The accuracy of PUMA questionnaire in combination with peak expiratory flow rate to identify at-risk, undiagnosed COPD patients

Kang-Cheng Su Yi-Han Hsiao Hsin-Kuo Ko Kun-Ta Chou Tien-Hsin Jeng Diahn-Warng Perng MD PhD

PIL: S0300-2896(24)00234-5
DOI: https://doi.org/doi:10.1016/j.arbres.2024.06.013
Reference: ARBRES 3608
To appear in: Archivos de Bronconeumología
Received Date: 8 April 2024
Accepted Date: 15 June 2024

Please cite this article as: Su K-Cheng, Hsiao Y-Han, Ko H-Kuo, Chou K-Ta, Jeng T-Hsin, Perng D-Warng, The accuracy of PUMA questionnaire in combination with peak expiratory flow rate to identify at-risk, undiagnosed COPD patients, Archivos de Bronconeumología (2024), doi: https://doi.org/10.1016/j.arbres.2024.06.013

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 The Author(s). Published by Elsevier España, S.L.U. on behalf of SEPAR.
The accuracy of PUMA questionnaire in combination with peak expiratory flow rate to identify at-risk, undiagnosed COPD patients

Running title: PUMA-PEFR prediction model for COPD case findings

Kang-Cheng Su\textsuperscript{a,b,c}, Yi-Han Hsiao\textsuperscript{a,c}, Hsin-Kuo Ko\textsuperscript{a,c}, Kun-Ta Chou\textsuperscript{a,b,c}, Tien-Hsin Jeng\textsuperscript{d}, Diahn-Warng Perng\textsuperscript{a,c}

\textsuperscript{a} Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, R.O.C.
\textsuperscript{b} Center of Sleep Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, R.O.C.
\textsuperscript{c} School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, R.O.C.
\textsuperscript{d} Medical Department, Ditmanson Medical Foundation, Chia-Yi Christian Hospital, Chia-Yi, Taiwan, R.O.C.

Correspondence: Diahn-Warng Perng, M.D., PhD,
Professor, School of Medicine, College of Medicine, Yangming Campus, National Yang Ming Chiao Tung University, Taipei, Taiwan. R.O.C.
Chief, Division of General Chest Medicine, Department of Chest Medicine, Taipei Veterans General Hospital, R.O.C., No.201, Sec. 2, Shipai Rd., Beitou District, Taipei, Taiwan 11217, R.O.C.
E-mail: dwperng@vghtpe.gov.tw

Word count of the main manuscript body: 3713

GRAPHICAL ABSTRACT

Highlights

- The Chinese PUMA questionnaire (C-PUMA) is linguistically and clinically validated.
- The best C-PUMA cutoff score is \( \geq 6 \) (sensitivity/specificity/NNS=77%/64%/3).
- C-PUMA \( \geq 5 \) is adequate to detect more COPD patients in high-incidence primary care.
- The PUMA-PEFR model is more accurate and cost-effective than the C-PUMA alone.
- The best P\text{COPD} cutoff value is \( \geq 0.39 \) (sensitivity/specificity/NNS=79%/88%/2).

ABSTRACT

Introduction: The English PUMA questionnaire emerges as an effective COPD case-finding tool. We aimed to use the PUMA questionnaire in combination with peak expiratory flow rate (PEFR) to
improve case-finding efficacy in Chinese population.

Methods: This cross-sectional, observational study included 2 stages: translating English to Chinese PUMA (C-PUMA) questionnaire with linguistic validation and psychometric evaluation, followed by clinical validation. Eligible participants (with age ≥40 years, respiratory symptoms, smoking history ≥10 pack-years) were enrolled and subjected to 3 questionnaires (C-PUMA, COPD assessment test [CAT], and generic health survey [SF-12V2]), PEFR measurement, and confirmatory spirometry. The C-PUMA score and PEFR were incorporated into a PUMA-PEFR prediction model applying binary logistic regression coefficients to estimate the probability of COPD ($P_{\text{COPD}}$).

Results: C-PUMA was finalized through standard forward-backward translation processes and confirmation of good readability, comprehensibility, and reliability. In clinical validation, 240 participants completed the study. 78/240 (32.5%) were diagnosed with COPD. C-PUMA exhibited significant validity (correlated with CAT or physical component scores of SF-12V2, both $P<0.05$, respectively). PUMA-PEFR model had higher diagnostic accuracy than C-PUMA alone (area under ROC curve, 0.893 vs. 0.749, $P<0.05$). The best cutoff values of C-PUMA and PUMA-PEFR model ($P_{\text{COPD}}$) were ≥6 and ≥0.39, accounting for a sensitivity/specificity/numbers needed to screen of 77%/64%/3 and 79%/88%/2, respectively. C-PUMA ≥5 detected more underdiagnosed patients, up to 11.5% (vs. C-PUMA ≥6).

Conclusion: C-PUMA is well-validated. The PUMA-PEFR model provides more accurate and cost-effective case-finding efficacy than C-PUMA alone in at-risk, undiagnosed COPD patients. These tools can be useful to detect COPD early.

Keywords: diagnostic accuracy; linguistic validation; peak expiratory flow rate; prediction model; predictive performance; PUMA questionnaire
ABBREVIATIONS: C-PUMA: Chinese PUMA questionnaire; CAT: COPD assessment test; COPD: chronic obstructive lung disease; DA: diagnostic accuracy; FEV$_1$: forced expiratory volume in the first second; FVC: forced vital capacity; MCS: mental component score; NNS: numbers needed to screen; NPV: negative predictive value; PC: primary care; P$_{COPD}$: probability of COPD; PCS: physical component score; PEFR: peak expiratory flow rate; PPV: positive predictive value; SF-12V2: short-form 12-item health survey version 2; %PEFR: percent predicted peak expiratory flow rate.

Introduction

Chronic obstructive pulmonary disease (COPD) is the seventh leading cause of poor health and the third leading mortality disease worldwide.$^1$ Despite increasing awareness of COPD burden, COPD remains highly underdiagnosed worldwide, ranging from 65%-80% in the community,$^2$ and a substantially high proportion of underdiagnosis occurs in primary care (PC).$^3$-$^5$ Underuse of spirometry is the strongest predictor for underdiagnosis.$^6$ Low symptom perception and insufficient awareness of risk factors by the patients and their clinicians are also important contributing factors.$^2$-$^7$ Additionally, a large-scale study (n=30874) by Lamprecht et al. discovered that COPD underdiagnosis was also associated with male gender, younger age, never and current smoking, lower education, no previous spirometry, and less severe airflow limitation.$^6$ Hence, an effective COPD case-finding strategy to identify at-risk persons for confirmatory spirometry is urgently needed.

Currently, screening for COPD in asymptomatic adults is not recommended by the US Preventive Services Task Force.$^8,^9$ However, case-finding for early COPD identification in those with respiratory symptoms or exposure risks is advocated.$^{10}$ The latter often present to healthcare
facilities, particularly in PC, where are the optimal places to catch COPD early. Case-finding tools in PC include questionnaires, handheld devices, or a combination of both.\textsuperscript{11,12} Symptom-based questionnaires, such as the COPD diagnostic questionnaire,\textsuperscript{13,14} COPD Population Screener questionnaire (COPD-PS),\textsuperscript{15,16} and lung function questionnaire,\textsuperscript{17} are commonly used tools. These questionnaires exerted a poor to fair diagnostic accuracy (DA, indicated by the area under the receiver operating characteristic curve [AUROC] = 0.64-0.71), a wide range of sensitivity (63%-91%), and specificity (45%-70%). Compared with questionnaires, case-finding tools using handheld devices, such as simplified spirometers,\textsuperscript{14,16} or peak expiratory flow meters\textsuperscript{18,19} exhibited better DA (AUROC = 0.80-0.88), similar sensitivity (76%-88%) and higher specificity (72%-95%). Additionally, combined questionnaires and handheld devices improved DA (AUROC = 0.87-0.91).\textsuperscript{14,18,19} Taken together, questionnaires are simple and sensitive but less specific case-finding tools. A questionnaire needs linguistic validation before being widely used in people who speak different languages. Applying handheld devices can elevate specificity and reduce the numbers needed to screen (NNS). Different countries should establish their own case-finding strategies.

Recently, the English PUMA questionnaire (7 items, score 0-9, higher scores indicating a higher diagnostic rate) has emerged as a new effective COPD case-finding tool with acceptable DA and predictive performance (AUROC = 0.76, sensitivity/specificity/NNS = 74%/65%/4 at the cutoff score $\geq 5$) in Latin American countries.\textsuperscript{20} Additionally, Au-Duong et al. reported that a translated Chinese PUMA version (for Cantonese) exhibited similar performance (AUROC = 0.753; sensitivity/specificity = 77%/63% at the cutoff score $\geq 6$) in Hong Kong.\textsuperscript{21} In Taiwan, a recent nationwide telephone interview survey of the general population revealed that COPD was largely underdiagnosed.\textsuperscript{22} Taiwan lacks an effective, validated Chinese questionnaire for early identification of COPD so far. Thus, Chinese PUMA (for Taiwanese) can be a feasible solution. Additionally, monitoring peak expiratory flow rate (PEFR) is widely used in PC for asthma control in Taiwan. The
The present study's primary objective was to form a linguistically validated Chinese PUMA (C-PUMA, Mandarin version) and examine its DA with or without the combination of PEFR in Taiwan. The other objectives included the reliability and validity of C-PUMA and the predictive performance of different diagnostic modalities.

**Methods**

*Study design*

This cross-sectional, observational study was conducted in a medical center in Taiwan, from December 2019 and April 2022. The study comprised 2 stages (Fig. 1), including linguistic validation and psychometric evaluation based on the guideline\(^23,24\) to form the finalized C-PUMA, which was applied for clinical validation in the targeted population (validation cohort). The linguistic validation involved the translation of the English PUMA to the traditional Chinese version and the adaption of Chinese culture. The psychometric evaluation examined the reliability and validity of the translated questionnaire. The targeted population was invited to pulmonary outpatient clinics, where their demographic information, questionnaires of C-PUMA, COPD assessment test (CAT, Chinese version, licensed from the Mapi Research Trust organization), short-form 12-item health survey (SF-12V2, traditional Chinese version 2, licensed from the QualityMetric Inc.), PEFR measurements, and confirmatory post-bronchodilation spirometry were obtained (Fig. 2A). All participants completed the study flow on the same day. This study was approved by the Institutional Review Board (approval number: 2019-05-009CC and 2019-08-007AC). All participants signed informed consents.

*Study participants*

The targeted population was at-risk persons without a previous diagnosis of COPD. They presented to our medical center-affiliated pulmonary clinics because of chronic respiratory symptoms
(e.g. cough, dyspnea, phlegm). These persons might come from the community without any referrals or being referred by non-pulmonologists. They were enrolled if they met all the following criteria: aged ≥ 40 years; current or ex-smokers with a smoking history of ≥10 pack-years, or ≥ 50 pipes/year, or ≥ 50 cigars/year, and/or biomass smoke (wood or coal, for cooking or heating) exposure ≥ 100 h/year; reporting any respiratory symptoms. The exclusion criteria are listed in Appendix A. The PEFR was measured using a Mini-Wright peak flow meter (Micropeak, Micro Medical Limited, Rochester, UK) before bronchodilation according to the ERS recommendations. The best PEFR was adopted from three correct blows when participants exerted maximal expiratory efforts in a standing position. Spirometry measurement was in accordance with the standards from the American Thoracic Society/European Respiratory Society (Appendix A). The diagnosis of COPD, stage of airflow limitation, and COPD grouping were based on the 2017 Global Initiative for Chronic Obstructive Lung Disease (GOLD) report.

Translation and formation of the Chinese PUMA questionnaire

The original English PUMA questionnaire (Appendix B) was translated to the Chinese versions using forward-backward translation methods with the assistance of the professional TransPerfect Translations, Inc. (New York, USA). The translation process is illustrated in Figure 1 and detailed in Appendix A. The final format of C-PUMA was determined after the pilot test.

Reliability and validity of the translated PUMA questionnaire

The reliability was examined by test-retest reproducibility and internal consistency, and the validity was presented as concurrent and predictive validity. The details are shown in Table 1. In the pilot test, the confusing wording was revised through the face-to-face feedback from the first 10 respondents, and the test-retest reproducibility was examined in 35 eligible participants who repeated to answer the draft Chinese PUMA (ChV2, Fig. 1) one or two weeks apart. Afterward, the C-PUMA
format was finalized. In the validation cohort, the concurrent validity tested the associations of the C-PUMA with accepted standards, such as the CAT, because CAT is a COPD-specific questionnaire representing symptom burdens, and CAT was reported to act as a COPD case-finding tool. The predictive validity indicates the ability of C-PUMA to predict health status, represented by the SF-12V2. The SF-12V2 measures generic health-related quality of life and is presented with physical and mental component summaries (PCS and MCS, respectively; higher scores indicate better health, Appendix A).

**Combined variables to generate a COPD prediction model**

A binary logistic regression model using the enter method was applied to examine the independent variables related to COPD diagnosis and to generate an equation for estimating the probability of COPD (P\textsubscript{COPD}). Subsequently, a logit model was generated using the independent variables to calculate P\textsubscript{COPD}. The log odds ratio of participants with or without COPD is expressed as follows:

\[
\ln[P_{COPD}/(1-P_{COPD})] = \text{logit}(P_{COPD}) = f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i
\]

where \(\beta_0\) is the coefficient of the constant, and \(\beta_i\) is the coefficient(s) of the independent variable(s) \(X_i\). This equation can be transformed as follows:

\[
P_{COPD} = e^{f(x)}/ (1 + e^{f(x)})
\]

Where \(P_{COPD}\) can be directly calculated by computer software, and this model can serve as a case-finding tool.

**Statistical analysis**

Based on the recent review for cross-cultural adaption and psychometric validation research, and the epidemiological data from our similar study, the sample size was 35 for the pilot test and 240 for the validation cohort (Appendix A). The test-retest reproducibility was examined by the Pearson correlation coefficient and the intraclass correlation coefficient (ICC,
deemed excellent between 0.75 and 0.9).\textsuperscript{29,30} The internal consistency was determined by Cronbach’s \(\alpha\) coefficient (moderate if > 0.5; high if > 0.7).\textsuperscript{29} The concurrent and predictive validity was determined by Spearman’s rank correlation coefficient\textsuperscript{29,31} between the C-PUMA and CAT or SF-12V2, respectively.

Data are presented as means \(\pm\) SD or median (interquartile) or number (%), as appropriate. Continuous variables are compared using a t-test or Mann-Whitney U test. Categorical data were evaluated by a Chi-square test. The performance of different diagnostic modalities was determined and compared using AUROC analysis. An AUROC of 0.6-0.7, 0.7-0.8, 0.8-0.9, and 0.9-1.0 indicates poor, fair, good, and excellent DA, respectively.\textsuperscript{32} The best cutoff value of different modalities was calculated using the Youden index to determine the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and NNS. Statistical analysis was performed using SPSS Statistics for Windows, Version 20.0 (IBM Corp., Armonk, NY, USA). The comparisons of AUROC values were performed using MedCalc version 17.5.5 (MedCalc Software bvba, Ostend, Belgium). A two-sided \(P\) value < 0.05 was considered significant.

Results

\textit{Linguistic validation and psychometric evaluation of C-PUMA}

During the translation process, the expert panels selected the plainest Chinese expressions to maximize their readability and comprehensibility. In the pilot test, the first 10 participants indicated the ChV2 was clear, easily understood, and had no comments on its format during face-to-face feedback, and 35 participants confirmed its reliability. Thereby, the C-PUMA format was finalized (Appendix C). The reliability and validity are summarized in Table 1.

\textit{Patient characteristics in the validation cohort}
286 consecutive participants were invited, and 240 completed the study (Fig. 2A) without missing data. 78 of 240 (32%) were newly diagnosed COPD patients, who were categorized by stage I (37, 47%), II (34, 44%), and III (7, 9%) or by group A (60, 77%), B (15, 19%), C (2, 3%) and D (1, 1%), respectively. Compared with non-COPD cases, COPD patients were older and presented higher C-PUMA scores, more symptom burdens (higher CAT and less SF-12-PCS scores), and lower lung function parameters (Table 2). Based on the cutoff values of C-PUMA and PEFR, the participants were categorized into 4 subgroups: PUMA-/PEFR-, PUMA+/PEFR-, PUMA-/PEFR+, and PUMA+/PEFR+ (Fig. 2B), and their baseline characteristics were shown in Table A1 in Appendix A. The PUMA+/PEFR+ subgroup identified the most COPD patients (43 of 78, 55%) and contained the highest COPD proportion (43 of 49, 87.8%) across these 4 subgroups. Additionally, the majority of participants (132, 55%) directly came from the community without any referrals, followed by referrals from PC (86, 35.8%) and from non-pulmonary clinics (22, 9.2%) in the same hospital, respectively.

**Diagnostic accuracy and predictive performance of C-PUMA and PEFR**

C-PUMA demonstrated an acceptable DA (AUROC = 0.749) (Fig. 3) and the best cutoff score ≥ 6 (Table 3). Compared with score ≥ 5, which was the best cutoff score reported by the original English PUMA, score ≥ 6 demonstrated lower sensitivity, higher specificity, less NNS (Table 3), and higher accuracy (true positive plus true negative rate, 68.3% vs. 54.6%), but misdiagnosed more COPD patients up to 11.5% (false negatives, 18 vs. 9, Fig 4A and 4B). Looking deeply into the participants’ score distribution (Fig. 4C), 88.5% of COPD patients had scores ≥ 5, and 84.6% of non-COPD cases presented scores ≤ 6. Interestingly, a similar portion of COPD and non-COPD cases had their scores overlapped (score 5-6: 43.6% vs. 46.3%; score 4-7: 75.6% vs. 79%). Compared with C-PUMA, PEFR exhibited higher DA (Fig. 3) and predictive performance (Table 3) with the cutoff value of percent predicted PEFR (% PEFR) < 79%, which value was recently validated in another
similar case-finding study in Taiwan.18

Integrate PUMA and PEFR to form a prediction model

The independent factors to predict the diagnosis of COPD are listed in Table A2 (Appendix A). The univariate logistic regression indicated that age, smoking pack years, best PEFR, %PEFR, CAT, C-PUMA, and SF-12V2-PCS scores were significant variables. Among these, collinearity existed in the following pairs: best PEFR-%PEFR, C-PUMA-age, C-PUMA-smoking pack-years, C-PUMA-CAT, C-PUMA-SF-12-PCS. Thus, we adopted the C-PUMA score and %PEFR for multivariate logistic regression, and both variables were statistically independent and remained in the model (Table A1). The logit model, which incorporated PUMA score and %PEFR into a prediction model, namely the PUMA-PEFR model, was expressed as follows:

\[
\text{logit}(P_{\text{COPD}}) = f(x) = 4.274 + (0.493 \times \text{PUMA}) + (-0.091 \times \%\text{PEFR}).
\]

The aforementioned equation was transformed as follows:

\[
P_{\text{COPD}} = \frac{\exp[4.274 + (0.493 \times \text{PUMA}) + (-0.091 \times \%\text{PEFR})]}{1 + \exp[4.274 + (0.493 \times \text{PUMA}) + (-0.091 \times \%\text{PEFR})]}.
\]

The estimated \(P_{\text{COPD}}\) can be readily calculated by entering the 2 variables into preset computer software, as demonstrated in Table 4. The AUROC comparisons indicated that the PUMA-PEFR model \(P_{\text{COPD}}\) exhibited the best DA (Fig. 3). The best cutoff value for the PUMA-PEFR model is \(P_{\text{COPD}} \geq 0.39\), accounting for high accuracy (85.4%, Appendix A, Fig. A2) and good predictive performance (Table 3).

Discussion

The present study applied the standard forward-backward translation processes to form the C-PUMA, followed by confirmation of reliability, validity, and clinical performance. C-PUMA alone
showed fair DA, which was enhanced by the PUMA-PEFR prediction model. These tools might act as effective COPD case-finding tools in PC.

Linguistic validation and psychometric evaluation are necessary before a translated questionnaire is widely used in the targeted population. Although another Chinese PUMA questionnaire (HK-PUMA, for Cantonese) was reported by Au-Doung PLW et al. in Hong Kong, the details of the validation process were not reported. In the present study, excellent test-retest reproducibility (Table 1) indicates that C-PUMA measures a non-biased concept over time, and moderate-to-high internal consistency reflects that each item of C-PUMA constantly links to COPD diagnosis. This process ensured that C-PUMA was qualified before use in clinical validation. Additionally, C-PUMA exhibited weak but significant association with CAT (concurrent validity) and SF-12V2-PCS (predictive validity), indicating C-PUMA potentially links to COPD-specific symptom burdens and generic physical health status, respectively. Therefore, the translated C-PUMA is well-linguistically validated and suitable for use in the Chinese population.

Currently, PUMA has been reported in Latin American countries (original and externally validated English version), and in Hong Kong (Chinese validation for Cantonese). Compared with these PUMA studies (see the details in Table A3, Appendix A), our study had a similar study design except for the two major differences- respiratory symptoms and study settings. It might be concerned that in addition to age and smoke (tobacco or biomass) exposure, this study added respiratory symptoms to the inclusion criteria. The previous PUMA studies usually recruited subjects during routine spontaneous or scheduled appointments unrelated to the study in PC, irrespective of respiratory symptoms. However, the US Preventive Services Task Force recommended against screening for COPD in asymptomatic adults in 2016, which was reaffirmed in 2022. In contrast, the GOLD guideline recommended case-finding in symptomatic patients. The present study was conducted in a medical center, and patients seldom presented to our pulmonologist clinics without
respiratory symptoms. Moreover, these symptomatic patients might benefit from timely diagnosis and treatment. This concept was proved in a very recent multicenter, randomized, controlled trial reported by Aaron et al.\(^3\)\(^4\) They used a case-finding method to identify adults in the community with respiratory symptoms without a previous diagnosis of COPD or asthma. Patients with confirmed diagnosis was assigned to receive either pulmonologist-directed treatment or usual care for 1 year. They found that the interventional group had less healthcare utilization for respiratory illness in comparison with the usual-care group. Therefore, respiratory symptoms might play an important role while initiating a case-finding strategy for early COPD detection, and these case-finding tool-detected patients had beneficial health outcomes from pulmonologist care.

The other difference is clinical settings. The impact of clinical settings on the case-finding efficacy of any specific questionnaire has not been reported to date. Interestingly, the DA of our hospital-based PUMA study (AUROC 0.749) was similar to that in previously PC-based (AUROC 0.704-0.753) or population-based (AUROC 0.734) studies (Table A3). As to predictive performance, in terms of the best cutoff score/sensitivity/specificity, our data (≥ 6/77%/64%) are almost identical to that in the HK-PUMA cohort (≥ 6/76.5%/63.3%).\(^2\)\(^1\) However, there are variations in sensitivity (51.5%-77%), specificity (62.1%-81.6%), PPV (39.1%-59.9%), and NPV (63.3%-88%) across these PUMA studies (Table A3). Different clinical settings and COPD incidence (18.7%-45.1%) are the possible contributing factors. Collectively, PUMA exhibited similar DA but variable performance, indicating PUMA can be reliably used in English and Chinese populations, but different countries/regions must establish their own cutoff value before being widely used.

In the present study, the best C-PUMA cutoff score is ≥ 6, which had a more cost-effective NNS of 3 (vs. ≥ 5, NNS = 4), but at a cost of missing more COPD patients up to 11.5 %. Recently, CH Lin et al. reviewed different case-finding approaches and found that COPD incidence in PC ranged from 10%-26% in those aged ≥ 40 years with a smoking history.\(^1\) Considering the high incidence of
COPD in at-risk patients (e.g., 32% in the present study and 48.8% in our previous report\textsuperscript{18}) and large underdiagnosis in the general population in Taiwan,\textsuperscript{22} we suggest the C-PUMA cutoff score $\geq 5$ is appropriate for generalized use in PC to reduce COPD underdiagnosis if applying C-PUMA alone. Similarly, the HK-PUMA study also proposed that a cutoff score $\geq 5$, which elevated sensitivity (91.2%) and NPV (92.7%), was more suitable than the Youden index-determined cutoff score $\geq 6$ for their early detection strategy.\textsuperscript{21} Interestingly, the best English PUMA cutoff score also varied between $\geq 5$ and $\geq 6$ (Table A3).\textsuperscript{20,33} This might be speculated based on our data that a substantial and equal portion of COPD and non-COPD cases had their scores overlapped between this range. That is, the cutoff score of 5 or 6 has limited discrimination between COPD and non-COPD cases. Therefore, a secondary tool, such as PEFR, can enhance case discrimination in a case-finding approach.

This study found that PEFR was more accurate than C-PUMA, which was consistent with the recent studies that reported PEFR was also better than different questionnaires.\textsuperscript{18,19,35} However, the cutoff values of PEFR might largely vary across different studies, such as $\%\text{PEFR} < 79\%$\textsuperscript{18} or $< 80\%$,\textsuperscript{36} PEFR $< 350$, or $< 250$ L/min (for males and females, respectively).\textsuperscript{19} By contrast, COPD-PS was found to be more accurate than PEFR (cutoff value $\leq 2.2$ l/s m$^2$).\textsuperscript{37} These discrepancies might be ascribed to different peak flow meters, clinical settings (PC vs. hospital), and various COPD incidences (17.4%-49%) (summarized in Table A4, Appendix A). Additionally, Perez-Padilla et al. reported that $\%\text{PEFR} \geq 70\%$ could effectively exclude those with severe airflow limitation (FEV$_1$ $< 50\%$ predicted).\textsuperscript{38} Jithoo et al. suggested using PEFR with appropriate cutoff values could effectively reduce confirmatory spirometry numbers.\textsuperscript{39} Take together, PEFR is a useful case-finding tool but needs validation accordingly.

Combined various questionnaires and PEFR has been reported to serve as effective case-finding approaches in different study designs. In Spain, Soriano et al. reported that a combined modality
selecting those with COPD-PS score ≥ 4 and pre-bronchodilation PEFR ≤ 2.2 l/s.m² exhibited higher DA (AUROC 0.761) than that with COPD-PS (AUROC 0.715) or PEFR (AUROC 0.664) alone. This combined modality can reduce 90% of spirometry tests. In the US, Martinez et al. developed the novel CAPTURE questionnaire and classified the risk levels by scores (low: 0-1; middle:2-4; high 5-6) or by PEFR (high: males < 350 L/min, females < 250 L/min) for identification of clinically significant COPD (those with predicted FEV₁ < 60% and/or exacerbation risk). They reported that the combined modality resulted in the best DA (AUROC 0.9057), followed by PEFR (AUROC 0.8783) and CAPTURE (AUROC 0.7954). They concluded that those with middle scores should measure PEFR and those with high scores or middle scores with low PEFR should undergo spirometry. Identical case-finding protocol using the translated Spanish CAPTURE plus PEFR in those who spoke Spanish replicated identical results. Following these case-finding approaches, clinicians should independently evaluate two separate parameters (questionnaire scores and PEFR) to determine who should be subjected to confirmatory spirometry. This approach is likely to miss COPD patients who lack symptom perception (less symptom scores, such as case 2 in Table 4) or have preserved or mildly impaired lung function (less PEFR reduction, such as case 4 in Table 4). A prediction model possibly minimizes this risk. A prediction model can incorporate different factors into a single number, which number represents an integrated evaluation. Various COPD prediction models have been used to predict the risks of underdiagnosis and the probability of diagnosis. Recently, we incorporated age, CAT scores, smoking pack-years, and %PEFR to predict P_{COPD}. This integrated model exhibited the best DA (AUROC 0.866), compared to CAT (AUROC 0.666) or % PEFR (AUROC 0.832) alone. In this study, the integrated PUMA-PEFR model exhibited the best DA again. However, regarding combined questionnaires and PEFR, the evidence of whether a prediction model exerts better case-finding efficacy than evaluating 2 independent variables is lacking. Clinicians can develop their own combined modality based on available tools.
The strength of this study is that C-PUMA was well-translated and validated before use in clinical validation. There are some limitations. First, the data were driven in a medical center, not in PC. Actually, the healthcare system in Taiwan lacks mandatory referral regulations, and any outpatient has the freedom to visit any specialist in any hospital without a referral. Therefore, in the present study, 90.8% of participants came from the community, that is, our data will be very similar to those generated in PC. Nevertheless, validation in genuine PC is required to expand its generalizability. Second, a substantial portion of newly diagnosed COPD patients had mild disease (Group A 77% or stage I 47%). Although new evidence has shown that case-finding tool-detected patients might benefit from specialist care, the optimal medication choice has not been well-established.

**Conclusions**

Our data are in line with the conclusions of a recent review by Aaron SD et al. that the use of combined tools has shown better diagnostic accuracy than either tool alone and that the combined use of a questionnaire and a handheld device is a more effective strategy for identifying individuals at increased risk for COPD, although questionnaires used alone are still valuable tools to predict COPD. In Taiwan, C-PUMA is well linguistically and clinically validated. The best cutoff score is $\geq 6$. However, the cutoff score $\geq 5$ is considered appropriate to detect more undiagnosed COPD patients if applying C-PUMA alone. Furthermore, the PUMA-PEFR prediction model exhibits the best DA. Applying these new tools in PC can provide an opportunity to reduce COPD underdiagnosis.

**Ethical issue:** This study was approved by the Institutional Review Board of Taipei Veterans General Hospital (approval number: 2019-05-009CC and 2019-08-007AC). All participants signed informed consent.
**Funding/support:** This work was supported by AstraZeneca Inc. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Authors’ contributions:** Conception and design: D.W. Perng, K.C. Su, Y.H. Hsiao; Grant application: D.W. Perng and K.C. Su; acquisition, analysis, or interpretation of data: K.C. Su, H.K. Ko, K.T. Chou, Y.H. Hsiao, and T.H. Jeng; drafting of the manuscript: K.C. Su, H.K. Ko, Chou, Y.H. Hsiao, and T.H. Jeng; critical revision of the manuscript for important intellectual content: D.W. Perng. All the authors discussed the results and approved the final version of the manuscript.

**Prior abstract presentation:** Poster presentation, 2023 ATS conference, Washington DC, May 21, 2023

**Conflicts of Interest**

None declared.

**Competing interests:** None declared.

**Data availability statement:**

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors thank all study participants. We are grateful to the statistical modeling consultant, Dr. Jack Chen, from the ESTAT Statistical Consulting Co., Ltd., for his
assistance with statistical analysis. We also thank TransPerfect Translations, Inc. for their professional language translation and the Mapi Research Trust organization for the kind permission of CAT to use in this study.

consent.
References

https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)
2023 [accessed 31 March 2024].


https://doi.org/10.1136/thx.2007.085456.


https://doi.org/10.2147/COPD.S206250.


Fig. 1 Framework for linguistic and clinical validation. C-PUMA, the finalized version of the Chinese PUMA questionnaire; EnV1, English version 1; ChV0, Chinese version 0, and so forth; PEFR, peak expiratory flow rate.
Figure 2. Study flowchart of participants (A) and the distribution of participants categorized by the cutoff values of the PUMA questionnaire and peak expiratory flow rate (B). An abnormal CXR includes overt bronchiectasis (n= 6), interstitial lung disease (n= 5), pleural effusion (n= 2), lobar infiltration (n= 4), pneumoconiosis (n=1), and lung mass (n= 4). Questionnaires include a COPD assessment test, a Chinese PUMA questionnaire (C-PUMA), and a short-form 12-item health survey, Chinese version 2. PUMA+ and PEFR+ indicate C-PUMA scores ≥ 6 and PEFR < 79% predicted value, respectively.
Figure 3. Diagnostic accuracy according to the ROC curve analysis. The ROC curve and AUROC value of the 3 diagnostic modalities in the validation cohorts. *$P < 0.05$, vs. PUMA; #$P < 0.05$ vs. %PEFR. Statistical evaluations were performed using MedCalc. ROC, receiver operating characteristic curve; AUROC, area under the ROC; CI conference interval; %PEFR, percent predicted peak expiratory flow rate; P_{COPD}, probability of COPD.
Figure 4. Distributions of study participants based on the PUMA cutoff score of 5 (A), 6 (B), and individual PUMA scores from 1 to 9 (C). The proportion of summation numbers in participants with PUMA scores ranging from 5 to 6 and from 4 to 7 in COPD or non-COPD groups is shown in the left upper corner of panel C.
### Table 1 Reliability and validity of translated Chinese PUMA in the pilot test and validation cohort

<table>
<thead>
<tr>
<th></th>
<th>Numbers</th>
<th>Statistical analysis</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Grade</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-retest reproducibility(^a)</td>
<td>35 (pilot test)</td>
<td>Pearson correlation</td>
<td>0.83</td>
<td>&lt; 0.001</td>
<td>very strong</td>
<td>test 1 correlated with test 2</td>
</tr>
<tr>
<td></td>
<td>35 (pilot test)</td>
<td>Intraclass correlation</td>
<td>0.90</td>
<td>&lt; 0.001</td>
<td>excellent</td>
<td>test 1 correlated with test 2</td>
</tr>
<tr>
<td>Internal consistency(^b)</td>
<td>35 (pilot test- test 1)</td>
<td>Cronbach’s α</td>
<td>0.64</td>
<td>-</td>
<td>moderate</td>
<td>the average inter-correlation among the C-PUMA items</td>
</tr>
<tr>
<td></td>
<td>35 (pilot test- test 2)</td>
<td>Cronbach’s α</td>
<td>0.72</td>
<td>-</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td></td>
<td>240 (validation cohort)</td>
<td>Cronbach’s α</td>
<td>0.63</td>
<td>-</td>
<td>moderate</td>
<td></td>
</tr>
<tr>
<td><strong>Validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concurrent validity(^c)</td>
<td>240 (validation cohort)</td>
<td>Spearman rho</td>
<td>0.25</td>
<td>&lt; 0.001</td>
<td>weak</td>
<td>C-PUMA correlated with CAT</td>
</tr>
<tr>
<td>Predictive validity(^d)</td>
<td>240 (validation cohort)</td>
<td>Spearman rho</td>
<td>-0.30</td>
<td>&lt; 0.001</td>
<td>weak</td>
<td>C-PUMA correlated with SF-12V2-PCS</td>
</tr>
<tr>
<td></td>
<td>240 (validation cohort)</td>
<td>Spearman rho</td>
<td>0.24</td>
<td>0.711</td>
<td></td>
<td>C-PUMA correlated with SF-12V2-MCS</td>
</tr>
</tbody>
</table>

\(^a\)The test-retest reproducibility, indicating the temporal stability of the translated PUMA (ChV2), reflected the extent to which the participants’ responses remained consistent on different occasions.

\(^b\)The internal consistency examined the homogeneity of the ChV2, which reflected whether the questionnaire items were consistent in measuring the same construct (the probability of being diagnosed with COPD).

\(^c\)The concurrent validity tested the associations of the C-PUMA with accepted standards, such as CAT, for COPD-specific symptoms.

\(^d\)The predictive validity indicates the ability of C-PUMA to predict health status, such as SF-12V2.

CAT: COPD assessment test; MCS: mental component score; PCS: physical component score; SF-12V2: short-form 12-item health survey version 2.
<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Non-COPD</th>
<th>COPD</th>
<th>P-value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers (%)</td>
<td>240</td>
<td>162 (68)</td>
<td>78 (32)</td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>64 ± 12</td>
<td>61 ± 11</td>
<td>69 ± 11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male, N (%)</td>
<td>224 (93)</td>
<td>150 (93)</td>
<td>74 (95)</td>
<td>0.592</td>
</tr>
<tr>
<td>BMI</td>
<td>25.7 (23.7, 28.3)</td>
<td>26 (24.2, 28.7)</td>
<td>25.1 (23.1, 27.3)</td>
<td>0.186</td>
</tr>
<tr>
<td>Current smoker, N (%)</td>
<td>76 (32)</td>
<td>52 (32)</td>
<td>24 (31)</td>
<td>0.118</td>
</tr>
<tr>
<td>Smoking pack-years</td>
<td>34 (19, 50)</td>
<td>30 (16, 42)</td>
<td>41 (26, 56)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Biomass exposure, N (%)</td>
<td>16 (6.7)</td>
<td>8 (4.9)</td>
<td>8 (10.3)</td>
<td>0.165</td>
</tr>
<tr>
<td>Questionnaire score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-PUMA</td>
<td>5 (4, 7)</td>
<td>5 (4, 6)</td>
<td>6 (6, 7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CAT</td>
<td>4 (2, 6)</td>
<td>3 (1, 6)</td>
<td>4 (3, 8)</td>
<td></td>
</tr>
<tr>
<td>SF-12V2 PCS</td>
<td>63 (52, 68)</td>
<td>65 (58, 69)</td>
<td>53 (43, 65)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SF-12V2 MCS</td>
<td>88 (74, 97)</td>
<td>88 (71, 95)</td>
<td>88 (78, 99)</td>
<td>0.184</td>
</tr>
<tr>
<td>Peak flow meter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best PEFR, L/min</td>
<td>460 (363, 530)</td>
<td>500 (440, 560)</td>
<td>330 (245, 400)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PEFR, % pred.</td>
<td>91 (77, 104)</td>
<td>98 (89, 107)</td>
<td>70 (54, 81)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pre-BD spirometry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, L</td>
<td>2.45 ± 0.72</td>
<td>2.77 ± 0.54</td>
<td>1.79 ± 0.58</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, % pred.</td>
<td>92 ± 21</td>
<td>101 ± 15</td>
<td>74 ± 21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC, L</td>
<td>3.31 ± 0.77</td>
<td>3.49 ± 0.69</td>
<td>2.93 ± 0.81</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC, % pred.</td>
<td>92 ± 16</td>
<td>95 ± 13</td>
<td>86 ± 19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;/FVC ratio</td>
<td>74 ± 11</td>
<td>79 ± 5</td>
<td>61 ± 10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Post-BD spirometry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, L</td>
<td>2.54 ± 0.71</td>
<td>2.83 ± 0.55</td>
<td>1.93 ± 0.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;, % pred.</td>
<td>95 ± 20</td>
<td>103 ± 14</td>
<td>79 ± 20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC, L</td>
<td>3.42 ± 0.75</td>
<td>3.55 ± 0.69</td>
<td>3.16 ± 0.82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC, % pred.</td>
<td>95 ± 16</td>
<td>96 ± 15</td>
<td>93 ± 17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;/FVC ratio</td>
<td>74 ± 11</td>
<td>80 ± 5</td>
<td>61 ± 10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Bronchoreversibility (+)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>28 (12)</td>
<td>4 (3)</td>
<td>24 (31)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data are presented as numbers (%) for categorical variables, or median (interquartile range) for continuous non-parametric variables, or mean ± SD for continuous parametric variables.

<sup>a</sup> P-value is estimated with Chi-square test for categorical variables, Mann-Whitney U test for continuous non-parametric variables, or T-test for continuous parametric variables, respectively.

<sup>b</sup> Positive bronchoreversibility indicates an increase in FEV<sub>1</sub> and/or FVC ≥ 200 mL and ≥ 12% of baseline level after bronchodilation.

BD: bronchodilation; BMI: body mass index; C-PUMA: Chinese PUMA questionnaire; CAT: COPD assessment test; FEV<sub>1</sub>: forced expiratory volume in first second; FVC: forced vital capacity; MCS: mental component score; PCS: physical component score; PEFR: peak expiratory flow rate; SF-12V2: short-form 12-item health survey questionnaire; % pred.: percent predicted value.
Table 3 Predictive performance of different diagnostic modalities in the validation cohort

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Youden index</th>
<th>NNS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PUMA score (C-PUMA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUMA ≥ 4</td>
<td>97</td>
<td>17</td>
<td>36</td>
<td>93</td>
<td>0.15</td>
<td>3</td>
</tr>
<tr>
<td>PUMA ≥ 5</td>
<td>88</td>
<td>38</td>
<td>41</td>
<td>87</td>
<td>0.27</td>
<td>4</td>
</tr>
<tr>
<td><strong>PUMA ≥ 6a</strong></td>
<td><strong>77</strong></td>
<td><strong>64</strong></td>
<td><strong>51</strong></td>
<td><strong>85</strong></td>
<td><strong>0.41</strong></td>
<td><strong>3</strong></td>
</tr>
<tr>
<td>PUMA ≥ 7</td>
<td>45</td>
<td>85</td>
<td>58</td>
<td>76</td>
<td>0.29</td>
<td>3</td>
</tr>
<tr>
<td><strong>PEFR predicted value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%PEFR &lt; 79%b</td>
<td>84</td>
<td>86</td>
<td>69</td>
<td>94</td>
<td>0.70</td>
<td>2</td>
</tr>
<tr>
<td><strong>PUMA-PEFR prediction model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P_{COPD} ≥ 0.38</td>
<td>79</td>
<td>87</td>
<td>75</td>
<td>90</td>
<td>0.67</td>
<td>2</td>
</tr>
<tr>
<td>P_{COPD} ≥ 0.39a</td>
<td><strong>79</strong></td>
<td><strong>88</strong></td>
<td><strong>77</strong></td>
<td><strong>90</strong></td>
<td><strong>0.68</strong></td>
<td><strong>2</strong></td>
</tr>
<tr>
<td>P_{COPD} ≥ 0.40</td>
<td>77</td>
<td>88</td>
<td>76</td>
<td>89</td>
<td>0.65</td>
<td>2</td>
</tr>
</tbody>
</table>

C-PUMA: Chinese PUMA questionnaire; NNS: numbers needed to screen; NPV: negative predictive value; P_{COPD}: probability of COPD; PEFR: peak expiratory flow rate; PPV: positive predictive value.

a Indicates the best cutoff value determined by the Youden index.

b The cutoff value of %PEFR was based on our recent study, which had an identical study design and similar participants.15
<table>
<thead>
<tr>
<th>Independent variables</th>
<th>C-PUMA</th>
<th>% PEFR</th>
<th>( P_{\text{COPD}} )^a</th>
<th>COPD diagnosis</th>
<th>Post-bronchodilation value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FEV₁/FVC</td>
</tr>
<tr>
<td>Median value of different group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-COPD</td>
<td>5</td>
<td>98</td>
<td>0.102</td>
<td>No</td>
<td>0.80</td>
</tr>
<tr>
<td>COPD</td>
<td>6</td>
<td>70</td>
<td>0.703</td>
<td>Yes</td>
<td>0.61</td>
</tr>
<tr>
<td>Demonstration in selected cases from the validation cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 1</td>
<td>3</td>
<td>87</td>
<td>0.103</td>
<td>No</td>
<td>0.72</td>
</tr>
<tr>
<td>Case 2</td>
<td>3</td>
<td>49</td>
<td>0.785</td>
<td>Yes</td>
<td>0.64</td>
</tr>
<tr>
<td>Case 3</td>
<td>5</td>
<td>84</td>
<td>0.288</td>
<td>No</td>
<td>0.83</td>
</tr>
<tr>
<td>Case 4</td>
<td>5</td>
<td>79</td>
<td>0.694</td>
<td>Yes</td>
<td>0.61</td>
</tr>
<tr>
<td>Case 5</td>
<td>8</td>
<td>102</td>
<td>0.257</td>
<td>No</td>
<td>0.79</td>
</tr>
<tr>
<td>Case 6</td>
<td>8</td>
<td>66</td>
<td>0.901</td>
<td>Yes</td>
<td>0.69</td>
</tr>
</tbody>
</table>

% PEFR: percent predicted peak expiratory flow rate; \( P_{\text{COPD}} \): probability of COPD.

^a Entering the values of PUMA and %PEFR into a preset computer program immediately calculates the probability of COPD.
Appendix A, B, C. Supplementary data