.4 (-28.3, -18.6; <0.0001)
C-SBQ + microspirometry		-18.3 (-22.8,-13.9; <0.0001)
COPD-SQ + microspirometry		-23.4 (-28.3,-18.6; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, sensitivity for CAPTURE + peak flow is 25.5% lower than for peak flow (95%CI -30.5, -20.5; <0.0001).

Appendix 3-TABLE 2: SERIAL (AND) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

Strategies	Peak flow	Microspirometry
CAPTURE + peak flow	11.1 (9.7, 12.5; <0.0001)	
CDQ + peak flow	10.0 (8.7, 11.4; <0.0001)	
C-SBQ + peak flow	8.7 (7.5, 10.0; <0.0001)	
COPD-SQ + peak flow	9.8 (8.5, 11.2; <0.0001)	
CAPTURE + microspirometry		6.4 (5.3, 7.5; <0.0001)
CDQ + microspirometry		6.3 (5.3, 7.4; <0.0001)
C-SBQ + microspirometry		6.2 (5.1, 7.2; <0.0001)
COPD-SQ + microspirometry		6.7 (5.6, 7.8; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 11.1% higher than for peak flow (95%CI 9.7, 12.5; <0.0001).

Appendix 3-TABLE 3: SERIAL (AND) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

Strategies	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	-9.9 (-13.4, -6.4; <0.0001)			
CAPTURE + microspirometry	-13.8 (-17.8, -9.8; <0.0001)			
CDQ + peak flow		-10.5 (-14.1, -6.9; <0.0001)		
CDQ + microspirometry		-13.5 (-17.5, -9.5; <0.0001)		
C-SBQ + peak flow			-13.2 (-17.2, -9.3; <0.0001)	
C-SBQ + microspirometry			-16.5 (-20.8, -12.2; <0.0001)	
COPD-SQ + peak flow				-10.5 (-14.1, -6.9; <0.0001)
COPD-SQ + microspirometry				-13.8 (-17.8, 9.8; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, sensitivity for CAPTURE is 3.3% lower than for CDQ (95%CI -9.6, 2.9; 0.3245).

Appendix 3-TABLE 4: SERIAL (AND) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	23.4 (21.6,25.3; <0.0001)			
CAPTURE + microspirometry	25.9 (24.0,27.8; <0.0001)			
CDQ + peak flow		14.0 (12.4, 15.5; <0.0001)		
CDQ + microspirometry		17.4 (15.8, 19.1; <0.0001)		
C-SBQ + peak flow			17.1 (15.4, 18.7; <0.0001)	
C-SBQ + microspirometry			21.7 (19.9, 23.5; <0.0001)	
COPD-SQ + peak flow				15.1 (13.5, 16.7; <0.0001)
COPD-SQ + microspirometry				19.1 (17.4, 20.8; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 23.4% higher than for CAPTURE (95%CI 21.6, 25.3; <0.0001).

Appendix 3-TABLE 5: PARALLEL (OR) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

	Peak flow	Microspirometry
CAPTURE + peak flow	9.9 (6.4, 13.4; <0.0001)	
CDQ + peak flow	10.5 (6.9, 14.1; <0.0001)	
C-SBQ + peak flow	13.2 (9.3, 17.2; <0.0001)	
COPD-SQ + peak flow	10.5 (6.9, 14.1; <0.0001)	
CAPTURE + microspirometry		13.8 (9.8, 17.8; <0.0001)
CDQ + microspirometry		13.5 (9.5, 17.5; <0.0001)
C-SBQ + microspirometry		16.5 (12.2, 20.8; <0.0001)
COPD-SQ + microspirometry		13.8 (9.8, 17.8; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against index tests in the row. For example, sensitivity for CAPTURE + peak flow is 9.9% higher than for peak flow (95%CI 6.4, 13.4; <0.0001).

Appendix 3-TABLE 6: PARALLEL (OR) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

	Peak flow	Microspirometry
CAPTURE + peak flow	-23.4 (-25.3, -21.6; <0.0001)	
CDQ + peak flow	-14.0 (-15.5, -12.4; <0.0001)	
C-SBQ + peak flow	-17.1 (-18.7, -15.4; <0.0001)	
COPD-SQ + peak flow	-15.1 (-16.7, -13.5; <0.0001)	
CAPTURE + microspirometry		-25.9 (-27.8, -24.0; <0.0001)
CDQ + microspirometry		-17.4 (-19.1,-15.8; <0.0001)
C-SBQ + microspirometry		-21.7 (-23.5, -19.9; <0.0001)
COPD-SQ + microspirometry		-19.1 (-20.8, -17.4; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 23.4% lower than for peak flow (95%CI -25.3, -21.6; <0.0001).

Appendix 3-TABLE 7: PARALLEL (OR) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

Strategies	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	25.5 (20.5, 30.5; <0.0001)			
CAPTURE + microspirometry	27.0 (22.0, 32.1; <0.0001)			
CDQ + peak flow		22.8 (18.1, 27.6; <0.0001)		
CDQ + microspirometry		23.4 (18.6, 28.3; <0.0001)		
C-SBQ + peak flow			17.4 (13.0, 21.8; <0.0001)	
C-SBQ + microspirometry			18.3 (13.9, 22.8; <0.0001)	
COPD-SQ + peak flow				22.5 (17.7, 27.3; <0.0001)
COPD-SQ + microspirometry				23.4 (18.6, .28.3; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies tests in the column against strategies in the row. For example, sensitivity for CAPTURE + peak flow is 25.5% higher than for CAPTURE (95%CI 20.5, 30.5; <0.0001).

Appendix 3-TABLE 8: PARALLEL (OR) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	-11.1 (-12.5,-9.7; <0.0001)			
CAPTURE + microspirometry	-6.4 (-7.5, -5.3; <0.0001)			
CDQ + peak flow		-10.0 (-11.4,-8.7; <0.0001)		
CDQ + microspirometry		-6.3 (-7.4, -5.3; <0.0001)		
C-SBQ + peak flow			-8.7 (-10.0, -7.5; <0.0001)	
C-SBQ + microspirometry			-6.2 (-7.2, -5.1; <0.0001)	
COPD-SQ + peak flow				-9.8 (-11.2, -8.5; <0.0001)
COPD-SQ + microspirometry				-6.7 (-7.8, -5.6; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against index tests in the row. For example, specificity for CAPTURE + peak flow is 11.1% lower than for CAPTURE (95%CI -12.5, -9.7; <0.0001).

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1~~1~~
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3~~2~~
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4~~3~~
	4	Study objectives and hypotheses	4~~3~~
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5~~3~~
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5~~4~~
	9	Whether participants formed a consecutive, random or convenience series	5~~4~~
Test methods	**10a**	Index test, in sufficient detail to allow replication	5~~4~~
	10b	Reference standard, in sufficient detail to allow replication	6~~5~~
	11	Rationale for choosing the reference standard (if alternatives exist)	6~~5~~
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5-6~~4~~
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6~~5~~
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6~~5~~
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6~~5~~
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7~~5~~-6
	15	How indeterminate index test or reference standard results were handled	NA~~No report~~
	16	How missing data on the index test and reference standard were handled	NA~~No report~~
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	NA~~No report~~
	18	Intended sample size and how it was determined	6~~5~~
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Figure 2
	20	Baseline demographic and clinical characteristics of participants	Table 1
	21a	Distribution of severity of disease in those with the target condition	Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	Table 1
	22	Time interval and any clinical interventions between index test and reference standard	NA~~5~~
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Table 2
	25	Any adverse events from performing the index test or the reference standard	NA~~12~~
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	10-11~~10-11~~
	27	Implications for practice, including the intended use and clinical role of the index test	12~~No report~~
OTHER INFORMATION			
	28	Registration number and name of registry	15~~12~~
	29	Where the full study protocol can be accessed	15~~12~~
	30	Sources of funding and other support; role of funders	15~~12~~

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Accuracy and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among primary care patients (≥40 years) in China: a cross-sectional screening test accuracy study. Findings from the Breathe Well group

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-051811.R1
Article Type:	Original research
Date Submitted by the Author:	16-Jun-2021
Complete List of Authors:	Pan, Zihan; Peking University First Hospital; Peking University Third Hospital, Pulmonary and Critical Care Medicine Dickens, Andrew; University of Birmingham, Institute of Applied Health Research Chi, Chunhua; Peking University First Hospital, General Practice Department, Peking University First Hospital, Beijing, 100034, China Kong, Xia; Peking University First Hospital, General Practice Enocson, Alexandra; University of Birmingham, Applied Health Research Cooper, Brendan; Lung Investigation Unit, University Hospitals Birmingham NHS Foundation Trust Adab, Peymane; University of Birmingham, Public Health Cheng, Kar Keung; University of Birmingham, Department of Public Health and Epidemiology Sitch, Alice; University of Birmingham, Public Health Jowett, Sue; University of Birmingham, Health Economics Unit Adams, Rachel; University of Birmingham, Institute of Applied Health Research Correia-de-Sousa, Jaime; Life and Health Sciences Research Institute (ICVS)/3B's — PT Government Associate Laboratory, University of Minho; Horizonte Family Health Unit, Farley, Amanda; University of Birmingham, Gale, Nicola K.; Univ Birmingham, Health Services Management Centre Jolly, Kate; University of Birmingham Maglakelidze, Mariam; Georgian Respiratory Association Maglakelidze, Tamaz; Georgian Respiratory Association Martins, Sonia; ABC Medical School, Family Medicine Stavrikj, Katarina; Medical Faculty, Centre for Family Medicine Stelmach, Rafael ; University of São Paulo, (5) Pulmonary Department, Heart Institute (InCor), School of Medicine Turner, Alice; University of Birmingham, Institute of Applied Health Research Williams, Sian; National Heart and Lung Institute Jordan, Rachel; University of Birmingham, Public Health and Epidemiology
Primary Subject Heading:	Respiratory medicine

Secondary Subject Heading:	Diagnostics, Public health, General practice / Family practice, Health economics
Keywords:	GENERAL MEDICINE (see Internal Medicine), RESPIRATORY MEDICINE (see Thoracic Medicine), Chronic airways disease < THORACIC MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page

Accuracy and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among primary care patients (≥40 years) in China: a cross-sectional screening test accuracy study. Findings from the Breathe Well group

†Zihan Pan[1,2]; †Andrew P Dickens[3]; *Chunhua Chi[1]; Xia Kong[1]; Alexandra Enocson[3]; Brendan G Cooper[3,4]; *Peymané Adab[3]; Kar Keung Cheng[3,5]; Alice Sitch[3,6]; Sue Jowett[3], Rachel Adams[3], Jaime Correia-de-Sousa[7,8], Amanda Farley[3], Nicola Gale[9], Kate Jolly[3], Mariam Maglakelidze[10,11], Tamaz Maglakelidze[10], Sonia M Martins[12], Katarina Stavrikj[13], Rafael Stelmach[14], Alice M Turner[3], Siân Williams[7], Rachel E Jordan[3].

†Joint first authors

*Joint corresponding authors:

Name: Chunhua Chi

Postal address: Peking University First Hospital, No.8 XiShiKu Street, Xicheng District, Beijing, 100034, China

E-mail: chichunhua2012@qq.com

Telephone: +86 13910987530,

Fax numbers: +86 010-66158996

Name: Peymane Adab

Postal address: Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK B15 2TT

Email: p.adab@bham.ac.uk

Telephone: 0121 4143777

Co-authors details:

1. Department of General Practice, Peking University First Hospital, Beijing, China
2. Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China
3. Institute of Applied Health Research, University of Birmingham, Birmingham, UK.

4. University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital Birmingham, Birmingham, UK
5. General Practice Development and Research Centre, Peking University Health Science Centre, Beijing, China
6. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, UK
7. International Primary Care Respiratory Group
8. Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga Portugal. ICVS/3B's, PT Government Associate Laboratory, Braga/Guimarães, Portugal
9. Health Services Management Centre, School of Social Policy, College of Social Sciences, University of Birmingham, Birmingham, UK
10. Georgian Respiratory Association, Georgia
11. Petre Shotadze Tbilisi Medical Academy, Georgia
12. Family Medicine, ABC Medical School, Sao Paolo, Brazil
13. Center for Family Medicine, Faculty of Medicine, Ss.Cyril and Methodius University in Skopje, North Macedonia
14. Pulmonary Division, Heart Institute (InCor), Hospital das Clinicas da Faculdade de Medicina da Uinversidade de Sao Paulo, Brazil

Key word: COPD; screening test accuracy; screening strategies, health economics; primary care; multicentre study

Word count: 3425 words

Abstract

Objectives: To examine the accuracy and cost-effectiveness of various COPD screening tests and combinations within a Chinese primary care population.

Design Screening test accuracy study

Setting: Urban and rural community health centres in four municipalities of China: Beijing (north), Chengdu (southwest), Guangzhou (south) and Shenyang (northeast).

Participants: Community residents aged 40 years and above who attended community health centres for any reason were invited to participate. 2445 participants (mean age 59.8 [SD 9.6] years, 39.1% [n=956] male) completed the study (February-December 2019), 68.9% (n=1684) were never-smokers and 3.6% (n=88) had an existing COPD diagnosis. 13.7% (n=333) of participants had spirometry-confirmed airflow obstruction.

Interventions: Participants completed six index tests (screening questionnaires [CDQ, CAPTURE, Chinese Symptom-based questionnaire or C-SBQ, COPD-SQ], microspirometry [COPD-6], peak flow [USPE]) and the reference test (ndd Easy On-PC).

Primary and secondary outcomes: Cases were defined as those with FEV_1/FVC below the lower limit of normal (LLN-GLI) on the reference test. Performance of individual screening tests and their combinations was evaluated, with cost-effectiveness analyses providing cost per additional true case detected.

Results: Airflow measurement devices (sensitivities 64.9% [95% CI 59.5, 70.0] and 67.3% [61.9, 72.3], specificities 89.7% [88.4, 91.0] and 82.6% [80.9, 84.2] for microspirometry and peak flow respectively) generally performed better than questionnaires, the most accurate of which was C-SBQ (sensitivity 63.1% [57.6%, 68.3%], specificity 74.2% [72.3%, 76.1%]). The combination of C-SBQ and microspirometry used in parallel maximised sensitivity (81.4%) [76.8, 85.4] and had specificity of 68.0% [66.0, 70.0], with an incremental cost-effectiveness ratio of £64.20 (CNY385) per additional case detected compared with peak flow.

Conclusions: Simple screening tests to identify undiagnosed COPD within the primary care setting in China is possible, and a combination of C-SBQ and microspirometry is the most sensitive and cost-effective. Further work is required to explore optimal cut-points and effectiveness of programme implementation.

Trial registration: ISRCTN13357135

Article summary

Strengths and limitations of this study

- Defining airflow obstruction according to the lower limit of normal increased the likelihood that identified cases were true COPD.
- Recruiting participants from both urban and rural community hospitals maximised the generalisability of our findings to primary care patients.
- This study did not explore optimal cut-points for index tests, thus further work is required.
- While the study was conducted in four geographically disparate municipalities, our findings may not be generalisable to all adults ≥40 years old in China.

Introduction

Chronic obstructive pulmonary disease (COPD) is a common long-term condition characterized by persistent respiratory symptoms and airflow limitation[1]. Nearly one-third of the 3.2 million annual global deaths from COPD are from China[2, 3] where COPD ranks among the top three leading causes of death with associated direct medical costs of 118% of local average annual income[4]. COPD develops slowly, resulting in delays in symptom recognition and high rates of underdiagnosis. Ninety percent of the estimated 100 million people with COPD in China are undiagnosed; slightly higher than the 60-80% underdiagnosis rate worldwide[5-9]. Symptom reporting and recognition are lower in China, with 60% of diagnosed patients not reporting symptoms such as cough, expectoration and wheeze[10].

While COPD screening programmes are not currently endorsed in the United States and UK[11-13], considering the high proportion and heavy burden of undiagnosed disease[4], early identification is being prioritised in China. National policies recommend screening for undiagnosed COPD[14], but do not specify which screening tests to use. Furthermore, though spirometry is required for clinical diagnosis[1], it is not widely available in primary care settings in China. Therefore screening could reduce the numbers needing spirometry referral.

Globally, various COPD screening tools have been developed, including questionnaires and airflow measurement devices[15-17]. However, accuracy studies were mainly conducted in Western countries and have not been validated in a Chinese population where the distribution and underlying causes of undiagnosed COPD may differ. Furthermore, the majority of Chinese studies have used secondary or tertiary care COPD populations rather than people from community settings[18, 19]. Finally, the cost-effectiveness of different screening tests have not been previously estimated in China; a crucial consideration given the high prevalence of COPD in this middle-income country.

We examined the accuracy and cost-effectiveness of various screening tests and combinations within a Chinese primary care population.

Methods

Study design and participants

We conducted a cross-sectional, multicentre study to evaluate the accuracy and cost-effectiveness of various COPD screening tests and test combinations in primary care in China. Full details of participant recruitment and study assessments are described in the published protocol[20].

Participants were recruited from one urban and one rural community health centre (CHC) in each of four municipalities: Beijing (North China), Chengdu (southwest China), Guangzhou (south China) and Shenyang (northeast China) (Figure 1). Between February-December 2019, community dwelling residents aged 40 years and above who attended CHCs for any reason were invited to participate, either directly by the attending clinician, or through poster or social media (WeChat) advertisements. Participants who were unable to give informed consent, had contraindications for spirometry or unable to perform the test for other reasons were excluded.

Eligible participants provided informed consent at the start of the assessment visit, prior to height and weight measurement and completion of all index and reference tests. Participants also completed a study questionnaire concerning demographics, smoking status, exposures, medical diagnoses, respiratory symptoms[21] and quality of life[22]. Data were entered into a secure online REDCap database[23, 24].

Participants with airflow obstruction on the reference test were offered health education, smoking cessation advice, influenza vaccination and inhalers if relevant, or referred to tertiary hospitals for further treatment including pharmacotherapy or pulmonary rehabilitation.

Study assessment

Index tests

The six index tests included four screening questionnaires: COPD Diagnostic Questionnaire (CDQ, cut-point ≥20)[16, 25], CAPTURE (cut-point ≥2)[26], COPD Screening Questionnaire (COPD-SQ, cut-point ≥16)[19] and, the Chinese symptom-based questionnaire (C-SBQ, cut-point ≥17)[18] and two airflow measurement devices: microspirometry (Vitalograph COPD-6, cut-point for positive test FEV_1/FEV_6 <0.78)[27, 28], peak flow (USPE, cut-point <350 l/min men, <250 l/min women)[26].

Questionnaires were selected to maximize symptom capture and minimize item duplication, whilst allowing comparison of the most relevant questionnaires (Appendix 1). Previously defined cut-points were used to identify participants at risk of COPD.

Trained researchers provided instructions before participants performed 3 pre-bronchodilator manoeuvres with each airflow measurement device. The order of administering peak flow or microspirometry alternated by participant, and the best FEV_1 and FEV_6 measure for each device were used for analyses, irrespective of which attempt they came from.

Participants completed the four screening questionnaires immediately after administration of a bronchodilator (400ug, Salbutamol). Questionnaires were intended to be self-completed, although researchers were available to assist if needed.

Reference test

The reference test was quality diagnostic spirometry (ndd Easy On-PC), performed 20-60 minutes after bronchodilation. Spirometry was administered by a second researcher not involved in the index tests and blind to their results. Participants performed a minimum of 3 blows, and a maximum of 6, until repeatability within 100mls or 5% [29]. Flow volume curves were classified according to the ATS/ERS[29] criteria. Tests with at least 3 curves meeting these criteria, were "Good." "Acceptable" tests contained at least one curve which concurred with the criteria, allowing accurate assessment of FEV_1. If accurate assessment was not possible the curves were classified as "unacceptable", and the test was excluded from analysis. All traces were over-read for quality by one of three independent respiratory experts and graded according to standard criteria[29], without knowledge of the index test results.

Airflow obstruction was defined as post-bronchodilator FEV_1/FVC ratio below the lower limit of normal (LLN) using the GLI equations.

Sample size

The Alonzo method[30] for paired test accuracy studies was used to calculate the sample size. Assuming independence of tests and prevalence of 12%, we required 1622 participants to detect a difference in sensitivity of 10% (95% vs 85%[16, 26, 31, 32] for the comparison of CAPTURE and peak flow for example) with 90% power. With lower test sensitivity (90% vs 80%) 2279 participants are

needed to detect this difference with 90% power.

Statistical analysis

The diagnostic performance of each index test was investigated by presenting 2x2 tables and calculating the sensitivity, specificity, positive predictive value and negative predictive value with 95% confidence intervals. Comparative test accuracy was assessed by calculating the difference in sensitivity and specificity, presenting 95% confidence intervals and using McNemar's test.

The primary analysis compared the sensitivity and specificity between the CAPTURE screening questionnaire and the peak flow meter. The comparison was specified a priori as CAPTURE was rigorously developed, accounted for exposures other than smoking and was intended for use in conjunction with peak flow. Secondary analyses evaluated the comparative performance of all other individual index tests, as well as plausible combination test strategies. Test strategies were formed using two pre-determined combinations for appropriate pairs of individual index tests (questionnaires and lung function tests); firstly, to maximise sensitivity, where a participant with a positive result for either index test would be positive for the strategy (parallel testing strategy) and secondly, to maximise specificity, where a participant would need a positive result on both index tests to be positive for the strategy (serial testing strategy).

All analyses were conducted in Stata v15.

Economic analysis

We conducted a cost-effectiveness analysis to calculate the cost per additional case detected for all tests and combination strategies. The strategies were ordered by the number of true cases detected, from least to greatest, and the principle of dominance was applied to eliminate redundant strategies (where they were more costly and less effective). Each test was then compared with the next best alternative. For the purpose of this paper, the individual index tests and the combination strategy with the highest sensitivity were compared.

The unit costs and quantity of any equipment, medication and consumables required, staff time (and salary costs) to deliver each individual test and use of facilities were determined to calculate the health care costs of delivering each screening test/strategy. Each individual test was timed at a sample of assessment clinics to estimate an overall mean time and range for each test.

Equipment costs were depreciated (at 3.5% a year) over the estimated lifespan of the equipment (ranging from 1 to 6 years). Cost per patient visit was calculated assuming the equipment would be used for 12,000 patients per clinic per year. It was also assumed that positive cases would be confirmed with quality diagnostic spirometry (assuming 4000 patients/year). Costs were calculated in UK£ for a price year of 2019, and converted to Chinese Yuan (¥) using Purchasing Power Parities (PPP[33]) with a conversion rate of 6.0 (Appendix 2).

The paper follows the STARD guidance[34] for reporting studies of diagnostic accuracy.

Results

Sample

We invited 6198 eligible people to the study. A total of 2560 (41.3%) consented, of whom 25 withdrew and 90 were excluded from analysis (86 because of incorrect inhaler technique, four had unusable spirometry data). 2445 participants with complete data on all index and reference test were included in the final analysis (Figure 2). Approximately two thirds (68.0%) were recruited through their attending clinician, 24.5% via advertisements and 7.5% through word of mouth.

The mean age of participants was 59.8 (SD 9.6), 39.1% (n=956) were male, two thirds (n=1684, 68.9%) were never smokers and over half lived in an urban area (1338, 54.7%). 46.7% had no diagnosed conditions (n=1142); the most common diagnosed condition was hypertension (n=842, 34.4%), 3.6% (n=88) had an existing COPD diagnosis and 8.4% (n=205) had an existing chronic bronchitis/emphysema diagnosis (Table 1). 99.8% of participants had an acceptable usable spirometry (with 63.3% (n=1547) defined as good). 13.6% (n=333) of participants had spirometry-confirmed airflow obstruction using the LLN criteria, of whom 175 (52.5%) had moderate to severe obstruction i.e. GOLD stage II or above [1]. Those with airflow obstruction were older (63.5 vs 69.2 years) and more likely to be male (59.8% vs 35.8%), have a positive smoking history (55.5% vs 27.3%) and childhood respiratory infections (14.7% vs 7.8%) compared to those without airflow obstruction. Respiratory symptoms of wheeze, productive cough or breathlessness (mMRC≥2) were reported by 52.9% of those with airflow obstruction (66.3% of those who were GOLD stage II or above), and 25.1% of those without. Amongst participants with no previously reported COPD diagnosis, the prevalence of obstruction was

9.9% (n=218), of whom 89 (40.8%) were GOLD stage II or above. Using the $FEV_1/FVC<0.7$ criteria[1], 17.4% (n=425) of all participants had airflow obstruction.

Performance of individual tests and screening strategies

Among the screening questionnaires, the C-SBQ had the highest sensitivity in detecting airflow obstruction at 63.1% (57.6%, 68.3%), CAPTURE the lowest sensitivity (51.7% [46.1, 57.1]), with CDQ (55.0% [49.4%, 60.4%]) similar to COPD-SQ [55.3% (49.7%, 60.7%)]. The CDQ had the highest specificity (78.6% [76.8%, 80.4%]). CAPTURE compared to CDQ had the most obvious difference in specificity of 8.4% (-10.7, -6.0; p<0.001) ((Table 2, Table 3, Table 4)).

Both peak flow and microspirometry devices had higher sensitivity and specificity compared to all questionnaires (Table 3, Table 4). Peak flow had the highest sensitivity (67.3%) and microspirometry the highest specificity (89.7%) (Table 3, Table 4).

Of the combined screening strategies, C-SBQ combined with airflow measurement devices in parallel (i.e. recorded as screen-positive if either test was positive) had the best performance, with sensitivities of 80.5%-81.4%, and specificities of 65.5%-68%. Parallel strategies (requiring either test to be positive) optimised sensitivity and serial strategies (requiring both tests to be positive) optimised specificity. Taking CAPTURE and peak flow as an example, a parallel combination had sensitivity of 77.2% compared to 41.7% in serial combination, while the specificity significantly increased from 59.1% to 93.7% (Table 2).

Overall, test performance was slightly higher when screening questionnaires were combined with microspirometry rather than peak flow. Strategies including CAPTURE performed less well compared to those based on other questionnaires. Parallel strategies including the C-SBQ had the highest sensitivities, whereas those based on the CDQ had the highest specificity (Table 2, Table 3). Full comparisons of serial and parallel strategies are described in Appendix 3.

Cost-effectiveness of preferred screening tests

Analysis of the C-SBQ parallel strategies revealed that the most costly strategy was the combination of C-SBQ and microspirometry, but this also detected the most true cases (Table 5). The C-SBQ alone was dominated by microspirometry (more costly, less effective). The incremental cost-effectiveness ratio (ICER) for C-SBQ and microspirometry (versus peak flow) was greatest at

£64.20 (CNY 385.20), but could be considered cost-effective if the threshold willingness to pay for an additional true case detected in China is at least CNY 385.

Discussion

This is the first study assessing the accuracy of individual screening tools and their combinations to identify undiagnosed COPD within Chinese community populations. We showed that the combination of a simple questionnaire and airflow measurement device could adequately identify adults requiring diagnostic spirometry. Our overall findings were consistent with a meta-analysis of studies from other countries[35], that airflow measurement devices were more accurate than questionnaires, and that combinations of screening tests improved ability to detect COPD in primary care. Within single test strategies, microspirometry had the best performance (sensitivity 64.9%, specificity 89.7%). For combination strategies, the C-SBQ and microspirometry used in parallel, maximised sensitivity (81.4%) with reasonable specificity (68%) and would be deemed cost-effective if the Chinese health service was willing to pay ≥CNY 385 per additional case detected.

C-SBQ had the highest sensitivity of all screening questionnaires in our study, with comparable specificity. However, accuracy of the C-SBQ was worse than reported in the validation paper of the Chinese tool, with lower sensitivity (63.1% vs 82.5%) but slightly higher specificity (74.2% vs 72.9%). The observed discrepancy may be due to differences in the spectrum of clinical characteristics[36] (community sample rather than tertiary care population in previous study) and airflow obstruction criteria used (we used the lower limit of normal rather than the GOLD criteria).

Inclusion of the C-SBQ and the CDQ from which it was derived allowed direct comparison of the two measures, confirming that C-SBQ was more accurate for use in Chinese community populations when prioritising sensitivity (sensitivity 63.1% vs 55.0% with slightly lower specificity 74.2% vs 78.6%).

Direct comparison between our findings and those of previous studies was limited by differences in populations and pre-test probabilities. COPD among never smokers is more common in China than in western countries and we included never smokers in this study to maximise the range of potential COPD risk factors represented e.g. environmental exposures such as dust,

biomass fumes and passive smoking, as well as active smoking. Inevitably this contributed to the lower test performance observed. Furthermore, the CAPTURE questionnaire was originally designed to detect more severe COPD. The different case definition in our study therefore precludes direct comparison with previous studies (we plan to report accuracy for detecting more severe clinically significant COPD in a future publication).

Our test accuracy study has highlighted the strengths of different screening tests, which can be used to evaluate future screening programmes. We recruited a large number of participants from urban and rural settings in four geographically diverse municipalities in China, and the proportion of never smokers in our sample (68.9%) was similar to that found in a recent nationally representative cross-sectional study in China (71.4%)[10], which included a younger population (age 20+). We demonstrated that lung function tests and diagnosis of COPD can be implemented by GPs and nurses after a structured training course with regular quality over reading and feedback, as evidenced by 99% usable spirometry and consistently good quality spirometry in most GP sites. The fully paired study design enabled us to compare the accuracy of multiple index tests and strategies. Alternating the order of peak flow and microspirometry tests during assessments decreased the potential training effect that could have been introduced when conducting consecutive lung function tests in a research context.

We defined the reference test as airflow obstruction regardless of clinical symptoms, to reflect the methods of previous studies and also account for the differing symptom profile reported among Chinese populations, where chronic respiratory symptoms are less recognised. In our study, just over half of those with obstruction were likely to benefit from some treatment due to reported symptoms, and a further quarter of those obstructed would benefit from smoking cessation advice as they had a positive smoking history but no respiratory symptoms.

Accuracy might have differed if the GOLD criteria were used, though unlikely to substantially change the comparative performance of the tests. Defining airflow obstruction according to the LLN criteria increased the likelihood that participants testing positive on study spirometry were true COPD cases, rather than detecting comorbidities with similar clinical presentations such as cardiovascular disease[37]. As pre-bronchodilator spirometry was omitted from the study assessment to minimise participant burden and increase uptake in this large community-based study, we could not assess airflow reversibility.

Our study population included slightly more women than men (60% women). As smoking prevalence is also much lower among women, our study cannot provide an accurate estimate of COPD prevalence. However this should not impact on the estimate of screening test accuracy, which was the primary objective. It was not possible to exclude diagnosed COPD patients from this study, as Chinese community health centres do not have COPD registers and patients are frequently unaware of their condition. However, as the aim of our study was to determine accuracy of different screening tests by comparing all tests against a reference standard, rather than to evaluate the implementation of a screening programme, inclusion of COPD patients was justified. By including some people with known COPD, we maximised the number of test positives in the study sample.

Although China has recently introduced a national policy of COPD screening, there is no current guidance regarding the tests to use or which test characteristics (i.e. sensitivity / specificity) to prioritise. Considering the estimated high prevalence of undiagnosed COPD in China, highly sensitive strategies may be preferred to maximise the number of detected cases, although this would result in large numbers being referred for diagnostic spirometry, many of whom would be false positives. However, the potential inefficiency may be offset by a recent policy to include spirometry in routine primary care health consultations; avoiding the need to refer patients to hospital for diagnostic assessment. While the more sensitive parallel strategies may be preferential in the Chinese healthcare setting, there is a trade-off between sensitivity and specificity according to epidemiology, resources and context; hence, serial strategies may be considered optimal in other settings.

If the strategy of C-SBQ and microspirometry were used in practice and had the same accuracy as reported here, it is likely that true COPD cases who were not detected (false negatives) would have mild disease and would re-attend with recurring symptoms, offering further opportunities for referral to diagnostic spirometry.

While our analyses used recommended cut-points for the index tests, it is important to explore their optimal cut-points when applied in this context, as many tests were developed with alternate purposes and/or populations in mind. Thresholds used to indicate airflow obstruction (either in the screening tests or reference test) may not be valid in the whole Chinese population as adequate reference values for lung function are currently unreliable.

Although we have determined the accuracy of different tests when used for screening Chinese community populations for undiagnosed COPD, we did not evaluate the implementation of a screening programme. A recently published model-based cost-effectiveness analysis from China which used international data on QALYs, demonstrated that use of a screening questionnaire combined with a hand-held spirometer was cost-saving compared to no screening, but this did not compare different screening strategies and was not based on data from an implementation trial[38]. It is important to undertake a trial to compare the effectiveness and cost-effectiveness of the most efficient screening strategy identified in this study (maximising yield with acceptable false positive rate) against usual care on yield and clinical outcomes. Such a trial would need to assess uptake of screening and incorporate pathways for clinical assessment and subsequent treatment for test positive cases. In our study sample >75% had potential to benefit; >half with obstruction had treatable symptoms and a further quarter with obstruction and no symptoms would benefit from smoking cessation advice. We presented cost per additional true case detected, however no country has, to date, stated a willingness to pay threshold for this outcome. The quality-adjusted life year (QALY) is a more common metric in health economic analyses, with established cost per QALY thresholds. Although outside the remit of our test accuracy study, future work should attempt to extrapolate cases detected to the management of patients with COPD, to assess the impact on quality of life and survival to allow the calculation of QALYs.

In conclusion, we have demonstrated that within the primary care setting in China, the most efficient screening test strategy was a combination of the C-SBQ and microspirometry where a positive test in either would result in a referral for diagnostic spirometry. Further work is required to explore optimal cut-points and there is a need for a clinical trial to evaluate whether a screening programme using this test combination is clinically and cost-effective.

Contributors

Rachel E Jordan and Peymane Adab co-led the study design, with contributions and advice from all other authors. Chunhua Chi, Xia Kong, KK Cheng contributed to decisions on outcome measures. Chunhua Chi and KK Cheng advised on involving GP practices. Brendan Cooper, Andrew P Dickens, Alexandra Enocson, Rachel E Jordan and Peymane Adab advised on lung function testing. Brendan

Cooper and Alexandra Enocson provided training and oversaw the quality assessment for lung function testing. Andrew P Dickens, Rachel E Jordan, Alice Sitch and Peymane Adab designed the testing strategy. Alice Sitch and Sue Jowett designed the analysis plan and economic evaluations respectively. Zihan Pan coordinated the data collection, with support from Andrew P Dickens, Rachel E Jordan and Peymane Adab. Zihan Pan conducted the statistical analysis, supported by Alice Sitch, Sue Jowett and Andrew P Dickens. Zihan Pan and Andrew P Dickens wrote the manuscript with input from all other authors. Chunhua Chi was the local PI and oversaw all activities in China. Rachel Adams, Jaime Correia-de-Sousa, Amanda Farley, Nicola Gale, Kate Jolly, Mariam Maglakelidze, Tamaz Maglakelidze, Sonia M Martins, Katarina Stavrikj, Rafael Stelmach, Alice M Turner, and Siân Williams contributed to the development and oversight of this study. As part of the Breathe Well Global Health Research Group, all authors contributed to and approved the final version.

Acknowledgement

The investigator and collaborative team include: R Jordan, P Adab (CIs), Z Pan, AP Dickens, C Chi, X Kong, A Enocson, BG Cooper, KK Cheng, A Sitch, S Jowett, R Adams, J Correia de Sousa, A Farley, N Gale, K Jolly, M Maglakelidze, T Maglakelidze, SM Martins, K Stavrikj, R Stelmach, A Turner and S Williams. We thank the dedicated team of researchers at Department of General Practice, Peking University First Hospital for managing and co-ordinating the project. We are also grateful for support from our Trial Steering Committee (Dr. Semira Manaseki-Holland [Chair], Prof. David Mannino, Mr. Zhiwei Zhang [patient], Dr. Xueying Zhou [clinician], Prof. Chunhua Chi, Dr. Rachel Jordan), and) and The International Scientific Advisory Committee (Prof. Debbie Jarvis, Dr. Semira Manaseki-Holland, Prof. David Mannino, Prof. Niels Chevannes). We especially want to thank the patients included in the study and the clinicians at all community hospitals for their assistance, without whom this project would not have been possible.

The clinicians involved in this study included: Wenyan Duan, Mengmeng Guo, Zhi Jin (Caiyu Community Hospital, Beijing); Yuling Li, Shuo Chen, Kang Rong, Yi Cai, Jing Bai (Xinjiekou Community Hospital, Beijing); Wei Xiong, Honglin Luo, Juan Chen, Na Li (Hezuo Community Hospital, Chengdu); Min Chen, Yaling Luo, Xiaolong Zhou (Nanxin Community Hospital, Chengdu), Shixing Liu, Guoxiong Han, Shuhua Sun, Jian Long Li, Jianlong Cai, Haimei Chen, Yuanping Chen

(Shayuan Community Hospital, Guangzhou); Bin Hu, Huilian Xuan, Jieya Zhang, Huanping Chen, Fengqi Liu, Haixia Yin (Jiulong Community Hospital, Guangzhou); Ying Chen, Xinzheng Cui, Yao Chen, Hui Peng, Weihan Jin, Xiaoguang Xu (Beishi Community Hospital, Shenyang); Zhenguo Gao, Zhongning Kan, Lina Song, Yongsheng Tong, Yuxuan Zhang, Shuai Mu, Yang Gao (Taoxian Community Hospital, Shenyang); Zhennan Qi, Hui Pang, Anny Gao, Xue Jin (Peking University First Hospital).

We thank the patient advisory group for their useful comments, including Mr. Zhiwei Zhang, Mr. Chen Wang, Mr Baoyin Fang and Mr. Deqian Ma.

We thank the support of the Health Committees in Xicheng District, Beijing and in Guangzhou for identifying potential research sites.

We would also like to acknowledge Radmila Ristovska (1955-2020), also involved in the initiation of this study.

We gratefully acknowledge International Primary Care Respiratory Group (IPCRG) for introducing us to the primary care networks involved in this study and for its continued facilitation of clinical engagement.

We obtained appropriate permissions to use the Symptom Based Questionnaire, COPD Screening Questionnaire, COPD Diagnostic Questionnaire and CAPTURE.

Ethics approval

The study has been approved by Peking University First Hospital (2018-R-141, PUFH) and University of Birmingham (ERN_18-1177, UoB).

Patient and public involvement

The research team conducted a research prioritization exercise with patients, clinicians and policy makers, and the need to identify effective screening strategies for undiagnosed COPD was one of the research areas prioritized. The patient advisory group advised on the format of study material prior to recruitment commencing. All stakeholders involved in this exercise received study updates twice a year, and were kept informed of findings and consulted at the end of the study regarding implications for practice and policy decisions, as well as advice on appropriate dissemination of study findings.

In addition, the study has a Trial Steering Committee (TSC) that meets regularly and comprises various independent members, including a patient and a clinician representative as well as international experts in respiratory research and several members of the study research team.

Serious adverse events (SAE)

No SAE from performing the index tests or the reference test in the study.

Registration number and name of registry

The protocol for this study was previously published and registered on ISRCTN registry. The number was ISRCTN13357135 and the full study protocol can be accessed at *http://www.isrctn.com* (ISRCTN13357135).

Funding

This research was funded by the National Institute for Health Research (NIHR) NIHR global group on global COPD in primary care, University of Birmingham, (project reference: 16/137/95) using UK aid from the UK Government to support global health research. This paper presents independent research supported by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. The funder of the study had no role in study design, data collection, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

The authors declare no conflicts of interest.

Data availability statement

Data are available upon reasonable request. All data requests should be submitted to authors CC and PA for consideration. Access to anonymised data may be granted following review.

Additional file

Appendix 1. Screening questionnaires

Appendix 2. Costs, timings and assumptions for case-finding strategies

Appendix 3. Comparisons of serial and parallel strategies

Figure and table legends

Figure 1 Map of Breathe Well-China research sites

Figure 2 Study flow chart

Table 1 Characteristics of study participants

Table 2 Accuracy of Index tests and strategies

Table 3 Comparative sensitivity for individual tests

Table 4 Comparative Specificity for individual tests

Table 5 Per patient cost, effectiveness and cost-effectiveness of selected screening strategies

TABLE 1 Characteristics of study participants

Characteristic	Total sample (n=2445)	Reference test positive (n=333)	Reference test negative (n=2112)
Male sex, n (%)	956 (39.1%)	199 (59.8%)	757 (35.8%)
Age in years ; mean(SD)	59.8 (9.6)	63.5 (8.9)	59.2 (9.6)
BMI; mean (SD)	24.9 (3.5)	24.3 (3.4)	25.0 (3.4)
Education, n (%)			
High school or below	1879 (76.9)	277 (83.2%)	1602 (75.9%)
Above High school	566 (23.1)	56 (16.8%)	510 (24.1%)
Employment status, n(%)			
Employed	674 (27.6%)	54 (16.2%)	620 (29.4%)
Unemployed	665 (27.2%)	98 (29.4%)	567 (26.9%)
Retired	1106 (45.2%)	181 (54.4%)	925 (43.8%)
Living area, n(%)			
Urban	1338 (54.7%)	174 (52.3%)	1164 (55.1%)
Smoking status, n(%)			
Current smoker	472 (19.3%)	113 (33.9%)	359 (17.0%)
Ex-smoker	289 (11.8%)	72 (21.6%)	217 (10.3%)
Never smoker	1684 (68.9%)	148 (44.5%)	1536 (72.7%)
Male	--	27 (18.2%)	--
Female	--	121 (81.8%)	--
Pack years; mean (SD)	9.0 (17.8)	18.0 (21.0)	7.6 (16.8)
Health in general, n(%)			
Very Good-good	1255 (51.3%)	127 (38.1%)	1128 (53.4%)
Fair-very bad	1190 (48.7%)	206 (61.9%)	984 (46.6%)
Diagnosed conditions, n(%)			
COPD	88 (3.6%)	64 (19.2%)	24 (1.1%)
Chronic bronchitis/emphysema	205 (8.4%)	93 (27.9%)	112 (5.3%)
Asthma	105 (4.3%)	48 (14.4%)	57 (2.7%)
Tuberculosis	41 (1.7%)	12 (3.6%)	29 (1.4%)
Hypertension	842 (34.4%)	119 (35.7%)	723 (34.2%)
Diabetes Mellitus	330 (13.5%)	43 (12.9%)	287 (13.6%)
Heart disease	274 (11.2%)	43 (12.9%)	231 (10.9%)
Other	269 (11.0%)	31 (9.3%)	238 (11.3%)
None of the above	1142 (46.7%)	106 (31.8%)	1036 (49.1%)
Symptoms, n(%)			
At least occasional wheeze	322 (13.2)	110 (33.0)	212 (10.0)
Productive cough	457 (18.7)	117 (35.1)	340 (16.1)
mMRC, n(%)			

Grade 0-1	2222 (90.9%)	257 (77.2%)	1965 (93.0%)
Grade 2-4	223 (9.1%)	76 (22.8%)	147 (7.0%)
CAT, mean(SD)	6.1 (5.4%)	8.9 (6.9%)	5.6 (4.9%)
Bronchitis, pneumonia or severe whooping cough in childhood	169 (6.9%)	38 (11.4%)	131 (6.2%)
Tuberculosis in childhood	45 (1.8%)	11 (3.3%)	34 (1.6%)
Exposure to pollutants*, n (%)			
Current/past exposure	2256 (92.3%)	307 (92.2%)	1949 (92.3%)
Never	189 (7.7%)	26 (7.8%)	163 (7.7%)
Year(s) of exposure, mean (SD)	8.9 (6.4)	9.1 (6.6)	8.8 (6.4)
GOLD stage if <LLN†, n (%)			
I (FEV_1 ≥80% predicted)	--	158 (47. 5%)	--
II (FEV_1 50-79% predicted)	--	137 (41.1%)	--
III (FEV_1 30-49% predicted)	--	33 (9.9%)	--
IV (FEV_1 <30% predicted)	--	5 (1.5%)	--

* cooking fumes, biomass smoking, gas, steams, dust

† LLN = lower limit of normal

TABLE 2 Accuracy of Index tests and strategies

Part 1	Part 2	Strategy type	TP*	FP*	TN*	FN*	Sensitivity% (95% CI)	Specificity% (95% CI)	PPV%* (95% CI)	NPV%* (95% CI)
CAPTURE	n/a	Individual	172	628	1484	161	51.7 (46.1, 57.1)	70.3 (68.3, 72.2)	21.5 (18.7, 24.5)	90.2 (88.7, 91.6)
CDQ	n/a	Individual	183	451	1661	150	55.0 (49.4, 60.4)	78.6 (76.8, 80.4)	28.9 (25.4, 32.6)	91.7 (90.4, 92.9)
C-SBQ	n/a	Individual	210	545	1567	123	63.1 (57.6, 68.3)	74.2 (72.3, 76.1)	27.8 (24.6, 31.2)	92.7 (91.4, 3.9)
COPD-SQ	n/a	Individual	184	479	1633	149	55.3 (49.7, 60.7)	77.3 (75.5, 79.1)	27.8 (24.4, 31.3)	91.6 (90.3, 92.9)
Peak flow	n/a	Individual	224	368	1744	109	67.3 (61.9, 72.3)	82.6 (80.9, 84.2)	37.8 (33.9, 41.9)	94.1 (92.9, 95.1)
Microspirometry	n/a	Individual	216	217	1895	117	64.9 (59.5, 70.0)	89.7 (88.4, 91.0)	49.9 (45.1, 54.7)	94.2 (93.1, 95.2)
CAPTURE	Peak flow	Parallel (OR)	257	863	1249	76	77.2 (72.3,81.6)	59.1 (57.0, 61.2)	22.9 (20.5,25.5)	94.3 (92.9,95.5)
CDQ	Peak flow	Parallel (OR)	259	663	1449	74	77.8 (72.9, 82.1)	68.6 (66.6, 70.6)	28.1 (25.2, 31.1)	95.1 (93.9, 96.2)
C-SBQ	Peak flow	Parallel (OR)	268	729	1383	65	80.5 (75.8, 84.6)	65.5 (63.4, 67.5)	26.9 (24.2,29.7)	95.5 (94.3,96.5)
COPD-SQ	Peak flow	Parallel (OR)	259	687	1425	74	77.8 (72.9, 82.1)	67.5 (65.4, 69.5)	27.4 (24.6, 30.3)	95.1 (93.8, 96.1)
CAPTURE	Microspirometry	Parallel (OR)	262	764	1348	71	78.7 (73.9, 83.0)	63.8 (61.7, 65.9)	25.5 (22.9,28.3)	95.0 (93.7,96.1)
CDQ	Microspirometry	Parallel (OR)	261	585	1527	72	78.4 (73.6, 82.7)	72.3 (70.3, 74.2)	30.9 (2.8, 34.1)	95.5 (94.4, 96.5)
C-SBQ	Microspirometry	Parallel (OR)	271	675	1437	62	81.4 (76.8, 85.4)	68.0 (66.0, 70.0)	28.6 (25.8,31.6)	95.9 (94.7,96.8)
COPD-SQ	Microspirometry	Parallel (OR)	262	620	1492	71	78.7 (73.9, 83.0)	70.6 (68.7, 72.6)	29.7 (26.7, 32.8)	95.5 (94.3, 96.4)
CAPTURE	Peak flow	Serial (AND)	139	133	1979	194	41.7 (36.4, 47.2)	93.7 (92.6, 94.7)	51.1 (45, 57.2)	91.1 (89.8, 92.2)
CDQ	Peak flow	Serial (AND)	148	156	1956	185	44.4 (39.0, 50.0)	92.6 (91.4, 93.7)	48.7 (42.9, 54.5)	91.4 (90.1, 92.5)
C-SBQ	Peak flow	Serial (AND)	166	184	1928	167	49.8 (44.4, 55.4)	91.3 (90.0, 92.5)	47.4 (42.1, 52.8)	92 (90.8, 93.2)
COPD-SQ	Peak flow	Serial (AND)	149	160	1952	184	44.7 (39.3, 50.3)	92.4 (91.2, 93.5)	48.2 (42.5, 53.9)	91.4 (90.1, 92.5)
CAPTURE	Microspirometry	Serial (AND)	126	81	2031	207	37.8 (32.6, 43.3)	96.2 (95.3, 96.9)	60.9 (53.9, 67.6)	90.8 (89.5, 91.9)
CDQ	Microspirometry	Serial (AND)	138	83	2029	195	41.4 (36.1, 46.9)	96.1 (95.2, 96.9)	62.4 (55.7, 68.8)	91.2 (90.0, 92.4)
C-SBQ	Microspirometry	Serial	155	87	2025	178	46.5	95.9	64.0	91.9

		(AND)					(41.1, 52.1)	(94.9, 96.7)	(57.7, 70.1)	(90.7, 93)
COPD-SQ	Microspirometry	Serial (AND)	138	76	2036	195	41.4 (36.1, 46.9)	96.4 (95.5, 97.2)	64.5 (57.7, 70.9)	91.3 (90.0, 92.4)

*TP: True Positive

*FP: False Positive

*TN: True Negative

*FN: False Negative

*PPV: Positive Predictive Value

*NPV: Negative Predictive Value

Serial = positive on BOTH tests required for screen positivity

Parallel = positive on EITHER test required for screen positivity

TABLE 3: Comparative sensitivity for individual tests

Individual test	CAPTURE (95%CI,*P*)	CDQ (95%CI,*P*)	C-SBQ (95%CI,*P*)	COPD-SQ (95%CI,*P*)	Peak flow (95%CI,*P*)	Microspirometry (95%CI,*P*)
CAPTURE		-3.3(-9.6, 2.9; 0.3245)	-11.4(-16.9, 5.9; <0.0001)	-3.6(-9.6, 2.5; 0.2615)	-15.6(-22.1,-9.1; <0.0001)	-13.2(-20.2,-6.2; 0.0002)
CDQ			-8.1(-12.6,-3.6; 0.0003)	-0.3(-5.3, 4.7; 1.0000)	-12.3(-18.7, -6.0; 0.0001)	-9.9(-16.7,-3.2; 0.0037)
C-SBQ				7.8(3.2, 12.4; 0.0007)	-4.2(-10.4, 2.0; 0.1978)	-1.8(-8.4, 4.8; 0.6427)
COPD-SQ					-12.0(-18.3,-5.7; 0.0002)	-9.6(-16.4, -2.8; 0.0052)
Peak flow						2.4(-4.1, 8.9; 0.5047)
Microspirometry						

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing index tests in the column against index tests in the row. For example, sensitivity for CAPTURE is 3.3% lower than for CDQ (95%CI -9.6, 2.9; 0.3245).

TABLE 4: Comparative Specificity for individual tests

Individual test	CAPTURE (95%CI,*P*)	CDQ (95%CI,*P*)	C-SBQ (95%CI,*P*)	COPD-SQ (95%CI,*P*)	Peak flow (95%CI,*P*)	Microspirometry (95%CI,*P*)
CAPTURE		-8.4 (-10.7, -6.0; <0.0001)	-3.9 (-6.2, -1.6; 0.0008)	-7.1 (-9.3, -4.8; <0.0001)	-12.3 (-14.8, -9.8; <0.0001)	-19.5 (-21.8, -17.1; <0.0001)
CDQ			4.5 (3.0, 5.9; <0.0001)	1.3 (-0.4, 3.0; 0.1335)	-3.9 (-6.1, -1.8; 0.0003)	-11.1 (-13.2, -9.0; <0.0001)
C–SBQ				-3.1 (-4.8, -1.5; 0.0002)	-8.4 (-10.6, -.6.2; <0.0001)	-15.5 (-17.7, -13.3; <0.0001)
COPD-SQ					-5.3 (-7.4, -3.1; <0.0001)	-12.4 (-14.6, -10.3; <0.0001)
Peak flow						-7.1 (-9.1, -5.2; <0.0001)
Microspirometry						

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing index tests in the column against index tests in the row. For example, specificity for CAPTURE is 8.4% lower than for CDQ (95%CI -10.7, -6.0; <0.0001).

TABLE 5 Per patient cost, effectiveness and cost-effectiveness of selected screening strategies

Strategy	Cost per test UK£ (CNY)	Difference in cost UK£ (CNY)	True cases detected	Difference in true cases detected	ICER* UK£ (CNY) per additional true case detected
C-SBQ	2.22 (13.30)	-	0.0858	-	Dominated by microspirometry
Microspirometry	1.60 (9.60)	-0.62 (-3.70)	0.0883	0.0025	18.13 (108.78) vs no screening**
Peak flow	1.71 (10.25)	0.11 (0.64)	0.0915	0.0057	32.89 (197.36) vs microspirometry
C-SBQ and microspirometry	3.43 (20.59)	1.72 (10.35)	0.1184	0.0269	64.20 (385.20) vs peak flow

* ICER: Incremental cost-effectiveness ratio

**Due to the symptom-based question being excluded from the analysis, the next option is compared with no screening

References

[1] Global Initiative for Chronic Obstructive Lung Disease (GOLD).Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2018 report).https://goldcopd.org/wp-content/uploads/2019/12/GOLD-2020-FINAL-ver1.2-03Dec19_WMV.pdf.

[2] GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017; 5:691–706.

[3] Yin P, Wang H, Vos T, et al. A Subnational Analysis of Mortality and Prevalence of COPD in China From 1990 to 2013: Findings From the Global Burden of Disease Study 2013. Chest. 2016 Dec; 150(6):1269-1280. doi: 10.1016/j.chest.2016.08.1474. Epub 2016 Sep 29.

[4] Zhu B, Wang Y, Ming J, Chen W, Zhang L. Disease burden of COPD in China: a systematic review. Int J Chron Obstruct Pulmon Dis. 2018. 13: 1353-1364.

[5] Bednarek M, Maciejewski J, Wozniak M, Kuca P, Zielinski J. Prevalence, severity and underdiagnosis of COPD in the primary care setting. Thorax. 2008. 63(5): 402-7.

[6] Casas HA, de Oca M M, López VMV, Aguirre C, Schiavi E, Jardim JR. COPD Underdiagnosis and Misdiagnosis in a High-Risk Primary Care Population in Four Latin American Countries. A Key to Enhance Disease Diagnosis: The PUMA Study. PLoS One. 2016. 11(4): e0152266.

[7] Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. Lancet Respir Med. 2017. 5(5): 426-434.

[8] Determinants of Underdiagnosis of COPD in National and International Surveys . Lamprecht B, Soriano JB, Studnicka M, Kaiser B, Vanfleteren LE, Gnatiuc L, et al. Determinants of underdiagnosis of COPD in national and international surveys. Chest. 2015;148:971–85.

[9] Zhong N, Wang C, Yao W, et al. Prevalence of chronic obstructive pulmonary disease in China: a large, population-based survey.Am J Respir Crit Care Med 2007; 176: 753–60.

[10] Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonarydisease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet. 2018. 391(10131): 1706-1717. DOI: 10.1016/S0140-6736(18)30841-9.

[11] Screening for Chronic Obstructive Pulmonary Disease. US Preventive Services Task Force. Screening for Chronic Obstructive Pulmonary Disease: US Preventive Services Task Force Recommendation Statement. *JAMA* **315**, 1372-1377 (2016).

[12] UK National Screening Committee. UK National Screening Committee. An evaluation of screening for COPD against the National Screening Committee criteria. (2013).

[13] Screening for chronic obstructive pulmonary disease (COPD) in the general adult population. External review against programme appraisal criteria for the UK National Screening Committee. (2018).

[14] National Health and Family Planning Commission of the People. National Health and Family Planning Commission of the People's Republic of China. The 13th Five-Year Plan for Healthcare.Dec

27, 2016. http://www.gov.cn/zhengce/content/
[15] A. P. Dickens, D. A. Fitzmaurice, P. Adab, et al. Accuracy of Vitalograph lung monitor as a screening test for COPD in primary care. npj Primary Care Respiratory Medicine (2020) 30:2 ; https://doi.org/10.1038/s41533-019-0158-2
[16] Stanley AJ, Hasan I, Crockett AJ, van Schayck OC, Zwar NA. COPD Diagnostic Questionnaire (CDQ) for selecting at-risk patients for spirometry: a cross-sectional study in Australian general practice. NPJ Prim Care Respir Med. 2014. 24: 14024.
[17] Jackson H, Hubbard R. Detecting chronic obstructive pulmonary disease using peak flow rate: cross sectional survey. BMJ. 2003. 327(7416): 653-4.
[18] Zhang Q, Wang M, Li X, Wang H, Wang J. Do symptom-based questions help screen COPD among Chinese populations. Sci Rep. 2016. 6: 30419.
[19] Zhou YM, Chen SY, Tian J,et al. Development and validation of a chronic obstructive pulmonary disease screening questionnaire in China. Int J Tuberc Lung Dis. 2013 Dec;17(12):1645-51.
[20] Pan Z, Dickens AP, Chi C, et al. Study to evaluate the effectiveness and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among residents (⩾40 years) in four cities in China: protocol for a multicentre cross-sectional study on behalf of the Breathe Well group. BMJ Open 2020;10:e035738. doi:10.1136/bmjopen-2019-035738.
[21] Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 1978;118:1–120.
[22] Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. Eur Respir J 2009;34:648–654.
[23] Harris, PA, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377-381
[24] Harris, PA, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208.
[25] Price DB, Tinkelman DG, Halbert RJ, et al. Symptom-based questionnaire for identifying COPD in smokers. Respiration. 2006. 73(3): 285-95.
[26] Martinez FJ, Mannino D, Leidy NK, et al. A New Approach for Identifying Patients with Undiagnosed Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2017.195(6): 748-756.
[27] Frith P, Crockett A, Beilby J, et al. Simplified COPD screening: validation of the PiKo-6® in primary care. Prim Care Respir J. 2011. 20(2): 190-8, 2 p following 198.
[28] Labor M, Vrbica Ž, Gudelj I, Labor S, Plavec D. Diagnostic accuracy of a pocket screening spirometer in diagnosing chronic obstructive pulmonary disease in general practice: a cross sectional validation study using tertiary care as a reference. BMC Fam Pract. 2016. 17(1): 112.
[29] ATS/ERS task force: standardisation of lung function testing No 1 ERJ 2005: 26:153-161
[30] Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. Stat Med. 2002. 21(6): 835-52.
[31] Represas-Represas C, Fernández-Villar A, Ruano-Raviña A, Priegue-Carrera A, Botana-Rial M. Screening for Chronic Obstructive Pulmonary Disease: Validity and Reliability of a Portable Device in Non-Specialized Healthcare Settings. PLoS One. 2016. 11(1): e0145571.
[32] Van den Bemt L, Wouters BC, Grootens J,et al. Diagnostic accuracy of pre-bronchodilator

FEV1/FEV6 from microspirometry to detect airflow obstruction in primary care: a randomised cross-sectional study. NPJ Prim Care Respir Med. 2014 Aug 14;24:14033.

[33] https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm (accessed October 5th 2020)

[34] http://www.equator-network.org/reporting-guidelines/stard/(accessed June 1st 2020)

[35] Haroon S, Jordan R, Takwoingi Y, Adab P. Diagnostic accuracy of screening tests for COPD: a systematic review and meta-analysis. BMJ Open. 2015. 5(10): e008133.

[36] Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 2013 Aug 06;18511 (11).

[37] van Dijk W, Tan W, Li P, et al. Clinical relevance of fixed ratio vs lower limit of normal of FEV1/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. Ann Fam Med. 2015;13(1):41-48. doi:10.1370/afm.1714

[38] https://www.nature.com/articles/s41533-021-00233-z

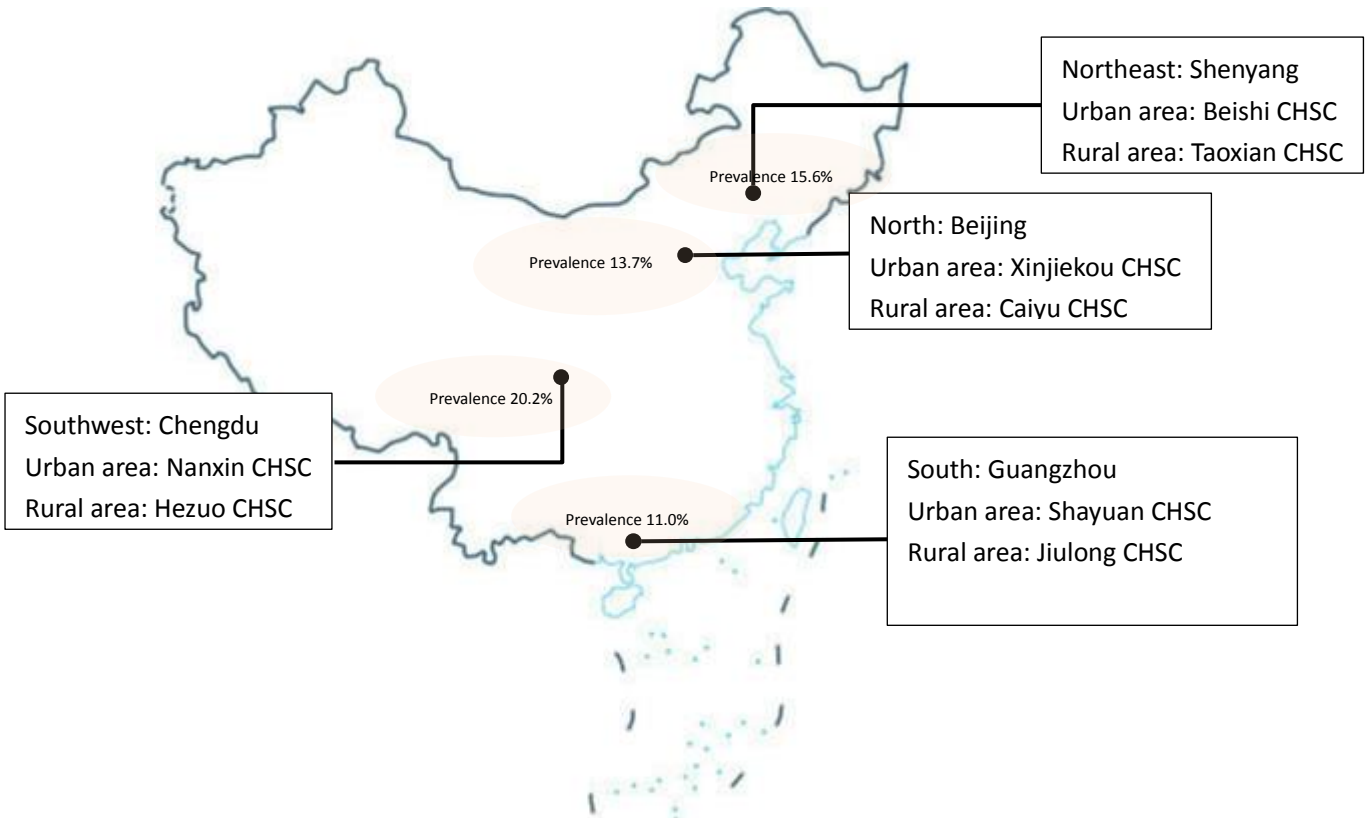
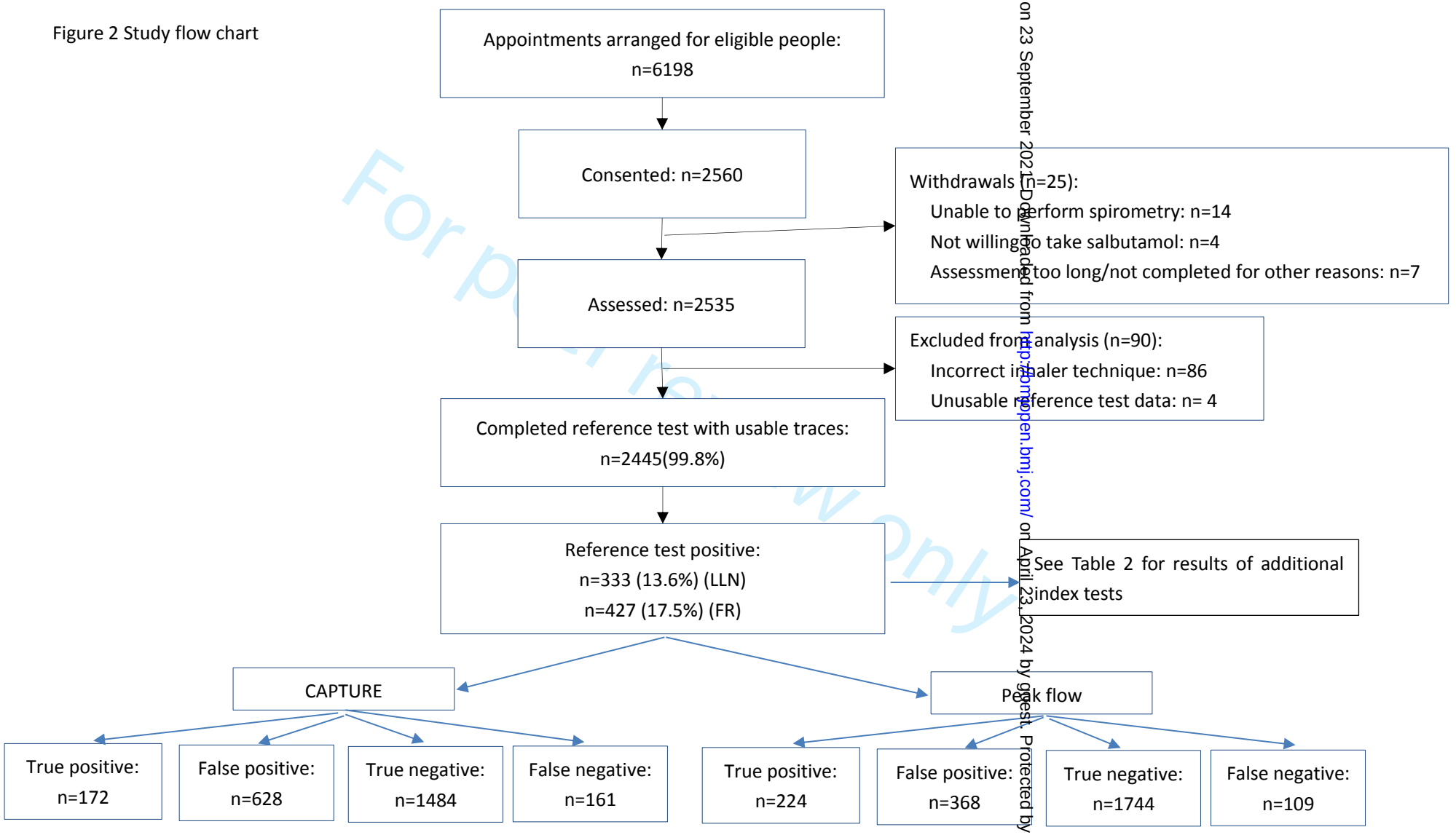


Figure 1 the map of Breathe Well-China research sites

Figure 2 Study flow chart



Evaluating screening strategies for identifying undiagnosed COPD in China: a Breathe Well project

中国慢阻肺筛查策略评估：健康呼吸 Breathe Well 研究项目

Lung health questionnaire

肺部健康问卷

Participant Initials 研究对象编号	
Study ID 问卷编号	
Date 填写日期	
Interviewer ID 研究人员编号	

Some questions in the following booklets may appear similar. However, it is important that we ask these questions in slightly different ways so please complete all questions, answering them as accurately as possible.

一些问题可能相似，但是我们以稍微不同的方式提出这些问题很重要。

因此，请您完成所有的问题，并尽可能准确地作答。

CDQ

1. Age group, years
   年龄

40–49 ☐ 50-59 ☐ 60-69 ☐ 70+ ☐

2. What is your weight in kilograms?
   您的体重（公斤）？

   _____kilograms
   _____ 公斤

   What is your height in meters?
   您的身高（米）？

   _____ metres
   _____ 米

3. Smoking
   吸烟强度，包年

   What is the total number of years you have smoked?
   您一共吸烟多少年？

   _____ years

   _____ 年

   How many cigarettes do you currently smoke each day (or 'did smoke each day' if ex-smoker)?
   目前您每天吸多少支烟？（或，如果是既往吸烟者，过去您每天吸多少支烟？）

   _____ cigarettes

   _____ 支

4. Does the weather affect your cough?
   您的咳嗽是否受天气影响？

Yes ☐ No ☐

是 ☐ 否 ☐

5. Do you ever cough up phlegm (sputum) from your chest when you don't have a cold?
您不感冒的时候，会从胸腔里咳出痰吗？（区别于从嗓子中咳痰）

Yes ☐ No ☐
是 ☐ 否 ☐

6. Do you usually cough up phlegm (sputum) from your chest first thing in the morning?
清晨您的第一件事是从胸腔里咳出痰吗？

Yes ☐ No ☐
是 ☐ 否 ☐

7. How frequently do you wheeze?
您喘息的次数是多少？

Occasionally or more often ☐ Never ☐
有时候或更频繁 ☐ 从不 ☐

8. Do you have or have you had any allergies?
目前或既往您有过敏物吗？

Yes ☐ No ☐
是 ☐ 否 ☐

CAPTURE

1. Have you ever lived or worked in a place with dirty or polluted water or air, smoke or second-hand smoke or dust?
您是否曾经在有脏的或受到污染的水或空气，烟雾或二手烟雾或灰尘的地方生活或工作？

Yes ☐ No ☐
是 ☐ 否 ☐

2. Does your breathing change with seasons, weather or air quality?
您的呼吸是否随着季节、天气或空气质量而变化？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Does your breathing make it difficult to do things such as carry heavy loads, shovel dirt or snow, jog, play tennis or swim?
您的呼吸是否会使您难以进行一些工作，比如提重物，铲土或积雪，慢跑，打网球或游泳等？

Yes ☐ No ☐
是 ☐ 否 ☐

4. Compared to others your age, do you tire easily?
和您的同龄人相比，您是否容易感到疲劳？

Yes ☐ No ☐
是 ☐ 否 ☐

5. In the past 12 months, how many times did you miss work, school, or other activities due to a cold, bronchitis, or pneumonia?
在过去的 12 个月里，您有多少次因感冒、支气管炎或肺炎而错过了工作、学校或其他活动？

0 ☐ 1 ☐ 2 or more ☐
0 ☐ 1 ☐ 2 或以上 ☐

Symptom-based questionnaire

1. How frequently are you exposed to second-hand smoking?
您接触二手烟的频率是多少？

<7hrs per week ☐ ≥7hrs per week ☐
< 7小时/周 ☐ > 7 小时/周 ☐

2. Do you often cough when you do not have a cold?
您是否在不感冒的时候经常咳嗽？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Do you have more signs of shortness of breath compared with others of the same age?
和同龄人相比，您是否有更多的呼吸急促的症状？

Yes ☐ No ☐
是 ☐ 否 ☐

4. Have you had long-term exposure to dust or chemical particles?
您是否长期地接触粉尘或化学颗粒？

Yes ☐ No ☐

是 ☐ 否 ☐

5. Did you have a history of chronic respiratory diseases when you were a child?
在您孩童时期，您是否有慢性呼吸疾病的病史？

Yes ☐是 No ☐
是 ☐ 否 ☐

COPD-SQ

1. Do you often cough?
您是否经常咳嗽？

Yes ☐ No ☐
是 ☐ 否 ☐

2. Family history of respiratory disease
是否有呼吸疾病家族史？

Yes ☐ No ☐
是 ☐ 否 ☐

3. Exposure to biomass smoke from cooking fires
是否接触烹饪产生的生物烟雾？

Yes ☐ No ☐
是 ☐ 否 ☐

Appendix 2: Costs, timings and assumptions for case-finding strategies

Assessment timings	**Minutes per patient**
Symptom questionnaire (completion and processing)	6
Peak flow	2
Microspirometry	4
Confirmatory NDD spirometry	30
Staff	**Hourly costs (UK £)**
Clinic staff	6.25
Additional unit costs	**(UK £)**
Symptom questionnaire	0.10
Peak flow	
Mouthpiece cost per patient	0.10
Overall equipment cost	8.00
Other consumable costs per patient	0.21
Microspirometry (COPD-6)	
Mouthpiece cost per patient	0.10
Overall equipment cost	75.00
Battery cost per year	5.00
Other consumable costs per patient	0.21
Confirmatory NDD spirometry	
Mouthpiece cost per patient	1.30
Overall equipment cost	1,095
Salbutamol cost per patient	0.70
Other consumable and equipment costs per patient	0.25
Assumptions	
Number of visits per year per case finding clinic (assuming 48 tests per day, 5 days a week, 50 weeks a year)	12,000
Number of visits per year per NDD spirometry clinic (assuming 16 tests per day, 5 days a week, 50 weeks a year)	4,000
Lifetime of peak flow meter	1 year
Lifetime of microspirometry	6 years
Lifetime of NDD spirometry	6 years
Proportion of patients requiring staff assistance with questionnaire	95%
Cost of case finding method per patient	**(UK £)**
Symptom questionnaire	0.70
Peak flow	0.52
Microspirometry	0.73
Confirmatory NDD spirometry	4.90

Appendix 3-TABLE 1: SERIAL (AND) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

Strategies	Peak flow	Microspirometry
CAPTURE + peak flow	-25.5 (-30.5,-20.5; <0.0001)	
CDQ + peak flow	-22.8 (-27.6,-18.0; <0.0001)	
C-SBQ + peak flow	-17.4 (-21.8,-13.0; <0.0001)	
COPD-SQ + peak flow	-22.5 (-27.3,-17.7; <0.0001)	
CAPTURE + microspirometry		-27.0 (-32.1,-22.0; <0.0001)
CDQ + microspirometry		-23.4 (-28.3, -18.6; <0.0001)
C-SBQ + microspirometry		-18.3 (-22.8,-13.9; <0.0001)
COPD-SQ + microspirometry		-23.4 (-28.3,-18.6; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, sensitivity for CAPTURE + peak flow is 25.5% lower than for peak flow (95%CI -30.5, -20.5; <0.0001).

Appendix 3-TABLE 2: SERIAL (AND) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

Strategies	Peak flow	Microspirometry
CAPTURE + peak flow	11.1 (9.7, 12.5; <0.0001)	
CDQ + peak flow	10.0 (8.7, 11.4; <0.0001)	
C-SBQ + peak flow	8.7 (7.5, 10.0; <0.0001)	
COPD-SQ + peak flow	9.8 (8.5, 11.2; <0.0001)	
CAPTURE + microspirometry		6.4 (5.3, 7.5; <0.0001)
CDQ + microspirometry		6.3 (5.3, 7.4; <0.0001)
C-SBQ + microspirometry		6.2 (5.1, 7.2; <0.0001)
COPD-SQ + microspirometry		6.7 (5.6, 7.8; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 11.1% higher than for peak flow (95%CI 9.7, 12.5; <0.0001).

Appendix 3-TABLE 3: SERIAL (AND) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

Strategies	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	-9.9 (-13.4, -6.4; <0.0001)			
CAPTURE + microspirometry	-13.8 (-17.8, -9.8; <0.0001)			
CDQ + peak flow		-10.5 (-14.1, -6.9; <0.0001)		
CDQ + microspirometry		-13.5 (-17.5, -9.5; <0.0001)		
C-SBQ + peak flow			-13.2 (-17.2, -9.3; <0.0001)	
C-SBQ + microspirometry			-16.5 (-20.8, -12.2; <0.0001)	
COPD-SQ + peak flow				-10.5 (-14.1, -6.9; <0.0001)
COPD-SQ + microspirometry				-13.8 (-17.8, 9.8; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, sensitivity for CAPTURE is 3.3% lower than for CDQ (95%CI -9.6, 2.9; 0.3245).

Appendix 3-TABLE 4: SERIAL (AND) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	23.4 (21.6,25.3; <0.0001)			
CAPTURE + microspirometry	25.9 (24.0,27.8; <0.0001)			
CDQ + peak flow		14.0 (12.4, 15.5; <0.0001)		
CDQ + microspirometry		17.4 (15.8, 19.1; <0.0001)		
C-SBQ + peak flow			17.1 (15.4, 18.7; <0.0001)	
C-SBQ + microspirometry			21.7 (19.9, 23.5; <0.0001)	
COPD-SQ + peak flow				15.1 (13.5, 16.7; <0.0001)
COPD-SQ + microspirometry				19.1 (17.4, 20.8; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 23.4% higher than for CAPTURE (95%CI 21.6, 25.3; <0.0001).

Appendix 3-TABLE 5: PARALLEL (OR) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

	Peak flow	Microspirometry
CAPTURE + peak flow	9.9 (6.4, 13.4; <0.0001)	
CDQ + peak flow	10.5 (6.9, 14.1; <0.0001)	
C-SBQ + peak flow	13.2 (9.3, 17.2; <0.0001)	
COPD-SQ + peak flow	10.5 (6.9, 14.1; <0.0001)	
CAPTURE + microspirometry		13.8 (9.8, 17.8; <0.0001)
CDQ + microspirometry		13.5 (9.5, 17.5; <0.0001)
C-SBQ + microspirometry		16.5 (12.2, 20.8; <0.0001)
COPD-SQ + microspirometry		13.8 (9.8, 17.8; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies in the column against index tests in the row. For example, sensitivity for CAPTURE + peak flow is 9.9% higher than for peak flow (95%CI 6.4, 13.4; <0.0001).

Appendix 3-TABLE 6: PARALLEL (OR) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **lung function test alone**

	Peak flow	Microspirometry
CAPTURE + peak flow	-23.4 (-25.3, -21.6; <0.0001)	
CDQ + peak flow	-14.0 (-15.5, -12.4; <0.0001)	
C-SBQ + peak flow	-17.1 (-18.7, -15.4; <0.0001)	
COPD-SQ + peak flow	-15.1 (-16.7, -13.5; <0.0001)	
CAPTURE + microspirometry		-25.9 (-27.8, -24.0; <0.0001)
CDQ + microspirometry		-17.4 (-19.1,-15.8; <0.0001)
C-SBQ + microspirometry		-21.7 (-23.5, -19.9; <0.0001)
COPD-SQ + microspirometry		-19.1 (-20.8, -17.4; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against strategies in the row. For example, specificity for CAPTURE + peak flow is 23.4% lower than for peak flow (95%CI -25.3, -21.6; <0.0001).

Appendix 3-TABLE 7: PARALLEL (OR) STRATEGIES (sensitivity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

Strategies	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	25.5 (20.5, 30.5; <0.0001)			
CAPTURE + microspirometry	27.0 (22.0, 32.1; <0.0001)			
CDQ + peak flow		22.8 (18.1, 27.6; <0.0001)		
CDQ + microspirometry		23.4 (18.6, 28.3; <0.0001)		
C-SBQ + peak flow			17.4 (13.0, 21.8; <0.0001)	
C-SBQ + microspirometry			18.3 (13.9, 22.8; <0.0001)	
COPD-SQ + peak flow				22.5 (17.7, 27.3; <0.0001)
COPD-SQ + microspirometry				23.4 (18.6, .28.3; <0.0001)

Note: Values indicate the difference in sensitivity (with 95% CI & p values), comparing strategies tests in the column against strategies in the row. For example, sensitivity for CAPTURE + peak flow is 25.5% higher than for CAPTURE (95%CI 20.5, 30.5; <0.0001).

Appendix 3-TABLE 8: PARALLEL (OR) STRATEGIES (specificity)

Comparing each **combination** (questionnaire & lung function test) against the **questionnaire alone**

	CAPTURE	CDQ	C-SBQ	COPD-SQ
CAPTURE + peak flow	-11.1 (-12.5,-9.7; <0.0001)			
CAPTURE + microspirometry	-6.4 (-7.5, -5.3; <0.0001)			
CDQ + peak flow		-10.0 (-11.4,-8.7; <0.0001)		
CDQ + microspirometry		-6.3 (-7.4, -5.3; <0.0001)		
C-SBQ + peak flow			-8.7 (-10.0, -7.5; <0.0001)	
C-SBQ + microspirometry			-6.2 (-7.2, -5.1; <0.0001)	
COPD-SQ + peak flow				-9.8 (-11.2, -8.5; <0.0001)
COPD-SQ + microspirometry				-6.7 (-7.8, -5.6; <0.0001)

Note: Values indicate the difference in specificity (with 95% CI & p values), comparing strategies in the column against index tests in the row. For example, specificity for CAPTURE + peak flow is 11.1% lower than for CAPTURE (95%CI -12.5, -9.7; <0.0001).

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1~~1~~
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3~~2~~
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4~~3~~
	4	Study objectives and hypotheses	4~~3~~
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5~~3~~
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5~~4~~
	9	Whether participants formed a consecutive, random or convenience series	5~~4~~
Test methods	**10a**	Index test, in sufficient detail to allow replication	5~~4~~
	10b	Reference standard, in sufficient detail to allow replication	6~~5~~
	11	Rationale for choosing the reference standard (if alternatives exist)	6~~5~~
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5-6~~4~~
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6~~5~~
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6~~5~~
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6~~5~~
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7~~5~~-6
	15	How indeterminate index test or reference standard results were handled	NA~~No report~~
	16	How missing data on the index test and reference standard were handled	NA~~No report~~
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	NA~~No report~~
	18	Intended sample size and how it was determined	6~~5~~
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Figure 2
	20	Baseline demographic and clinical characteristics of participants	Table 1
	21a	Distribution of severity of disease in those with the target condition	Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	Table 1
	22	Time interval and any clinical interventions between index test and reference standard	NA~~5~~
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Table 2
	25	Any adverse events from performing the index test or the reference standard	NA~~12~~
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	10-11~~10-11~~
	27	Implications for practice, including the intended use and clinical role of the index test	12~~No report~~
OTHER INFORMATION			
	28	Registration number and name of registry	15~~12~~
	29	Where the full study protocol can be accessed	15~~12~~
	30	Sources of funding and other support; role of funders	15~~12~~

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.