EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

Derivation and Validation of the UCAP-Q Case-finding Questionnaire to Detect Undiagnosed Asthma and COPD

Chau Huynh, G. A. Whitmore, Katherine L. Vandemheen, J. Mark FitzGerald, Celine Bergeron, Louis-Philippe Boulet, Andreanne Cote, Stephen K. Field, Erika Penz, R. Andrew McIvor, Catherine Lemière, Samir Gupta, Irvin Mayers, Mohit Bhutani, Paul Hernandez, M. Diane Lougheed, Christopher J. Licskai, Tanweer Azher, Martha Ainslie, Ian Fraser, Masoud Mahdavian, Gonzalo G. Alvarez, Tetyana Kendzerska, Shawn D. Aaron

Please cite this article as: Huynh C, Whitmore GA, Vandemheen KL, *et al*. Derivation and Validation of the UCAP-Q Case-finding Questionnaire to Detect Undiagnosed Asthma and COPD. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.03243-2021).

Derivation and Validation of the UCAP-Q Case-finding Questionnaire to Detect Undiagnosed Asthma and COPD

Chau Huynh, MSc
G. A. Whitmore, PhD
Katherine L. Vandemheen, MScN
J. Mark FitzGerald, MD
Celine Bergeron, MD
Louis-Philippe Boulet, MD
Andreanne Cote, MD
Stephen K. Field, MD
Erika Penz, MD
R. Andrew McIvor, MD
Catherine Lemière, MD
Samir Gupta, MD
Irvin Mayers, MD
Mohit Bhutani, MD
Paul Hernandez, MD
M. Diane Lougheed, MD
Christopher J. Licskai, MD
Tanweer Azher, MD
Martha Ainslie, MD
Ian Fraser, MD
Masoud Mahdavian, MD
Gonzalo G. Alvarez, MD
Tetyana Kendzerska, MD
Shawn D. Aaron, MD

The Ottawa Hospital Research Institute, University of Ottawa, (CH, KLV, GA, TK, SDA); Desautels Faculty of Management, McGill University, (GAW); Department of Medicine, The University of British Columbia (JMF, CB); Department of Medicine, Université de Montreal (CL); Centre de recherche, Hôpital Laval, Université Laval (LPB, AC); Department of Medicine, University of Saskatchewan (EP); Cumming School of Medicine, University of Calgary (SKF); Department of Medicine and the Li Ka Shing Knowledge Institute of St. Michael's Hospital University of Toronto (SG); Firestone Institute for Respiratory Health, McMaster University (RAM); Department of Medicine, University of Alberta (IM, MB); Department of Medicine, Dalhousie University (PH); Department of Medicine, Queen's University (MDL); Department of Medicine, University of Manitoba (MA); Department of Medicine, University of Western Ontario (CJL); Department of Medicine, Memorial University (TA), Michael Garron Hospital, Department of Medicine, University of Toronto (IF); Royal Victoria Regional Health Centre (MM); all in Canada.
This study is dedicated to the memory of our beloved friend and colleague Dr. Mark FitzGerald

Correspondence and request for reprints:
Dr. Shawn Aaron
The Ottawa Hospital, General Campus
501 Smyth Road
Ottawa, Ontario, Canada
K1H 8L6
Telephone: (613) 739-6636
Fax: (613) 739-6807
Email: **saaron@ohri.ca**

Conception and design: SDA, GAW.
Analysis and interpretation: CH, GAW, SDA.
Drafting the manuscript for important intellectual content: CH, GAW, SDA, KLV, JMF, CB, CL, LPB, AC, EP, SKF, SG, RAM, IM, MB, PH, MDL, MA, CJL, TA, IF, MM, GGA, TK.
Word count: 3420

Funding: Canadian Institutes of Health Research- FDN grant #154322

Conflict of Interest Statement for Each Author:

Ms. Huynh reports no conflicts of interest, Dr. Whitmore reports no conflicts of interest, Ms. Vandemheen reports no conflicts of interest, Dr. FitzGerald reports no conflicts of interest, Dr. Bergeron reports no conflicts of interest, Dr. Boulet reports receiving grants, consulting fees or honoraria from Amgen, AstraZeneca, GlaxoSmithKline, Merck, Novartis, and Sanofi-Regeneron, Dr. Cote reports receiving honoraria from GlaxoSmithKline, Astra-Zeneca, Sanofi and Covis, Dr. Field reports receiving consulting fees or honoraria from Merck, GlaxoSmithKline, Novartis and Boeringher Ingelheim, Dr. Penz reports receiving honoraria from Astra-Zeneca, GlaxoSmithKline, Novartis, ICEBM, and Boeringher Ingelheim, Dr. McIvor reports no conflicts of interest, Dr. Lemiere reports receiving consulting fees or honoraria from Sanofi, TEVA, GlaxoSmithKline, and Astra-Zeneca, Dr. Gupta reports no conflicts of interest, Dr. Mayers reports no conflicts of interest, Dr. Bhutani reports payments or honoraria from AZ, GSK, Novartis, Grifols, Sanofi, Covis, and Valeo, Dr. Hernandez reports payments or honoraria from AZ, GSK, Novartis, and Valeo, Dr. Lougheed reports receiving grants from AZ and GSK, Dr. Licskai reports receiving honoraria from Astra-Zeneca, GlaxoSmithKline, Novartis, and Boeringher Ingelheim, Dr. Azher reports no conflicts of interest, Dr. Ainslee reports receiving honoraria from Boeringher Ingelheim, Dr. Fraser reports no conflicts of interest, Dr. Mahdavian reports no conflicts of interest, Dr. Alvarez reports no conflicts of interest, Dr. Kendzerska reports speaker honoraria from Astra-Zeneca,

Dr. Aaron reports sitting on advisory boards and honoraria payments from GlaxoSmithKline, Sanofi, and Astra-Zeneca.

Abstract:

Background: Many people with asthma and COPD remain undiagnosed. We developed and validated a new case-finding questionnaire to identify symptomatic adults with undiagnosed obstructive lung disease.

Methods: Adults in the community with no prior history of physician-diagnosed lung disease who self-reported respiratory symptoms were contacted via random-digit dialing. Pre- and post-bronchodilator spirometry was used to confirm asthma or COPD. Predictive questions were selected using multinomial logistic regression with backward elimination. Questionnaire performance was assessed using sensitivity, predictive values, and area under the receiver operating curve (AUC). The questionnaire was assessed for test-retest reliability, acceptability, and readability. External validation was prospectively conducted in an independent sample and predictive performance re-evaluated.

Results: A 13-item UCAP-Q case-finding questionnaire to predict undiagnosed asthma or COPD was developed. The most appropriate risk cut-off was determined to be 6% for either disease. Applied to the derivation sample (N=1615), the questionnaire yielded a sensitivity of 92% for asthma and 97% for COPD, specificity of 17%, with an AUC of 0.69 (95% CI: 0.64-0.74) for asthma and 0.82 (95% CI: 0.78-0.86) for COPD. Prospective validation using an independent sample (n=471) showed sensitivities of 93% and 92% for asthma and COPD, respectively, specificity of 19%, with AUC's of 0.70 (95% CI: 0.62-0.79) for asthma and 0.81 (95% CI: 0.74-0.87) for COPD. AUC's for UCAP-Q were higher compared to AUC's for currently recommended case-finding questionnaires for asthma or COPD.

Conclusions: The UCAP-Q demonstrated high sensitivities and AUC's for identifying undiagnosed asthma or COPD. A web-based calculator allows for easy calculation of risk probabilities for each disease.

Key Words: obstructive lung diseases; asthma, COPD, screening; multinomial logistic regression, risk calculator

Introduction

Approximately 545 million people worldwide suffer from chronic respiratory diseases, with asthma and chronic obstructive pulmonary disease (COPD) accounting for the

majority of the global respiratory disease burden.[1] However, the true prevalence of asthma and COPD in the community is likely under-estimated.[2–5] A summary review of previous studies suggest up to 70% of obstructive lung disease (OLD) remains undiagnosed in the population.[4,6–9]

While spirometry testing is viewed as the gold standard for diagnosing OLD, its underuse and inaccessibility undermines efforts to appropriately identify and diagnose symptomatic persons.[10] Other factors, including underreporting of patients' symptoms, also contribute to missed opportunities for diagnosis.[11–13] The 2016 United States Preventive Services Task Force (USPSTF) advised against screening spirometry for asymptomatic persons, but encouraged active case-finding to identify those with symptoms who might suffer from undiagnosed OLD.[14] These above recommendations require consideration of the difference between screening, which involves testing large numbers of apparently healthy people to detect unrecognized disease; and case-finding, whereby people with respiratory symptoms who are at increased risk of having a disease are tested to make a diagnosis earlier than would occur by waiting for them to present with symptoms or signs. It is hoped that earlier detection of symptomatic OLD by case-finding with subsequent disease management may improve patients' symptoms and their health outcomes.[12,15]

Previous studies have attempted to develop case-finding tools for either asthma or COPD separately. A number of instruments have been developed for COPD; however, these were limited by small derivation sample size,[16–18] lack of external validation,[19–21] inclusion of self-reported cases,[22] or use of only pre-bronchodilator spirometry, resulting in misdiagnosed cases.[16,19] The CAPTURE questionnaire was tested in 186 patients with COPD and 160 control subjects and exhibited a sensitivity of 96% for differentiating COPD cases from controls[17]. The COPD Diagnostic Questionnaire (COPD-DQ)[23] is the most extensively studied questionnaire and has moderate overall

performance for COPD detection, with an area under the receiver operating characteristic curve (AUC) of 0.65 to 0.72.[24] Case-finding instruments developed specifically for adult asthma are sparse.[25,26] The Asthma Screening Questionnaire (ASQ) is one of the few diagnostic instruments developed for persons over 18 years of age. However, with a derivation sample size of only 50 participants, and no internal or external validation, the generalizability of the ASQ remains uncertain.[25]

Combining asthma and COPD detection into a single instrument is worthwhile, because despite their distinct pathophysiological differences, they present with the same symptoms (*i.e*., dyspnea, wheeze, cough, and mucus production) and health care providers often have trouble distinguishing one condition from the other upon initial presentation.[27] The first diagnostic test for both conditions is pre and post bronchodilator spirometry. Accordingly, a single questionnaire could identify patients at risk of either disease who require objective testing. To that end, this study describes the development and validation of a case-finding questionnaire to identify symptomatic adults with undiagnosed asthma or COPD in the community. This marks the first study to incorporate both diseases into a single case-finding instrument.

Methods

Study Population and Recruitment

The subject cohort was drawn from the Undiagnosed COPD and Asthma Population (UCAP) Study. The questionnaire developed here, which we refer to as the UCAP Questionnaire (UCAP-Q), was derived and then independently validated in two separate phases (the derivation phase and validation phase, respectively). Participants from the derivation phase were recruited via random-digit dialing of landlines and cellphones from sixteen sites across Canada between June 2017 and March 2020. [28] Participants were enrolled if they were 18 years of age or older, experienced one or more respiratory symptoms (*i.e*., shortness of breath, wheezing, increased mucus or sputum, or prolonged cough) in the past six months, and had no prior physician diagnosis of asthma, COPD, or

any other lung diseases. All potential participants completed the ASQ regardless of age. Participants ≥60 years, and participants <60 years with a score of <6 on the ASQ, also completed the COPD-DQ. Participants scoring ≥20 points on the COPD-DQ or ≥6 on the ASQ were invited to the local study site for pre and post bronchodilator spirometry to confirm, or rule out, OLD. Participants who did not report sufficient respiratory symptoms to pass the eligibility thresholds were not invited for spirometry. Subjects whose forced expiratory volume in one second (FEV1) improved by ⩾12% and ⩾200 mL following bronchodilator administration with 400 µg of salbutamol, were labelled as having "spirometry consistent with asthma". Subjects whose post-bronchodilator forced expiratory volume in one second/forced vital capacity (FEV1/FVC) ratio was below the lower 95% confidence limit for a healthy individual, adjusted for sex, age and height, were labelled as having "spirometry consistent with COPD". Subjects who met spirometry criteria for both conditions were considered to have COPD with partial bronchodilator reversibility and were classified as "spirometry consistent with COPD" for the purposes of the present analysis. Subjects meeting neither criterion were classified as "no obstructive lung disease." All subjects whose spirometry was consistent with COPD or asthma were re-tested with spirometry at a later date to ensure consistency of results over time.

After the UCAP-Q was developed and tested for reliability and acceptability it was prospectively validated using an independent sample of participants. Participants from the validation phase were recruited from October 2020 to September 2021. The process of recruitment was identical to the derivation phase with one exception, for the validation phase the UCAP-Q was also administered in addition to the ASQ and COPD-DQ questionnaires. Participants scoring ≥20 points on the COPD-DQ, or ≥6 on the ASQ, or whose UCAP-Q responses yielded an expected probability of asthma or COPD of 6% or higher, were invited to the local study site for pre and post bronchodilator spirometry to confirm, or rule out, OLD.

Ethics approval was obtained from each local study site and all subjects signed informed, written consent. This study was conducted and analyzed in accordance with TRIPOD guidelines.

Derivation of the Questionnaire- Candidate Pool of Predictors

The pool of potential questions used to develop the UCAP Questionnaire was selected from the six questionnaires listed below. Study participants completed these questionnaires during their spirometry visit:

1. COPD Assessment Test (CAT)[29]; 2. Short Form-36 Quality of Life Questionnaire (SF-36 QoL)[30]; 3. Work, Productivity and Impairment: General Health (WPAIGH) Questionnaire[31]; 4. The St. George's Respiratory Questionnaire (SGRQ)[32]; 5. ASQ[25]; 6. A data collection form containing demographic and clinical variables.

Only predictors demonstrating univariate significance of $p<0.25$[33] and considered clinically relevant were subsequently chosen for the candidate pool. Table E1 in the online supplement lists the candidate pool of predictors and associated univariate results.

Statistical Analyses

The UCAP-Q questionnaire was derived from a multinomial logistic regression analysis using backward elimination to select predictors from the candidate pool of predictors in Table E1. Predictors were removed sequentially until $p<0.05$. The outcome variables were asthma, COPD, and no obstructive lung disease (no OLD). An alternative potential model was developed using Classification and Regression Trees (CART), however the resultant questionnaire developed using the CART methodology lacked face validity and discriminative ability and was therefore abandoned in favor of the questionnaire generated via logistic regression.

The item scores for variables with ordinal rating scales in the regression model were obtained by fitting a logistic regression function to each disease outcome. Based on the relationships observed between item responses and outcomes of each disease, some questions with ordinal rating scales demonstrated non-monotonic associations between the outcome risk and ordinal item responses.

For example, some ordinal item responses followed a U-shaped risk curve for the disease outcome. To illustrate, subjects found to have undiagnosed asthma, when answering how many attacks of shortness of breath or wheezing they had experienced in the past 3 months, gave the extreme responses of either 'zero attacks' or 'more than three attacks' as their most common answers. Fitted logistic regression scoring of the item responses addresses this irregular characteristic. [28]

Performance of the instrument was assessed using the area under the receiver operating characteristic curve (AUC) for discrimination. Calibration was assessed using a Hosmer-Lemeshow goodness-of-fit test for multinomial logistic regression. Internal validation using 10-fold cross-validation was used to assess overfitting of the regression model. The specificity, sensitivity, positive predictive value, and negative predictive value of the questionnaire were assessed for the derivation and validation phases.

To determine the optimal risk cut-off for each disease, we applied specific opportunity costs to prediction errors, and chose risk thresholds that minimized the total incurred cost of prediction errors. We reasoned that the cost of a false negative for asthma (ie. if asthma was predicted to be no OLD by the UCAP-Q) would be relatively high, since undiagnosed and hence untreated asthma can be potentially associated with lifelong morbidity, while diagnosed asthma is usually controllable with treatment, and therefore we assigned this a cost of ten units. We assigned a slightly lower cost of eight units for a false negative for COPD, on the assumption that diagnosis and treatment of COPD while very important, usually does not completely control symptoms and correct activity limitation. We assigned a cost of one unit for a false positive for asthma or COPD (ie. if no OLD was predicted to be asthma or COPD), on the assumption that spirometry testing of someone who does not have obstructive lung disease is safe and relatively inexpensive.

Acceptability and Reliability of the Questionnaire

A sample of respiratory clinicians and a separate sample of UCAP study participants were recruited to obtain their feedback on the content and wording of UCAP-Q. Test-retest reliability was obtained

by having selected participants complete the UCAP-Q on two separate occasions, one week apart. Intraclass correlation coefficient (ICC) was used for continuous variables and Kappa was used to measure test-retest reliability for categorical variables (unweighted Kappa for ordinal variables).

Sample Size Estimate for the Validation Phase

Sample size for the validation phase was estimated based on the AUC as the performance measure of interest. For COPD, a total sample of 452 participants provided a two-sided 95% confidence interval with a 0.2 precision for a target AUC of 0.85. For asthma, a sample of 313 participants achieved the same precision for a target AUC of 0.75. Statistical analyses were performed using STATA (StataCorp, College Station, TX, USA) Version 16.1.

Results

Study Population

A total of 1,652 participants were enrolled in the derivation study (**Figure 1a**). Thirty-seven participants were unable to complete adequate spirometry and were excluded. The remaining 1,615 participants were included. One hundred and ninety-five participants (12.0%) were found to have newly diagnosed COPD and 136 (8.4%) were found to have newly diagnosed asthma.

A total of 471 participants were enrolled in the validation study (Figure 1b). Forty-nine participants (10.4%) were found to have newly diagnosed COPD and 42 were found to have newly diagnosed asthma (8.9%). The clinical and demographic characteristics of the derivation and validation cohorts are shown in **Table 1**.

Results- Derivation Phase

The results of the multinomial logistic regression are shown in **Tables E2** and **E3**. Thirteen questions were found to be predictive of undiagnosed asthma or COPD. A web-calculator depicts the 13-item questionnaire and automatically calculates predicted probabilities of asthma and COPD based on the participant's responses. **Figures 2a and 2b** display the calculator, which is accessible for public use at

https://omc.ohri.ca/UCAPquestionnaire/. The risk score for each disease is the predicted probability calculated from each of the multinomial logistic regression equations. **Table 2** depicts sensitivity, specificity, and predictive values at varying risk cut-offs for each disease using data from the derivation cohort. Using the assigned unit costs to minimize the cost of prediction errors, the most appropriate risk cut-off was determined to be 6% for each disease.

Table 3 depicts the performance of the UCAP-Q when applied to the derivation sample to predict asthma or COPD at a 6% risk cut-off for both diseases. This cut-off in the derivation sample yielded a sensitivity of 92% for asthma, a sensitivity of 97% for COPD, with an overall specificity of 17%.

The AUC of the UCAP-Q for the derivation exercise was 0.69 (95% CI: 0.64-0.74) for asthma and 0.82 (95% CI: 0.78-0.86) for COPD. Internal 10-fold cross-validation produced average AUC values of 0.64 (95% CI: 0.45-0.80) for asthma and 0.79 (95% CI: 0.70-0.90) for COPD. The *p*-value of the Hosmer-Lemeshow goodness of fit test was 0.258, indicating a well-fitted model. Calibration/discrimination plots of cumulative cases and cumulative probability estimates among the 1580 subjects in the derivation sample are depicted in **Figure 3.**

Acceptability, Reliability and Readability of the UCAP Questionnaire:

Prior to external validation of the questionnaire, the 13-item tool was sent to twelve pulmonary specialists and general internists, and a sample of 27 study participants for their feedback related to the content and clarity of the questions. Overall, only minor modifications to the questions were made, clarifying that sleep disruption was related to respiratory symptoms and describing examples of attacks of chest problems as attacks of shortness of breath or wheezing. The UCAP Questionnaire scored 7 on the Flesch-Kincaid Grade Level scale and 70.9 on the Flesch-Kincaid Reading Ease scale, indicating 'fairly easy' readability.[34]

Test-retest reliability was assessed in 27 UCAP participants who completed the UCAP Questionnaire at two time points, one-week apart. ICC and Kappa values for the 13-items were between 0.739 and

1.00, with an outlying ICC of 0.592 for one question. The test-retest reliability results can be found in Table E4 of the online supplement.

Results- Validation Phase

Table 4 depicts the performance of the UCAP-Q when applied prospectively to the independent validation sample to predict asthma or COPD. Applying UCAP risk prediction thresholds of ≥6% for each disease to the validation sample yielded a sensitivity of 93% for asthma, sensitivity of 92% for COPD, with an overall specificity of 19%. The AUC for the validation phase was 0.70 (95% CI: 0.62-0.79) for asthma and 0.81 (95% CI: 0.74-0.87) for COPD.

Table 5 shows a comparison of the performance of the UCAP-Q against the predictive performance of the combined ASQ and COPD-DQ questionnaires for the independent validation sample**.** The UCAP Questionnaire was able to predict a diagnosis of asthma with higher sensitivity compared to the ASQ/COPD-DQ (93% vs 76% sensitivity respectively). Sensitivity for prediction of undiagnosed COPD (92%) was identical between the questionnaires. The UCAP-Q exhibited a higher AUC for prediction of asthma (Asthma AUC for UCAP-Q = 0.70 vs Asthma AUC for ASQ/COPD-DQ = 0.65), and a higher AUC for prediction of COPD (COPD AUC for UCAP = 0.81 vs COPD AUC for ASQ/COPD-DQ = 0.77).

Discussion

We have developed and validated a case-finding tool to identify symptomatic community-dwelling adults suspected of undiagnosed asthma or COPD. We designed the UCAP-Q questionnaire to be easy to complete either by telephone or on-line. Computer scoring of the questionnaire is automatically done by an on-line program. The program provides expected probabilities for asthma and COPD separately, based on the participant's responses.

We designed the questionnaire to have relatively high sensitivities and high negative predictive values, at the expense of lower specificity, and we selected all subjects with an expected probability of asthma ≥ 6%, or of COPD ≥6%, for testing with spirometry. This decision was made because we did not want to miss diagnoses of asthma or COPD in subjects with significant respiratory symptoms. The opportunity costs of missing a diagnosis of asthma or COPD in patients suffering from respiratory symptoms are relatively large relative to the cost and risks of spirometry testing, which are relatively small. However, we acknowledge that individual clinicians might employ risk thresholds other than 6%, for deciding which patients should proceed to spirometry. The UCAP-Q generates an easy-to-understand probability for asthma and a similar probability for COPD, and this information can then be used by the clinician to decide independently on whether spirometry testing is warranted.

Findings from the external validation suggest the UCAP Questionnaire demonstrates better sensitivity and a better AUC value compared to existing ASQ and COPD-DQ questionnaires to detect cases of undiagnosed asthma. The AUC for detecting undiagnosed COPD was also better for the UCAP Questionnaire compared to the ASQ and COPD-DQ questionnaires.

While no difference in sensitivity was found between the ASQ/COPD-DQ and UCAP-Q for COPD, the UCAP-Q is a better case-finding tool than the ASQ for adult asthma. The UCAP-Q had 93% sensitivity for identifying asthma compared to 76% sensitivity for ASQ. Thus, our study serves the objective of case finding by identifying a larger proportion of true positives. The UCAP-Q tool also incorporates robust risk factors, such as occupational exposures, and is not limited to asking only questions about symptoms as seen in the ASQ. Given that barriers to diagnosis of lung disease are multifactorial, risk questions related to environmental influences and social factors allows for a better identification of undiagnosed cases of either disease.

Some strengths of our study include the large samples for model derivation and prospective external validation, the use of multiple sites, and representative population-based recruiting by random-digit dialing. We performed a separate validation of the model using an external independent prospective sample, and assessed the questionnaire for acceptability, readability and reliability. Previous case-finding tools were often developed by using data from pulmonary or primary care clinics, and most studies evaluating case-finding tools included participants who were previously diagnosed with asthma or COPD. Hence predictive performance measures (e.g., positive predictive value, negative predictive value) of previous studies were likely a reflection of the increased disease prevalence in the selected sample.[20–22] In contrast, data obtained for this study were gathered by random sampling of previously undiagnosed subjects from the general adult population and the prevalence rates obtained in this study for each disease are more likely to reflect the current prevalence of undiagnosed asthma and COPD in symptomatic individuals within the community.

UCAP sampling purposely avoided GP practices and clinics where patients with breathing problems may have been clinically assessed and perhaps already diagnosed with disease. Instead, UCAP used a case-finding strategy to identify at-risk individuals with undiagnosed respiratory

symptoms in the community. Thus, the derivation and validation samples were drawn from a random, representative population-based sample rather than an in-clinic sample. In-clinic samples would tend to have a higher prevalence of patients who are actively followed for respiratory problems or have diagnosed disease. While the UCAP-Q was externally validated using a random, population-based independent sample, future external validation of the UCAP-Q in other settings, such as primary care clinics and practices, can further evaluate its performance.

This study has several limitations. Because the UCAP-Q was designed as a case-finding tool we tested the questionnaire only in subjects who reported a recent history of respiratory symptoms. The UCAP-Q has not been validated in subjects who may have obstructive lung disease that is mild enough to be asymptomatic. We did not perform bronchial challenge tests in subjects with normal spirometry, therefore, some symptomatic individuals having asthma with airway hyper-responsiveness, but without airflow obstruction or responsiveness to bronchodilator, may have been missed. Performing bronchial challenge testing on every subject with normal spirometry who entered into this study would have been prohibitively expensive and was impossible to do during the pandemic because of infection control concerns related to aerosolization of methacholine. In addition, bronchial challenge testing is difficult to access in a real-life community healthcare setting, and we wanted to ensure that results from our study could be translated to practice within the community. Given that data gathered for this study were based on random digit dialing, only persons with access to a cellphone or a landline who lived within 90 minutes of each study site were included. Most subjects in the derivation and validation cohorts were of one ethnic descent, and more than 50% had a college or university education, suggesting that some subgroups within the population were not well represented.[35,36] These findings highlight the need for further validation to assess the usefulness of the UCAP case-finding tool in different sub-populations.

In conclusion, we have developed and externally validated a 13-item case-finding questionnaire to assess undiagnosed asthma or COPD in symptomatic, community-dwelling adults. Our findings suggest that the case-finding instrument is reliable and sensitive in detecting undiagnosed cases of asthma or COPD, which should then be confirmed with spirometry. Future case-finding approaches could implement use of the UCAP Questionnaire in public spaces or embedded within on-line social media platforms to identify persons within these settings with undiagnosed OLD who should be targeted with spirometry. The UCAP-Q is available online and the web-based calculator can be used by people in the community who are suffering from respiratory symptoms to assess their probability of undiagnosed asthma or COPD. We hope that this easy access will prompt symptomatic people to seek physician evaluation and request spirometry testing.

Figure Legends:

Figure 1. Study Flow diagrams: **Figure 1A)** derivation cohort, and **Figure 1B)** validation cohort

Figure 2A. The Online UCAP-Q questionnaire

Figure 2B: Example of calculated risk scores for asthma and COPD for a person using the Online UCAP Questionnaire

Figure 3: Calibration/discrimination plots of cumulative cases and cumulative probability estimates among the 1580 subjects in the derivation sample.

The two curves in each plot are cumulative cases and cumulative probability estimates among the 1580 subjects in the derivation sample, ordered from largest to smallest probability estimate. Close tracking of the two curves indicates good calibration for the multinomial logistic regression model. Strong bending of the curves toward the upper-left corner of the graph shows good discrimination between disease cases and subjects with no disease.

Table 1. Demographic and clinical characteristics of the derivation and validation cohorts

Characteristics	Derivation Cohort (N=1615)			Validation Cohort (N=471)		
	No OLD N=1284	Asthma N=136	COPD N=195	No OLD N=380	Asthma N=42	COPD N=49
Age, year*	61 (48-70)	60 (48-71)	67 (59-74)	64 (54-72)	64 (59-70)	67 (61-73)
Male, (n, %)	630 (49)	78 (57)	125 (64)	185 (49)	19 (45)	29 (59)
Race/Ethnicity (n, %)						
Caucasian	1,178 (92)	127 (93)	188 (97)	362 (95)	42 (100)	48 (98)
Asian	56 (4)	5 (4)	1 (0.5)	9 (2)	0	1 (2)
American Indian, Alaska Native, Native Hawaiian, Hispanic or Latino, Mixed	24 (2)	2 (1.5)	4 (2)	5 (1)	0	0
Black or African American	26 (2)	2 (1.5)	1 (0.5)	4 (1)	0	0
Level of education (n, %)						
High school or less	359 (28)	42 (31)	75 (38)	88 (23)	12 (29)	10 (20)
Some college/university	174 (14)	21 (15)	24 (12)	67 (18)	7 (16)	12 (25)
College/university	712 (55)	71 (52)	87 (45)	225 (59)	23 (55)	27 (55)
Smoking history (n, %)						
Current	223 (17)	20 (15)	84 (43)	49 (13)	5 (12)	18 (37)
Former	508 (40)	61 (45)	86 (44)	162 (43)	19 (45)	22 (45)
Never	533 (43)	55 (40)	25 (13)	169 (44)	18 (44)	9 (18)
Comorbidities (n, %)						
GERD	473 (37)	48 (36)	64 (34)	121 (32)	15 (36)	12 (24)
Stroke	53 (4)	1 (1)	8 (4)	11 (3)	1 (2)	3 (6)
Coronary artery disease	108 (8)	10 (8)	39 (20)	45 (12)	6 (14)	10 (20)
Hypertension	447 (35)	42 (31)	70 (37)	144 (38)	20 (48)	14 (29)
Depression/Anxiety	492 (39)	52 (39)	64 (34)	160 (42)	17 (40)	11 (22)
Diabetes mellitus	176 (14)	16 (12)	28 (15)	64 (17)	7 (17)	4 (8)
Pre-bronchodilator spirometry*						
FEV_1	2.78 (2.21-3.36)	2.45 (1.90-3.06)	1.98 (1.44-2.42)	2.73 (2.18-3.29)	2.32 (1.85-2.75)	2.02 (1.45-2.50)
FEV_1 % predicted	96 (86-106)	82 (74-91)	71 (59-82)	98 (91-108)	84 (73-92)	71 (59-82)
FEV_1/FVC	77 (73-80)	69 (64-74)	60 (54-64)	77 (72-80)	69 (65-75)	59 (53-62)
Post-bronchodilator spirometry*						
FEV_1	2.86 (2.28-3.47)	2.80 (2.21-	2.09 (1.57-2.62)	2.82 (2.25-	2.69 (2.23-	2.09 (1.56-2.64)

		3.49)		3.40)	3.13)	
FEV_1 % predicted	99 (89-109)	95 (87-103)	76 (65-86)	102 (92-111)	94 (87-102)	77 (68-88)
FEV_1/FVC	79 (75-83)	75 (70-80)	62 (57-66)	79 (75-83)	74 (70-78)	61 (56-65)

*Data are presented as median (interquartile range)

OLD= obstructive lung disease

Table 2. Associated sensitivity, specificity, PPV, and NPV values (%) for UCAP-Q at varying risk cut-offs for each disease (derivation cohort)

	Asthma	COPD	No OLD		
Risk Cut-Off (%)	**Sensitivity**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
5	97	99	10	22	95
6	92	97	17	23	92
10	64	82	47	27	88
15	38	70	72	34	86
20	24	61	85	44	85

PPV= positive predictive value

NPV= negative predictive value

Table 3. Classification table of predicted and true disease at a risk cut-off of 6% for asthma and COPD in the derivation cohort

	True Disease[†]				
Disease Prediction*	**No OLD**	**Asthma**	**COPD**	**Total**	**PPV/NPV**
Yes	1047 (FP)	124 (TP1)	184 (TP2)	1355	PPV 23%
No	208 (TN)	11 (FN1)	6 (FN2)	225	NPV 92%
Total	1255	135	190	N=1580	--
Sensitivity	--	92%	97%	--	--
Specificity	17%	--	--	--	--

FP=false positive, FN= false negative, TP= true positive, TN=true negative. [†]Gold standard reference defined by spirometry Subject counts in Table 3 (total n=1580) are smaller than counts shown in Table 1 because of missing values in predictor variables of the UCAP-Q model in 35 subjects. Imputation of missing values was not attempted to avoid estimation bias.

Table 4. Classification table of predicted and true disease at a risk cut-off of 6% for asthma and COPD for the prospective validation cohort

	True Disease†				
Disease Prediction*	**No OLD**	**Asthma**	**COPD**	**Total**	**PPV/NPV**
Yes	309 (FP)	39 (TP1)	45 (TP2)	393	PPV 21%
No	71 (TN)	3 (FN1)	4 (FN2)	78	NPV 91%
Total	380	42	49	N=471	--
Sensitivity	--	93%	92%	--	--
Specificity	19%	--	--	--	--

FP=false positive, FN= false negative, TP= true positive, TN=true negative. †Gold standard reference defined by spirometry

Table 5. Comparison of predictive performance for the UCAP Questionnaire and the ASQ/COPD-DQ

		Asthma		COPD			
Questionnaire	**Specificity**	**Sensitivity**	**AUC**	**Sensitivity**	**AUC**	**PPV**	**NPV**
UCAP	19%	93%	0.70	92%	0.81	21%	91%
ASQ/COPD-DQ	18%	76%	0.65	92%	0.77	20%	83%

AUC= area under the receiver operating curve. PPV= positive predictive value, NPV= negative predictive value

Figure 1. Study Flow diagrams: Figure 1a) derivation cohort, and Figure 1b) validation cohort

Figure 1a
Derivation Study:

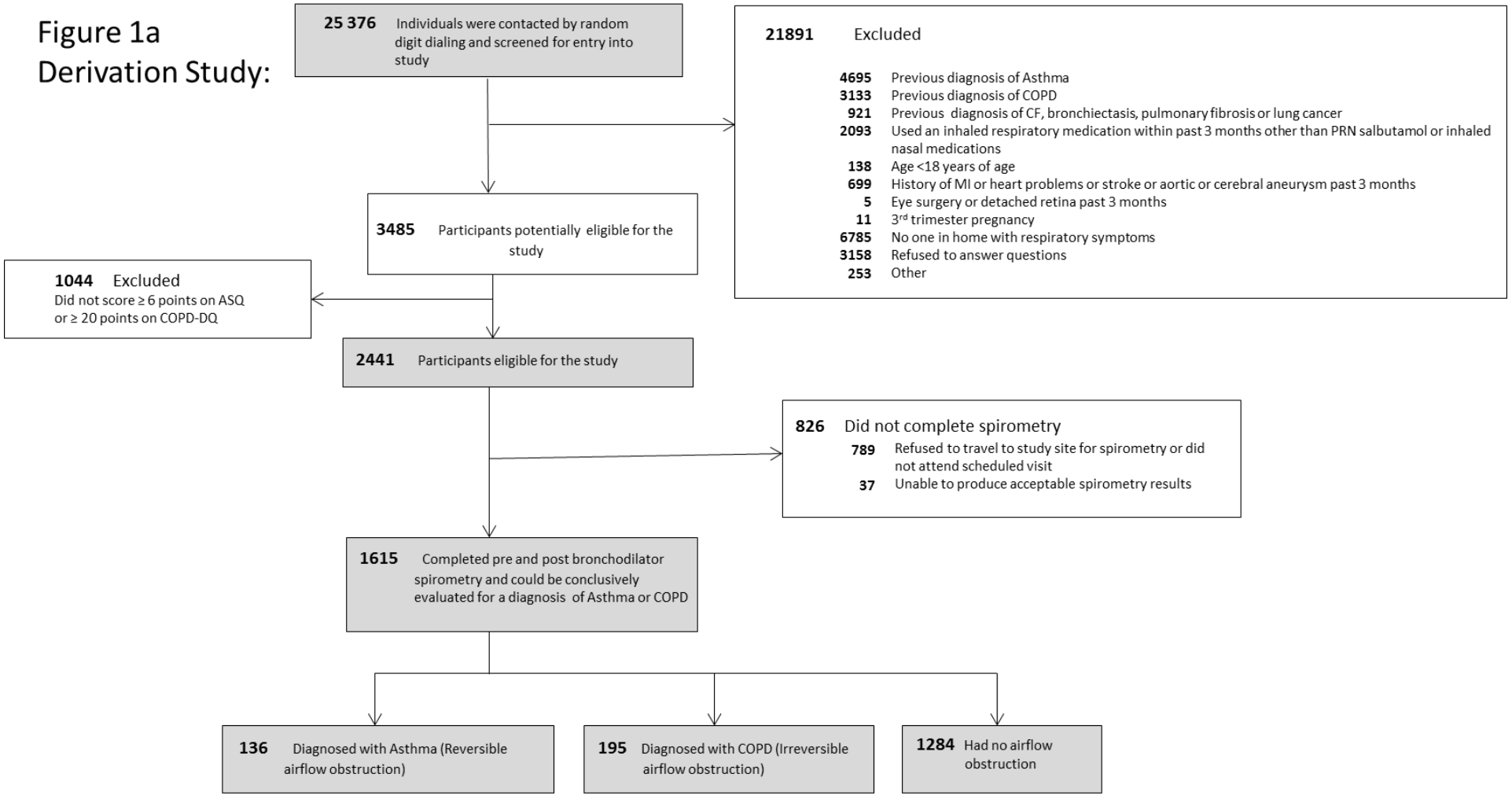


Figure 1b
Validation Study:

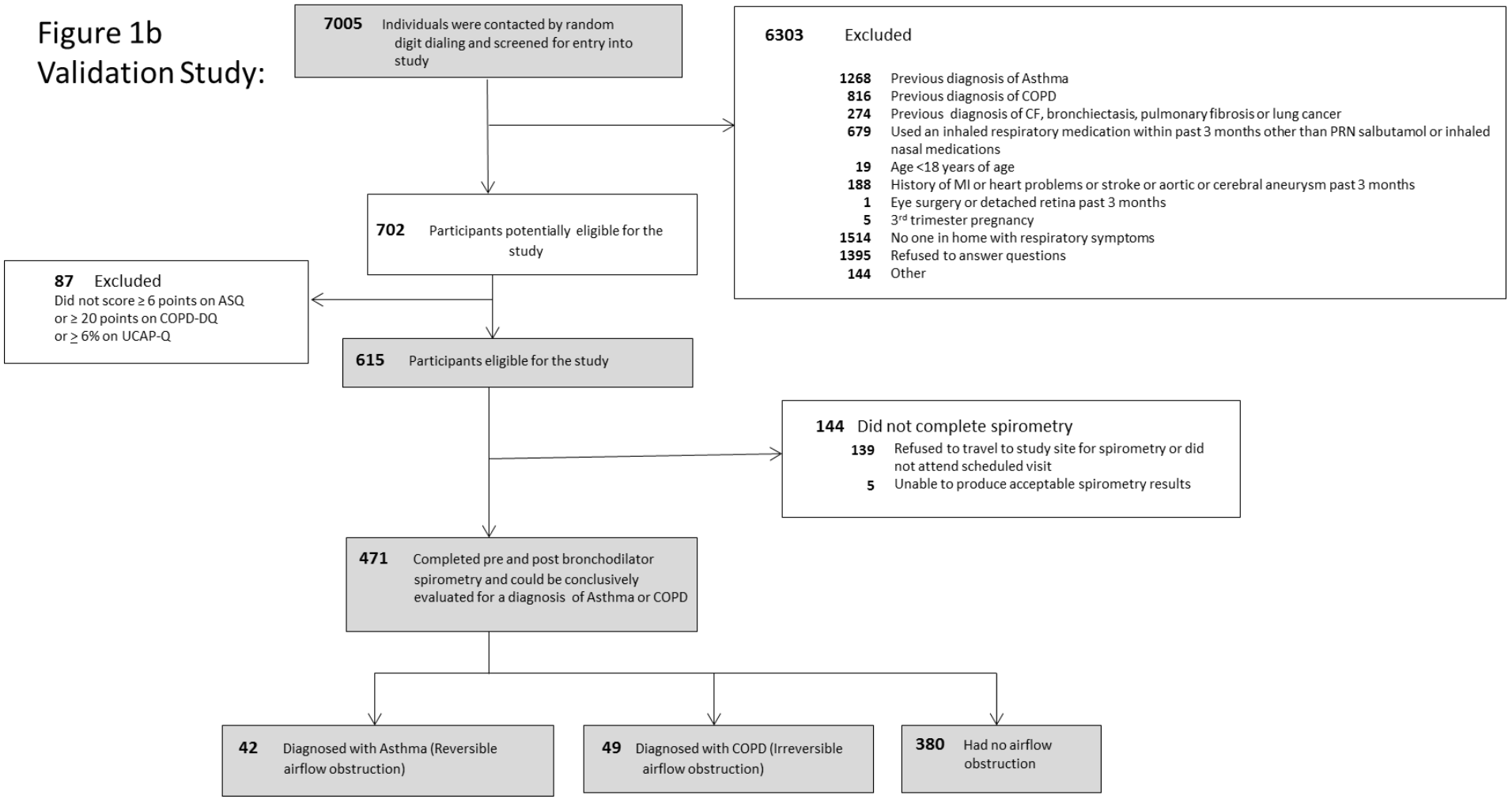
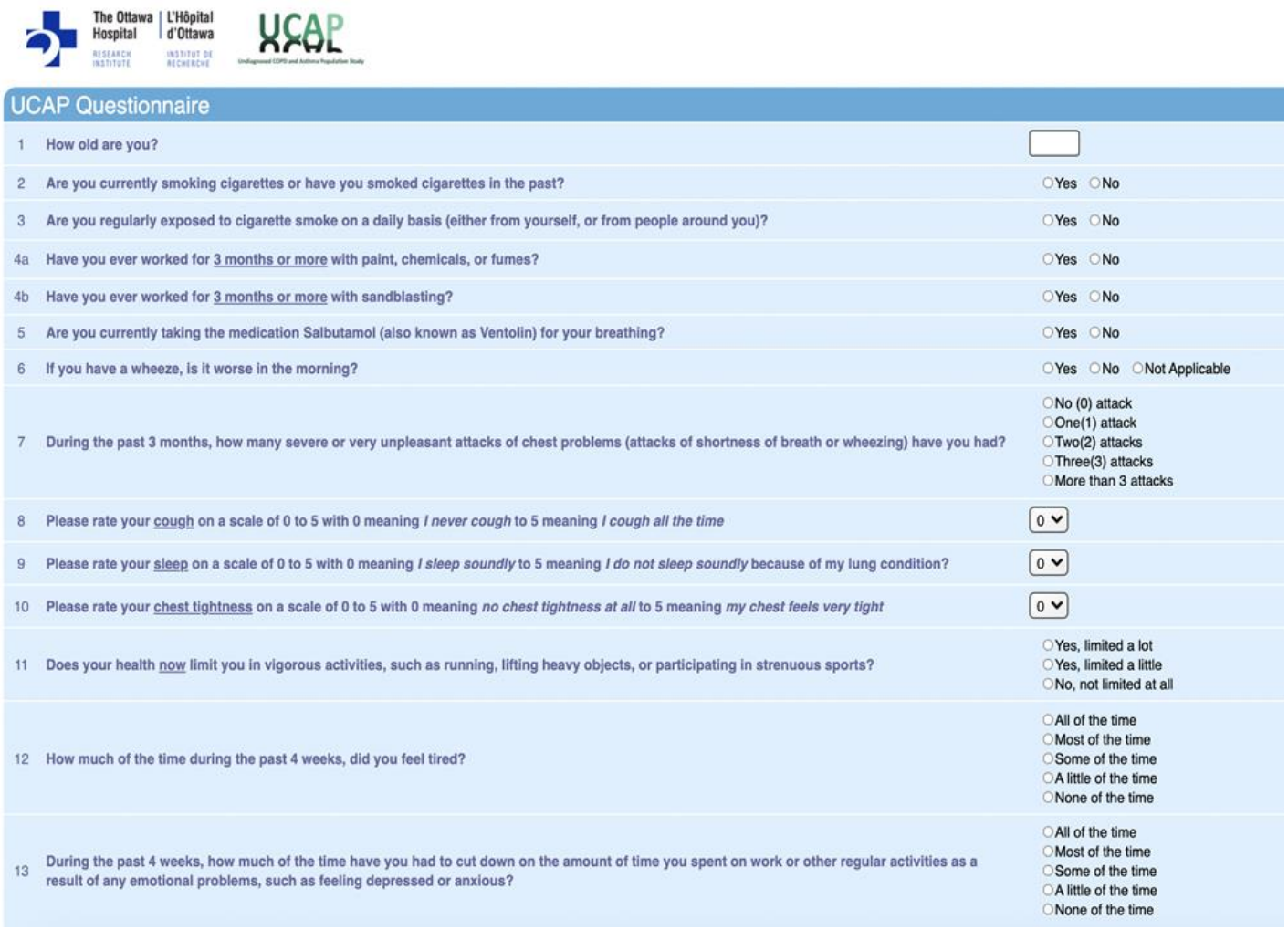


Figure 2A. The Online UCAP-Q questionnaire



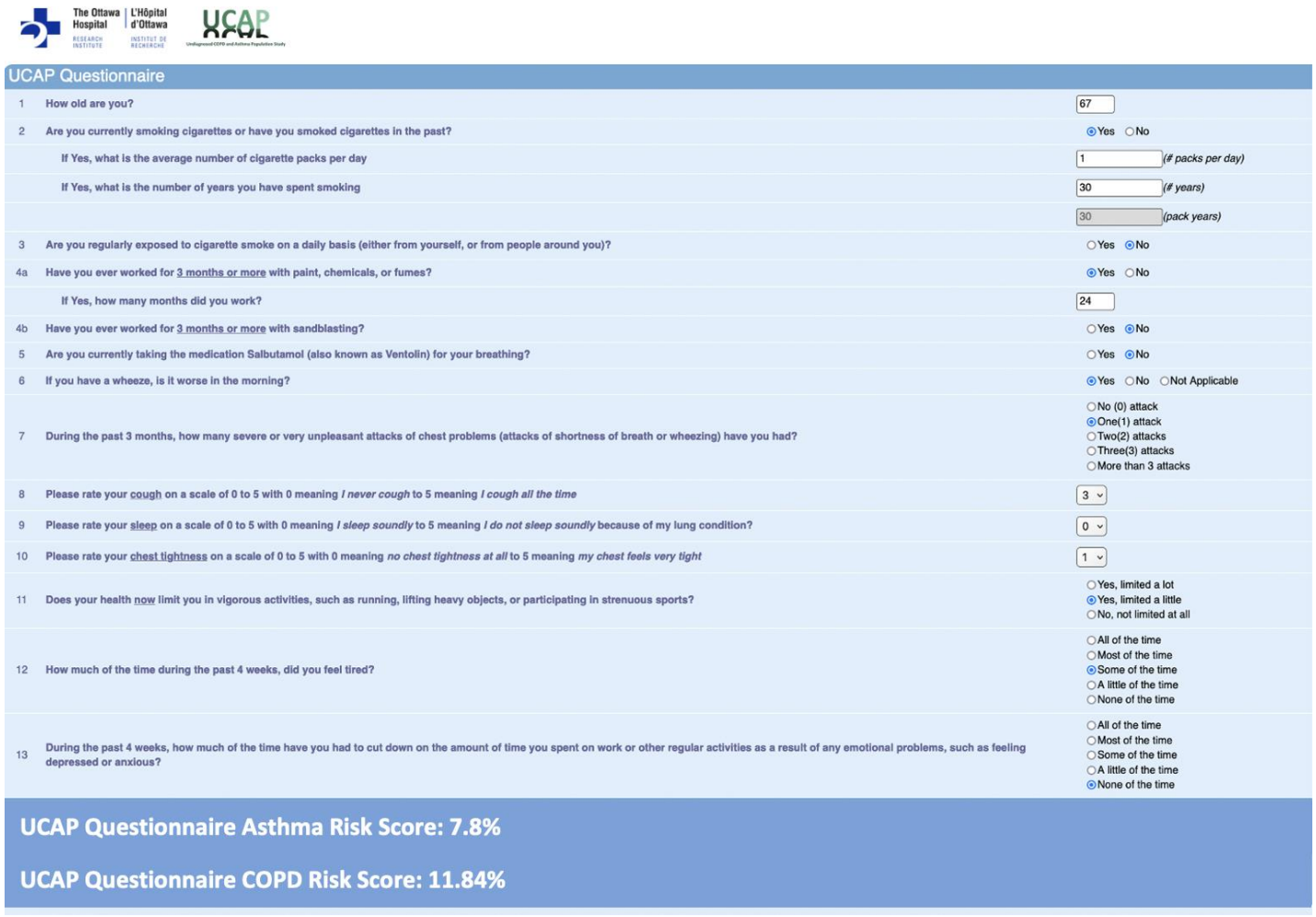
The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

UCAP

UCAP Questionnaire

1	How old are you?	☐
2	Are you currently smoking cigarettes or have you smoked cigarettes in the past?	○Yes ○No
3	Are you regularly exposed to cigarette smoke on a daily basis (either from yourself, or from people around you)?	○Yes ○No
4a	Have you ever worked for 3 months or more with paint, chemicals, or fumes?	○Yes ○No
4b	Have you ever worked for 3 months or more with sandblasting?	○Yes ○No
5	Are you currently taking the medication Salbutamol (also known as Ventolin) for your breathing?	○Yes ○No
6	If you have a wheeze, is it worse in the morning?	○Yes ○No ○Not Applicable
7	During the past 3 months, how many severe or very unpleasant attacks of chest problems (attacks of shortness of breath or wheezing) have you had?	○No (0) attack ○One(1) attack ○Two(2) attacks ○Three(3) attacks ○More than 3 attacks
8	Please rate your cough on a scale of 0 to 5 with 0 meaning *I never cough* to 5 meaning *I cough all the time*	0
9	Please rate your sleep on a scale of 0 to 5 with 0 meaning *I sleep soundly* to 5 meaning *I do not sleep soundly* because of my lung condition?	0
10	Please rate your chest tightness on a scale of 0 to 5 with 0 meaning *no chest tightness at all* to 5 meaning *my chest feels very tight*	0
11	Does your health now limit you in vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports?	○Yes, limited a lot ○Yes, limited a little ○No, not limited at all
12	How much of the time during the past 4 weeks, did you feel tired?	○All of the time ○Most of the time ○Some of the time ○A little of the time ○None of the time
13	During the past 4 weeks, how much of the time have you had to cut down on the amount of time you spent on work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious?	○All of the time ○Most of the time ○Some of the time ○A little of the time ○None of the time

Figure 2B: Example of calculated risk scores for asthma and COPD for a person using the Online UCAP Questionnaire



The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

UCAP

UCAP Questionnaire

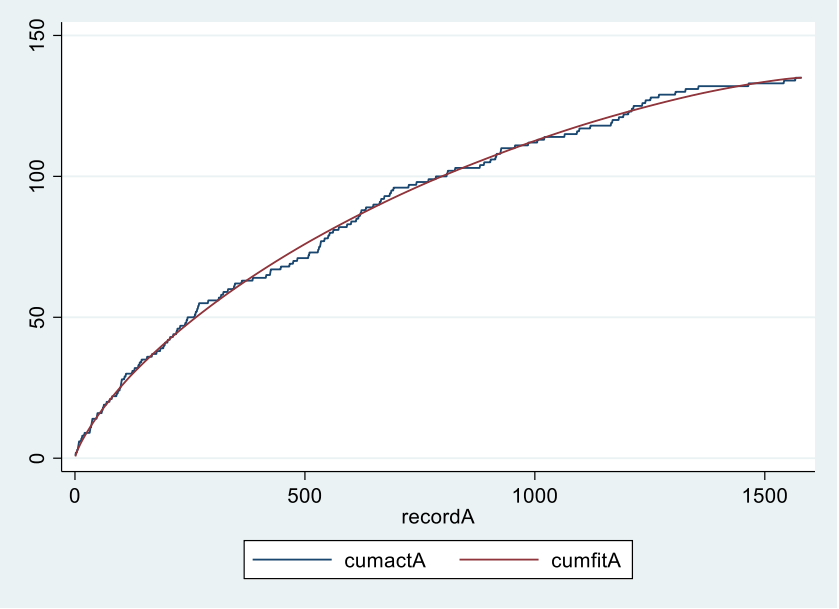
1	How old are you?	67
2	Are you currently smoking cigarettes or have you smoked cigarettes in the past?	◉ Yes ○ No
	If Yes, what is the average number of cigarette packs per day	1 *(# packs per day)*
	If Yes, what is the number of years you have spent smoking	30 *(# years)*
		30 *(pack years)*
3	Are you regularly exposed to cigarette smoke on a daily basis (either from yourself, or from people around you)?	○ Yes ◉ No
4a	Have you ever worked for 3 months or more with paint, chemicals, or fumes?	◉ Yes ○ No
	If Yes, how many months did you work?	24
4b	Have you ever worked for 3 months or more with sandblasting?	○ Yes ◉ No
5	Are you currently taking the medication Salbutamol (also known as Ventolin) for your breathing?	○ Yes ◉ No
6	If you have a wheeze, is it worse in the morning?	◉ Yes ○ No ○ Not Applicable
7	During the past 3 months, how many severe or very unpleasant attacks of chest problems (attacks of shortness of breath or wheezing) have you had?	○ No (0) attack ◉ One(1) attack ○ Two(2) attacks ○ Three(3) attacks ○ More than 3 attacks
8	Please rate your cough on a scale of 0 to 5 with 0 meaning *I never cough* to 5 meaning *I cough all the time*	3
9	Please rate your sleep on a scale of 0 to 5 with 0 meaning *I sleep soundly* to 5 meaning *I do not sleep soundly* because of my lung condition?	0
10	Please rate your chest tightness on a scale of 0 to 5 with 0 meaning *no chest tightness at all* to 5 meaning *my chest feels very tight*	1
11	Does your health now limit you in vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports?	○ Yes, limited a lot ◉ Yes, limited a little ○ No, not limited at all
12	How much of the time during the past 4 weeks, did you feel tired?	○ All of the time ○ Most of the time ◉ Some of the time ○ A little of the time ○ None of the time
13	During the past 4 weeks, how much of the time have you had to cut down on the amount of time you spent on work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious?	○ All of the time ○ Most of the time ○ Some of the time ○ A little of the time ◉ None of the time

UCAP Questionnaire Asthma Risk Score: 7.8%

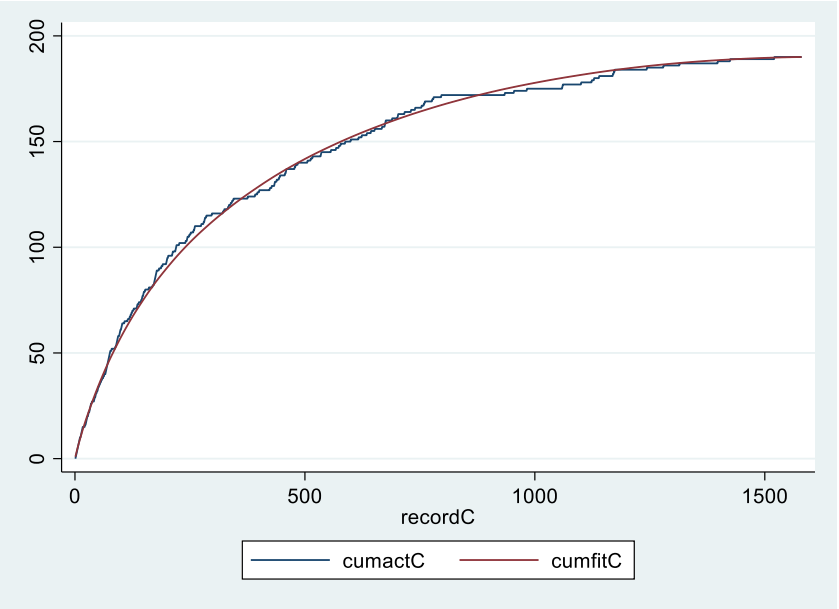
UCAP Questionnaire COPD Risk Score: 11.84%

Figure 3: Calibration/discrimination plots of cumulative cases and cumulative probability estimates among the 1580 subjects in the derivation sample.

A: Asthma



B: COPD



References:

1. Soriano JB, Abajobir AA, Abate KH, Abera SF, Agrawal A, Ahmed MB, et al. 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. The Lancet Respiratory Medicine. 2017 Sep;5(9):691–706.

2. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management and Prevention of COPD [Internet]. 2020 p. 1–141. Available from: https://goldcopd.org/

3. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2016 Jan;21(1):14–23.

4. 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 Oct;148(4):971–85.

5. Ho T, Cusack RP, Chaudhary N, Satia I, Kurmi OP. Under- and over-diagnosis of COPD: a global perspective. Breathe. 2019 Mar;15(1):24–35.

6. Evans J, Chen Y, Camp PG, Bowie DM, McRae L. Estimating the prevalence of COPD in Canada: Reported diagnosis versus measured airflow obstruction. Health Reports. 2014;25(82):1–11.

7. Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet L-P, et al. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov;198(9):1130–9.

8. Aaron SD, Boulet LP, Reddel HK, Gershon AS. Underdiagnosis and Overdiagnosis of Asthma. Am J Respir Crit Care Med. 2018 Oct 15;198(8):1012–20.

9. Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, et al. Health Services Burden of Undiagnosed and Overdiagnosed COPD. Chest. 2018 Jun;153(6):1336–46.

10. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. Respirology. 2017 Apr;22(3):575–601.

11. van Schayck CP. Underdiagnosis of asthma: is the doctor or the patient to blame? The DIMCA project. Thorax. 2000 Jul 1;55(7):562–5.

12. Price D, Freeman D, Cleland J, Kaplan A, Cerasoli F. Earlier diagnosis and earlier treatment of COPD in primary care. Primary Care Respiratory Journal. 2010 Sep 24;20(1):15–22.

13. Broekhuizen BD, Sachs AP, Hoes AW, Moons KG, van den Berg JW, Dalinghaus WH, et al. Undetected chronic obstructive pulmonary disease and asthma in people over 50 years with persistent cough. Br J Gen Pract. 2010 Jul 1;60(576):489–94.

14. US Preventive Services Task Force (USPSTF), Siu AL, Bibbins-Domingo K, Grossman DC, Davidson KW, Epling JW, et al. Screening for Chronic Obstructive Pulmonary Disease: US Preventive Services Task Force Recommendation Statement. JAMA. 2016 Apr 5;315(13):1372.

15. Ilmarinen P, Juboori H, Tuomisto LE, Niemelä O, Sintonen H, Kankaanranta H. Effect of asthma control on general health-related quality of life in patients diagnosed with adult-onset asthma. Scientific Reports. 2019 Nov 6;9(1):16107.

16. Yawn BP, Mapel DW, Mannino DM, Martinez FJ, Donohue JF, Hanania NA, et al. Development of the Lung Function Questionnaire (LFQ) to identify airflow obstruction. Int J Chron Obstruct Pulmon Dis. 2010;5:1–10.

17. Martinez FJ, Mannino D, Leidy NK, Malley KG, Bacci ED, Barr RG, et al. A New Approach for Identifying Patients with Undiagnosed Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2017 Mar 15;195(6):748–56.

18. Franco-Marina F, Fernandez-Plata R, Torre-Bouscoulet L, García-Sancho C, Sanchez-Gallen E, Martinez D, et al. Efficient screening for COPD using three steps: a cross-sectional study in Mexico City. Primary Care Respiratory Medicine. 2014 May 20;24(1):1–8.

19. Calverley PMA, Nordyke RJ, Halbert RJ, Isonaka S, Nonikov D. Development of a Population-Based Screening Questionnaire for COPD. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2005 Jan;2(2):225–32.

20. Price DB, Tinkelman DG, Nordyke RJ, Isonaka S, Halbert RJ. Scoring System and Clinical Application of COPD Diagnostic Questionnaires. Chest. 2006 Jun 1;129(6):1531–9.

21. Varela MVL, Oca MM de, Rey A, Casas A, Stirbulov R, Boscio VD. Development of a simple screening tool for opportunistic COPD case finding in primary care in Latin America: The PUMA study. Respirology. 2016;21(7):1227–34.

22. Martinez FJ, Raczek AE, Seifer FD, Conoscenti CS, Curtice TG, D'Eletto T, et al. Development and Initial Validation of a Self-Scored COPD Population Screener Questionnaire (COPD-PS). COPD: Journal of Chronic Obstructive Pulmonary Disease. 2008 Jan 1;5(2):85–95.

23. Kotz D, Nelemans P, van Schayck CP, Wesseling GJ. External validation of a COPD diagnostic questionnaire. European Respiratory Journal. 2008 Feb 1;31(2):298–303.

24. Guirguis-Blake JM, Senger CA, Webber EM, Mularski RA, Whitlock EP. Screening for Chronic Obstructive Pulmonary Disease: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2016 Apr 5;315(13):1378–93.

25. Shin B, Cole S, Park S-J, Ledford D, Lockey R. A New Symptom-Based Questionnaire for Predicting the Presence of Asthma. J Investig Allergol Clin Immunol. 2010;20:8.

26. Sá-Sousa A, Pereira AM, Almeida R, Araújo L, Couto M, Jacinto T, et al. Adult Asthma Scores—Development and Validation of Multivariable Scores to Identify Asthma in

Surveys. The Journal of Allergy and Clinical Immunology: In Practice. 2019 Jan 1;7(1):183–90.

27. Buist AS. Similarities and differences between asthma and chronic obstructive pulmonary disease: treatment and early outcomes. European Respiratory Journal. 2003 Jan 1;21:30–5.

28. Preteroti M, Whitmore GA, Vandemheen KL, FitzGerald JM, Lemière C, Boulet L-P, et al. Population-based case-finding to identify subjects with undiagnosed asthma or COPD. European Respiratory Journal. 2020 Jun 1;55(6):2000024.

29. Jones PW, Harding G, Berry P, Wiklund I, Chen W-H, Kline Leidy N. Development and first validation of the COPD Assessment Test. European Respiratory Journal. 2009 Sep 1;34(3):648–54.

30. RAND Health Care. 36-Item Short Form Survey Instrument (SF-36) [Internet]. RAND Corporation. Available from: https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form/survey-instrument.html

31. Reilly M, Zbrozek A, Dukes E. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4(5):353–65.

32. Jones P, Littlejohns L, Quirk FH. A self-complete measure of health status for chronic airflow limitation. Am Rev Respir Dis. 1992;145(6):1321–7.

33. W. Hosmer D, Lemeshow S. Applied Logistic Regression. 3rd ed. Hoboken, New Jersey: John Wiley & Sons; 2013. 518 p.

34. Flesch R. A new readability yardstick. Journal of Applied Psychology. 1948 Jun;32(3):221–33.

35. Gorman BK, Chu M. Racial and ethnic differences in adult asthma prevalence, problems, and medical care. Ethn Health. 2009 Oct;14(5):527–52.

36. Fragoso CAV, McAvay G, Gill TM, Concato J, Quanjer PH, Ness PHV. Ethnic differences in respiratory impairment. Thorax. 2014 Jan 1;69(1):55–62.

Derivation and Validation of the UCAP-Q Case-finding Questionnaire to Detect Undiagnosed Asthma and COPD

Online Data Supplement:

eTable 1. List of predictor variables assessed for the candidate pool based on their univariate significance levels for association with OLD.

Do you have worsening of the following symptoms when you lie down to sleep? • Cough • Chest tightness • Wheeze	 0.227 0.002 0.040
Do you have worsening of the following symptoms after exercise or physical activity? • Chest tightness • Wheeze • Short of breath	 0.066 0.144 0.003
Do you have worsening of the following symptoms after laughing or crying? • Cough • Chest tightness • Wheeze	 0.002 0.152 0.186
Age, years	0.000
Sex	0.000
Have you ever worked for 3 months or more at any of the following occupations: hard-rock mining, coal mining, sandblasting, working with asbestos, chemical or plastics manufacturing, flour, feed or grain milling, cotton or jute processing, foundry or steel milling, welding, firefighting, farming, forestry, saw-milling, work with paint, chemicals, or fumes	0.000
Are you regularly exposed to any of the following daily? • Smoking (primary or second-hand)	 0.000
Are you currently smoking cigarettes?	0.000
Have you every smoked pot/marijuana?	0.231
In the past 12 months, were you hospitalized for any breathing problems or respiratory illness?	0.050
Are you currently taking any medications for your breathing? Ventolin	0.005
I never cough to I cough all the time	0.318
I have no phlegm (mucus) in my chest at all to My chest is completely full of phlegm	0.047
My chest does not feel tight at all to My chest feels very tight	0.091
I am confident leaving my home despite my lung condition to I am not at all confident leaving my home because of my lung condition	0.185

I sleep soundly to I don't sleep soundly because of my lung condition	0.136
Over the past 3 months, I have had shortness of breath	0.047
During the past 3 months, how many severe or very unpleasant attacks of chest problems have you had?	0.014
If you have a wheeze, is it worse in the morning?	0.000
Questions about what activities usually make you feel breathless these days: • Sitting or lying still	0.005
Some more questions about your cough and breathlessness these days: • My cough hurts	0.013
I feel that I am not in control of my chest problem	0.079
• If I climb up one flight or stairs, I have to go slow or stop • If I hurry or walk fast, I have to stop or slow down	0.062 0.076
Does your health now limit you in these activities? If so, how much? • Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	0.000
• Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf	0.002
• Walking more than a kilometer	0.072
During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious? • Cut down on the amount of time you spent on work or other activities	0.227
How much of the time during the past 4 weeks…?	
• Have you been happy? • Did you feel worn out?	0.249 0.153
• Did you feel tired?	0.052
During the past 7 days, how many hours did you miss from work because of your breathing problems?	0.038
During the past 7 days, how much did your breathing problems affect your productivity while you were working?	0.011

eTable 2. Risk Scoring for Asthma based on the Multinomial Logistic Regression Model

Question [Scale]	β	95% CI	p-value
How old are you? [years]	-0.0044	-0.018-0.009	0.514
Are you currently smoking cigarettes or have you smoked cigarettes in the past? [pack-years]	0.0022	-0.010-0.014	0.714
Are you regularly exposed to cigarette smoke (either from yourself, or from people around you) on a daily basis? [Yes/No]	-0.4176	-0.916-0.081	0.101
Have you ever worked for 3 months or more with paint, chemicals, or fumes? Have you ever worked for 3 months or more with sandblasting? [months]	0.9238	0.323-1.524	0.003
Are you currently taking the medication Salbutamol (also known as Ventolin) for your breathing? [Yes/No]	0.4024	-0.1051-0.910	0.120
If you have a wheeze, is it worse in the morning? [Not Applicable/Yes/No]	0.5882	0.192-0.985	0.004
During the past 3 months, how many severe or very unpleasant attacks of chest problems (attacks of shortness of breath or wheezing) have you had? [More than 3 attacks/3 attacks/2 attacks/1 attack/No attack]	0.8841	0.242-1.526	0.007
Please rate your cough on a scale of 0 to 5 with 0 meaning I never cough to 5 meaning I cough all the time.	1.448	0.299-2.60	0.013
Please rate your sleep on a scale of 0 to 5 with 0 meaning I sleep soundly to 5 meaning I do not sleep soundly because of my lung condition.	0.1511	-1.023-1.325	0.801
Please rate your chest tightness on a scale of 0 to 5 with 0 meaning no chest tightness at all to 5 meaning my chest feels very tight.	0.8968	0.195-1.60	0.012

Does your health now limit you in vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports? [Yes, limited a lot/Yes, limited a little/No, not limited at all]	0.0877	-0.366-0.542	0.705
How much of the time during the past 4 weeks, did you feel tired? [All of the time/Most of the time/Some of the time/A little of the time/None]	-0.2959	-0.962-0.370	0.384
During the past 4 weeks, how much of the time have you had to cut down on the amount of time you spent on work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious? [All of the time/Most of the time/Some of the time/A little of the time/None]	1.0958	0.334-1.86	0.005
Constant	9.53375		

eTable 3. Risk Scoring for COPD based on Multinomial Logistic Regression Model

Question [Scale]	β	95% CI	p-value
How old are you? [years]	0.0420	0.025-0.059	0.000
Are you currently smoking cigarettes or have you smoked cigarettes in the past? [pack-years]	0.0308	0.023-0.039	0.000
Are you regularly exposed to cigarette smoke (either from yourself, or from people around you) on a daily basis? [Yes/No]	1.2363	0.833-1.64	0.000
Have you ever worked for 3 months or more with paint, chemicals, or fumes? Have you ever worked for 3 months or more with sandblasting? [months]	0.0116	-0.941-0.965	0.981
Are you currently taking the medication Salbutamol (also known as Ventolin) for your breathing? [Yes/No]	0.8238	0.341-1.31	0.001
If you have a wheeze, is it worse in the morning? [Not Applicable/Yes/No]	0.5788	0.198-0.960	0.003
During the past 3 months, how many severe or very unpleasant attacks of chest problems (attacks of shortness of breath or wheezing) have you had? [More than 3 attacks/3 attacks/2 attacks/1 attack/No attack]	-0.1735	-0.874-0.527	0.627
Please rate your cough on a scale of 0 to 5 with 0 meaning I never cough to 5 meaning I cough all the time.	0.8255	-0.214-1.86	0.119
Please rate your sleep on a scale of 0 to 5 with 0 meaning I sleep soundly to 5 meaning I do not sleep soundly because of my lung condition.	1.4280	0.316-2.54	0.012
Please rate your chest tightness on a scale of 0 to 5 with 0 meaning no chest tightness at all to 5 meaning my chest feels very tight.	-0.1256	-0.740-0.489	0.689
Does your health now limit you in vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports? [Yes, limited a lot/Yes,	0.9037	0.413-1.39	0.000

limited a little/No, not limited at all]			
How much of the time during the past 4 weeks, did you feel tired? [All of the time/Most of the time/Some of the time/A little of the time/None]	1.1871	0.476-1.90	0.001
During the past 4 weeks, how much of the time have you had to cut down on the amount of time you spent on work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious? [All of the time/Most of the time/Some of the time/A little of the time/None]	0.0011	-0.635-0.638	0.997
Constant	1.9512		

eTable 4. Associated sensitivity, specificity, PPV, and NPV values (%) for UCAP-Q at varying risk cut-offs for each disease (derivation cohort)

	Asthma	**COPD**	**No OLD**		
Risk Cut-Off (%)	**Sensitivity**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
5	97	99	10	22	95
6	92	97	17	23	92
10	64	82	47	27	88
15	38	70	72	34	86
20	24	61	85	44	85

PPV= positive predictive value

NPV= negative predictive value

eTable 5. Test-retest reliability of the UCAP-Q (N=27)

Question	ICC	95% CI
How old are you?	0.999	0.999-0.999
Are you currently smoking cigarettes or have you smoked cigarettes in the past? If yes, calculate total pack-years.	0.895	0.783-0.951
Have you ever worked for 3 months or more with paint, chemicals, or fumes? If yes, how many months did you work?	0.592	0.280-0.790
Have you ever worked for 3 months or more with sandblasting? If yes, how many months did you work?	1.00	1.00-1.00
	Weighted Kappa	**95% CI**
Are you regularly exposed to cigarette smoke (either from yourself, or from people around you) on a daily basis?	1.00†	1.00-1.00
Are you currently taking the medication Salbutamol (also known as Ventolin) for your breathing?	1.00†	1.00-1.00
If you have a wheeze, is it worse in the morning?	1.00†	1.00-1.00
During the past 3 months, how many severe or very unpleasant attacks of chest problems (attacks of shortness of breath or wheezing) have you had?	0.763	0.528-0.999
Please rate your cough on a scale of 0 to 5 with 0 meaning I never cough to 5 meaning I cough all the time.	0.877	0.766-0.987
Please rate your sleep on a scale of 0 to 5 with 0 meaning I sleep soundly to 5 meaning I do not sleep soundly only because of my lung condition.	0.871	0.745-0.996
Please rate your chest tightness on a scale of 0 to 5 with 0 meaning no chest tightness at all to 5 meaning my chest feels very tight.	0.929	0.845-1.00

Does your health now limit you in vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports?	0.923	0.825-1.00
How much of the time during the past 4 weeks, did you feel tired?	0.739	0.399-1.00
During the past 4 weeks, how much of the time have you had to cut down on the amount of time you spent on work or other regular activities as a result of any emotional problems, such as feeling depressed or anxious?	0.853	0.749-0.956

[†]Unweighted Kappa

Section E.1. Development of a Case Finding Model by Stepwise Methods

We have used stepwise multinomial logistic regression to select our case-finding model. Use of stepwise variable selection for model development in case finding can be very effective and reliable if employed appropriately. Here are aspects of our specific application and implementation that make this method a good choice to meet the requirements of the application:

1. Our variables are derived from a standard demographic questionnaire and a set of clinically tested questionnaires from the field of respirology. Thus, our candidate pool of variables is a collection of clinically relevant variables that have been found pertinent in related research contexts.
2. Our application requires a final questionnaire with only a few clinically plausible questions that will be easily understood by subjects. The selection method generates such a minimal set from our candidate pool of variables.
3. The variables in our large initial pool have varying degrees of similarity, overlap and collinearity that make it necessary to process the pool in a manner that takes account of their statistical interrelationships. The method must cull variables in a statistically coherent manner.
4. The model building is aimed at case finding. The variable selection should guard against both false positives and false negatives for the variables chosen.
5. Our priority is to discover true cases of disease; not to make inferences about direct effects of predictor variables on the case outcomes. Variables that are selected for the model are not expected to represent only their own direct association with the outcome. Rather, each selected variable captures the varied influences of variables correlated with the selected variable that are displaced by it in the selection process. The true effect size for the selected variable alone (as a population parameter) is indeterminable in practical terms. For example, the selected SGRQ question, “If you have wheeze, is it worse in the morning?”, stands in for several other similar questions that are also found in the standard questionnaires from which our variable pool was derived (such as “ Do you have worsening of [wheeze] when you lie down to sleep?” from the ASQ).
6. We have just made the point that selected variables are often chosen over similar variables by the selection process. We do not attempt to build composites of related variables (such as principal components, for instance) because our experience shows that it is sometimes subtle differences in the wording of a question that gives it a predictive edge. Forming a composite variable tends to dilute these nuanced effects of wording.

7. Multinomial logistic regression has the advantage of providing a logical probability model for the three mutually exclusive diagnostic outcomes for each eligible subject, which is the reality of our case-finding scenario where both cases of asthma and COPD are to be discovered. Case finding for multiple related variables is an uncommon study context.

Other model selection methods were considered for use in this research project. A full application of Classification and Regression Trees (CART) was developed in Chau (2021) but was eventually set aside as not suited to our needs, as described in her thesis. Although the statistical performance of the CART model roughly matched the stepwise multinomial logistic model, it was decided that the clinical implementation of the model and its clinical acceptability would prove problematic.
Variable selection by backward elimination, forward selection and removal-entry stepwise methods are sequential testing applications. The literature gives clear examples of situations where these methods have been inappropriately used. Yet, there are many applications that are built around effective and suitable use of sequential testing of the type found in stepwise regression. For example, Aaron et al (2015) and Stanojevic et al (2019) present a successful application (with external validation) for a stepwise logistic predictive model for cystic fibrosis mortality. As another example from a different field, the multiple testing challenges of gene discovery research have led to important research advances in statistical inference. An early advance is found in the analysis of microarray data looking at gene expression in large collections of genes (Lee, 2004). Benjamini and Hochberg (1995) developed a sequential testing framework that adapts readily to the stepwise context. Several variations of their procedure have also been developed. These and other methods have been found to be robust, flexible and reliable applications of sequential testing in model development.

Section E.2. Calibration/Discrimination Plots for the Multinomial Logistic Model

The multinomial logistic model is well calibrated if probability estimates for outcomes of subjects accurately represent the true probabilities that the outcomes will occur. The model is discriminating if probability estimates for outcomes are higher for subjects with disease than subjects with no disease.
Separate calibration/discrimination plots for asthma and COPD are given in panels (a) and (b) of Figure 3 of the main text. Our calibration/discrimination plot for each outcome type (asthma, COPD) is a graph of two cumulative sums. First, we sort the probability estimates of the disease outcome for all subjects from largest to smallest. Second, we calculate the cumulative sum of probability estimates across the records of the ordered subjects. Third, we calculate the cumulative sum of the indicator values (0 or 1) for the actual disease outcomes across the records of all subjects. Finally, we plot the cumulative sum of disease cases and the cumulative sum of the probability estimates against the ordered record index of all subjects. The extent to which the two curves track each other is an indication of the calibration of our multinomial logistic model. The degree to which the curves bend toward the upper

left-hand corner of the graph measures the extent to which the model discriminates disease cases from subjects with no disease. Refer to Aaron et al (2015) and their supplement for technical background and further illustrations of these goodness-of-fit plots for prediction models.

Section E.3. Hosmer-Lemeshow Goodness-of-Fit Table for the Multinomial Logistic Model

The table below compares observed and expected frequencies across three disease classes for 10 categories of estimated outcome probabilities. The categories are constructed to have equal numbers of subjects (n=1580 in total). The frequency data show few major discrepancies between observed and expected numbers across the probability range. The codes for the three classification outcomes for disease status are: no OLD (code 0), asthma (code 1) and COPD (code 2). The table also shows the test statistic, degrees of freedom, and P-value for the Hosmer-Lemeshow test of goodness of fit.

```
. mlogitgof, table

Goodness-of-fit test for a multinomial logistic regression model
Dependent variable: AsthmaCOPD1

Table: observed and expected frequencies
+-------------------------------------------------------------------+
|Group |   Prob | Obs_2  Exp_2 | Obs_1  Exp_1 | Obs_0  Exp_0 | Total |
|------+--------+--------------+--------------+--------------+-------|
|    1 | 0.0783 |     3   2.83 |     6   6.88 |   149 148.29 |   158 |
|    2 | 0.1007 |     5   4.76 |    11   9.29 |   142 143.95 |   158 |
|    3 | 0.1193 |     3   6.17 |    12  11.25 |   143 140.58 |   158 |
|    4 | 0.1369 |    15   8.12 |    10  12.18 |   133 137.70 |   158 |
|    5 | 0.1600 |    14   9.21 |    14  14.26 |   130 134.53 |   158 |
|------+--------+--------------+--------------+--------------+-------|
|    6 | 0.1914 |     6  12.41 |    20  14.99 |   132 130.60 |   158 |
|    7 | 0.2292 |    13  16.76 |    14  16.27 |   131 124.97 |   158 |
|    8 | 0.2824 |    18  21.65 |    18  18.63 |   122 117.72 |   158 |
|    9 | 0.3890 |    38  33.63 |    18  18.73 |   102 105.63 |   158 |
|   10 | 0.9210 |    75  74.44 |    12  12.53 |    71  71.03 |   158 |
+-------------------------------------------------------------------+
  number of observations =   1580
number of outcome values =      3
      base outcome value =      0
        number of groups =     10
```

```
chi-squared statistic =     19.208
   degrees of freedom =     16
   Prob > chi-squared =      0.258
```

Section E.4. Establishing Risk Thresholds That Minimize Total Cost of Classification Errors

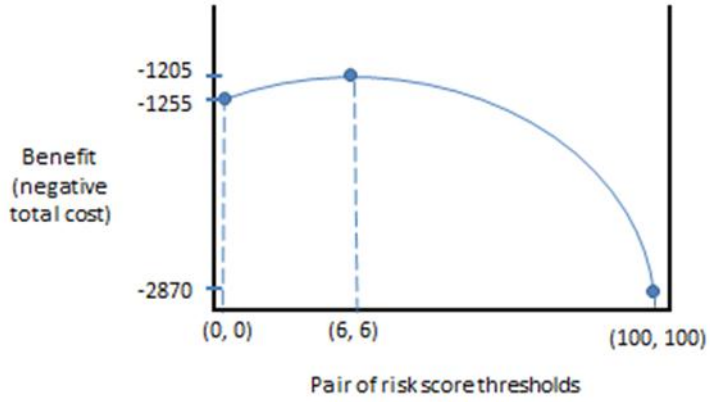
The main text explains that clinical judgment was used to assess the unit costs of prediction or classification errors for a false positive outcome (FP) and for the two types of false negative outcomes (FN1 for asthma and FN2 for COPD). Without loss of generality, the FP unit cost was set at c0=1. The unit costs for the false negatives were set at c1=10 for asthma and c2=8 for COPD. We set out to find two risk thresholds that minimize the total cost of classification errors. We denote the thresholds by t1 for the asthma risk score and t2 for the COPD risk score. The risk scores are percentages so we search for over all 10,000 pairs (t1, t2) of the natural numbers 1 to 100, evaluating the following cost function for each pair:

Total Cost(t1,t2)=c0*FP(t1,t2)+c1*FN!(t1,t2)+c2*FN(t1,t2)

As each subject in the derivation sample has a pair of calculated risk scores (r1, r2), a simple computer program was written to count subjects representing FP, FN1 and FN2 classification errors by comparing (r1, r2) to (t1, t2) for each subject. The evaluation of the cost function for all 10,000 (t1, t2) threshold pairs takes little computing time and the minimum cost pair is easily identified. In principle, multiple optima may occur but not in our application. The 10,000-pair search gives thresholds that are integer percentages, which are accurate enough for this application. The optimal threshold pair for our derivation sample in this study is (5, 7). We note, however, that in the actual implementation of this approach for the UCAP case-finding study, this calculation was made before fixing the final data set so that the study could operationally adopt the optimal threshold approach. The optimal thresholds at that earlier calculation were (6, 6) so a 6 percent threshold was adopted for each disease. Sensitivity analysis of the optimal thresholds and the assigned unit costs c1 and c2 for classification errors shows that the total cost level does not vary greatly in the neighborhood of the exact optimum so our approximate thresholds and unit costs are adequate for operational purposes.

Section E.5. Decision curve showing the tradeoff between benefit (negative total cost of classification errors) and the choice of the thresholds for risk scores used in case finding.

This visual is a stylistic representation, in that the threshold pair is actually a point on a plane rather than a point on a single axis of the graph. The plotting is not to scale.


-1205
-1255
Benefit
(negative
total cost)
-2870
(0, 0)
(6, 6)
(100, 100)
Pair of risk score thresholds

Cited references

Aaron SD, Stephenson AL, Cameron DW, Whitmore GA (2015). A statistical model to predict one-year risk of death in patients with cystic fibrosis, Journal of Clinical Epidemiology, 68, 1336-1345 (and online supplement).

Stanojevic S, Sykes J, Stephenson AL, Aaron SD, Whitmore GA (2019). Development and external validation of 1- and 2- year mortality prediction models in cystic fibrosis, European Respiratory Journal, 54:1900224 [https://doi.org/ 10.1183/13993003.00224-2019].

Lee M-LT (2004). *Analysis of Microarray Gene Expression Data*, Springer

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.