PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	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
AUTHORS	Pan, Zihan; Dickens, Andrew; Chi, Chunhua; Kong, Xia; Enocson, Alexandra; Cooper, Brendan; Adab, Peymane; Cheng, Kar Keung; Sitch, Alice; Jowett, Sue; Adams, Rachel; Correia-de-Sousa, Jaime; Farley, Amanda; Gale, Nicola K.; Jolly, Kate; Maglakelidze, Mariam; Maglakelidze, Tamaz; Martins, Sonia; Stavrikj, Katarina; Stelmach, Rafael; Turner, Alice; Williams, Sian; Jordan, Rachel

VERSION 1 – REVIEW

REVIEWER	Chappell, Francesca University of Edinburgh, Clinical Neurosciences
REVIEW RETURNED	03-May-2021

GENERAL COMMENTS	The authors have written a paper screening tests for undiagnosed COPD for use in China, with the clinical justification that undiagnosed COPD is common in the Chinese population. Please note that this review is restricted to the statistical aspects of the paper as I am not a clinician. Generally, I thought this was a very well written manuscript. The authors have clearly made an effort to conduct and report the study carefully, and I do not have many comments to make. 1. The aim of the study is to detest "undiagnosed COPD", but the participant sample includes 88 people with a previous COPD diagnosis. Should these people be excluded? 2. For the sample size calculation, were the “independent” tests a questionnaire and an airflow measurement device? Did they have actual tests in mind when doing the sample size calculation, presumably the CAPTURE questionnaire and the peak flow meter? Could they be explicit? 3. In the Index Tests section, it says, “Previously defined cut points were used to identify participants at risk of COPD.” Are there references for this? I'm assuming the cut points are standard clinical cut points. 4. The Study Flowchart could be extended to include the TP, FN, TN, and FP of the CAPTURE questionnaire and peak flow meter, as recommended by the authors of the STARD statement.

REVIEWER	Vliegenthart, Rozemarijn Medical University of South Carolina, Department of Clinical Radiology and Nuclear Medicine

REVIEW RETURNED	04-May-2021

GENERAL COMMENTS	In this study, authors investigated different low cost screening techniques for diagnosis of COPD in the general population, in comparison to bronchodilator-mediated spirometry as reference standard. Combination of a short questionnaire with microspirometry was found most sensitive, and more cost-effective than some other combinations. Strengths: COPD is a large problem in China and the rest of the world, which is often underdiagnosed. Early diagnosis in a primary care setting could help early intervention and lifestyle advice. The authors used multiple methods for COPD screening including questionnaires and short air flow exams, and compared to a reference standard for lung function. Extensive (cost) effectiveness analysis including different tests in parallel or sequential. Major comments: The population taking part in the study, seems to a certain extent biased and not completely representative of the general Chinese population. The overall prevalence of smoking is over 50% in Chinese men, and in this cohort, the never smokers amount to two thirds. Also, nearly 60% were female. So probably there is underrepresentation of (smoking) men? This needs more Discussion (representativeness and generalizability), in particular in relation to the prevalence of COPD found. Major factors in lung disease, in particular in non-smokers, are currently not taken into account into the questionnaires (inside fumes due to cooking and passive smoking). Could these be included in the questionnaires for higher sensitivity? The pros and cons of a serial vs parallel approach are not discussed. It may be that a serial approach, with first short questionnaire through digital medium, and only spirometry in case of elevated risk of COPD, could reach more/most individuals. For the suggested approach of parallel testing, an individual should first visit primary care. To what extent do authors think this limits applicability/numbers of individuals that are reached? Specific comments: Abstract: -Objectives should be more specific for this study. F.e. COPD diagnosis is not mentioned. -Results should mention 95% confidence interval for all measures. -Perhaps add a conclusion on cost effectiveness?

REVIEWER	Orso, Massimiliano Umbria Region, Health Planning Service of Perugia
REVIEW RETURNED	01-Jun-2021

GENERAL COMMENTS	I think that this paper is well written; the study objectives are clear and the methods are appropriate. I have just few minor comments/questions to the Authors:

	Page 5, line 23: “3.6% (n=88) had an existing COPD diagnosis”. The study objective is not to individuate undiagnosed COPD? Page 7, line 8: It would be better to start the sentence with “Ninety percent…” Page 7, line 37: Are there cost-effectiveness studies conducted in other countries? If so, their results could be applied in the China setting or there are specific differences? Page 11, line 10: numbers should be spelled out at the start of sentences Page 11, line 25: “3.6% (n=88) had an existing COPD diagnosis”: see my first comment Page 11, lines 27 and following: I would add a brief comparisons of the characteristics of individuals positive to the reference test and individuals negative (e.g. if there are significant differences between the two groups in terms of age, sex, setting, ecc) Page 13, lines 37-39: Could you please explain to me the meaning of this sentence: “we included never smokers in this study to maximise the range of potential cases. Inevitably this contributed to the lower test performance observed” Page 14, lines 45-47: “By including some people with known COPD, we maximised the number of test positives in the study sample”: this sentence explain why you included patients with a diagnosis of COPD. I am still not convinced about this inclusion, due to the objective of the study is to identify undiagnosed patients. Could you clarify this?

VERSION 1 – AUTHOR RESPONSE

Reviewers' comments	Response
Reviewer: 1	
1. The aim of the study is to detest "undiagnosed COPD", but the participant sample includes 88 people with a previous COPD diagnosis. Should these people be excluded?	Our original submission did discuss this, but we have slightly edited the first sentence of the relevant paragraph to improve clarity. It now reads “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.” (page 13) Our study was not designed to be the *implementation* of a screening strategy, rather to determine the accuracy of the screening tests used. In China, there are generally no lists of people with diagnosed disease and patients are frequently not aware of their condition. Therefore, knowledge of whether someone has a diagnosed condition is not available in advance. Within the context of assessing the **accuracy** of screening tests,

	including those with an existing diagnosis is fully justified, as we are comparing tests with a reference test. Furthermore, retaining those with an existing COPD diagnosis improves efficiency and allowed us to maximise the proportion of “test positives” in the sample.
2. For the sample size calculation, were the “independent” tests a questionnaire and an airflow measurement device? Did they have actual tests in mind when doing the sample size calculation, presumably the CAPTURE questionnaire and the peak flow meter? Could they be explicit?	Within the context of calculating sample sizes for test accuracy studies, ‘independence’ refers to the fact that screening questionnaires and airflow measurement devices are measuring different aspects; respiratory symptoms and airflow respectively. We did have CAPTURE and peak flow in mind, but the calculation was for a general comparison of a questionnaire and an airflow measurement device. We have revised the text to make this more explicit. (page 7)
3. In the Index Tests section, it says, “Previously defined cut points were used to identify participants at risk of COPD.” Are there references for this? I'm assuming the cut points are standard clinical cut points.	The cut-points were derived from validation studies of the index tests, and we cited relevant references within the manuscript. Standard clinical cut-points are not available as the index tests are not used extensively in clinical practice.
4. The Study Flowchart could be extended to include the TP, FN, TN, and FP of the CAPTURE questionnaire and peak flow meter, as recommended by the authors of the STARD statement.	We have revised the flowchart to include TP, FP, TN & FN for CAPTURE and peak flow.
Reviewer: 2	
Major comments: The population taking part in the study, seems to a certain extent biased and not completely representative of the general Chinese population. The overall prevalence of smoking is over 50% in Chinese men, and in this cohort, the never smokers amount to two thirds. Also, nearly 60% were female. So probably there is underrepresentation of (smoking) men? This needs more Discussion (representativeness and generalizability), in particular in relation to the prevalence of COPD found.	Although the prevalence of never smokers in study population seems high this is accounted for by the difference in smoking prevalence among men and women. The proportion of never smokers was much lower among men (~30%) compared to women (~96%), which, when combined results in the observed never smoking prevalence overall of 69%. As we state in the Discussion, this closely reflects the findings from a recent nationally representative cross-sectional study in China, which included a younger population, and never smoking rates were 71.4%. We therefore believe our sample is representative of the Chinese population in terms of smoking. The slightly higher proportion of women does mean the sample is slightly less representative. There is no

	reason to expect this to affect the accuracy of screening, but may have led to underestimation of the prevalence of undiagnosed COPD. We have added a sentence to this effect in the discussion (page 13).
Major factors in lung disease, in particular in non-smokers, are currently not taken into account into the questionnaires (inside fumes due to cooking and passive smoking). Could these be included in the questionnaires for higher sensitivity?	The aim of this study was not to develop a new screening questionnaire, but rather to test the accuracy of the numerous existing tools within the China context. We agree with the importance of using screening questionnaires that capture relevant exposures, and intentionally selected questionnaires with this in mind. Passive smoking is included in the symptom-based questionnaire (SBQ) and CAPTURE, while exposure to dust/biomass smoke/air pollution is included in the SBQ, CAPTURE and COPD-SQ. We accept that questionnaire test performance is dependent on the items included, and we are confident that we selected appropriate questionnaires from those available
The pros and cons of a serial vs parallel approach are not discussed. It may be that a serial approach, with first short questionnaire through digital medium, and only spirometry in case of elevated risk of COPD, could reach more/most individuals. For the suggested approach of parallel testing, an individual should first visit primary care. To what extent do authors think this limits applicability/numbers of individuals that are reached?	We agree that the choice of serial/parallel approach is an important decision that has significant healthcare implications. In light of the recent national COPD screening policy in China, we believe a highly sensitive parallel strategy would be preferable in this context. However, we acknowledge that the prioritizing of sensitivity/specificity is a policy decision and dependent on the local healthcare setting, and we have inserted the following sentence reflecting this in the Discussion. “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.” (page 13)
Specific comments: Abstract: -Objectives should be more specific for this study. F.e. COPD diagnosis is not mentioned.	We altered the objectives to “…of various **COPD** screening tests…” (page 3)
-Results should mention 95% confidence interval for all measures.	We have now inserted 95% confidence interval for all measures. (page 3)

-Perhaps add a conclusion on cost effectiveness?	We have added relevant wording on cost-effectiveness in the Conclusion. (page 3)
Reviewer: 3	
Page 5, line 23: “3.6% (n=88) had an existing COPD diagnosis”. The study objective is not to individuate undiagnosed COPD?	Please see our response to reviewer 1, where we have addressed this issue. The objective was not to identify undiagnosed COPD, but to test the accuracy of screening tests.
Page 7, line 8: It would be better to start the sentence with “Ninety percent…”	We have altered the wording as suggested. (page 5)
Page 7, line 37: Are there cost-effectiveness studies conducted in other countries? If so, their results could be applied in the China setting or there are specific differences?	There are relatively few cost-effectiveness studies of COPD screening. Most previous health economic analyses are model based, rather than directly collected field data as in our study. Furthermore, we are not aware of any that have compared cost-effectiveness of different screening strategies from a test-accuracy study. Thus comparison with other studies is complex and our findings are novel in this context. It is not possible to directly apply findings from other settings here. However, since submitting this paper, we are aware of another model-based health economic evaluation of COPD screening from China, which suggests that screening using a questionnaire and hand-held spirometry is cost-saving. We have added this to our discussion. (page 14)
Page 11, line 10: numbers should be spelled out at the start of sentences	We have altered the wording to “We invited 6198 eligible people to the study” (page 9)
Page 11, line 25: “3.6% (n=88) had an existing COPD diagnosis”: see my first comment	Please see our response to reviewer 1, where we have addressed this issue..
Page 11, lines 27 and following: I would add a brief comparisons of the characteristics of individuals positive to the reference test and individuals negative (e.g. if there are significant differences between the two groups in terms of age, sex, setting, ecc)	We have inserted the following sentence when describing the sample “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.” (page 9)
Page 13, lines 37-39: Could you please explain to me the meaning of this sentence: “we included never smokers in this study to maximise the range	In comparison other countries, China has a higher prevalence of COPD amongst never smokers. It was therefore important to include never smokers in our study, so that our

of potential cases. Inevitably this contributed to the lower test performance observed”	findings were generalizable to the wider China population (not only those with a positive smoking history). This also allows for the inclusion of COPD cases caused by risk factors other than smoking, such as environmental exposures (as mentioned by reviewer 2). To improve clarity in the manuscript, we have revised the first sentence you queried: “… 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.” (page 12) If we had restricted the study to ever smokers, the pre-test probability of true COPD would have been higher than our actual study sample (smoking is still the main cause of COPD in China); hence the test performance of index tests would have been higher. While restricting to ever smokers is appropriate in some settings, it is not the case in China for the reasons provided above.
Page 14, lines 45-47: “By including some people with known COPD, we maximised the number of test positives in the study sample”: this sentence explain why you included patients with a diagnosis of COPD. I am still not convinced about this inclusion, due to the objective of the study is to identify undiagnosed patients. Could you clarify this?	As per our response to reviewer 1, our study was not designed to diagnose new cases of COPD (ie not an evaluation of a screening programme, where the outcome is the number of new cases of COPD identified, and inputs would be dependent on who and how people are invited, uptake and attendance for screening); instead, we explored the accuracy of different screening tests and strategies that can be used to identify potential COPD within a screening programme (i.e. the outcome is a comparison of sensitivity and specificity, to inform which test should be implemented in a programme). When assessing the accuracy of screening tests it is justifiable to include those with an existing diagnosis, as we are comparing index tests with a reference test and therefore need sufficient numbers of true cases for the comparison.

VERSION 2 – REVIEW

REVIEWER	Vliegenthart, Rozemarijn Medical University of South Carolina, Department of Clinical Radiology and Nuclear Medicine

REVIEW RETURNED	01-Jul-2021

GENERAL COMMENTS	The authors have addressed my comments in a satisfactory manner. I have no more comments.