

Supplement for:

Population-Based Case-Finding to Identify Subjects with Undiagnosed Asthma or COPD

S.1 Table of Sensitivity and Specificity Values for the 8-item Questionnaire

The 8-item questionnaire and associated logistic regression coefficients for assessing probability of OLD is given in Table 4 of the main article. The ROC curve for this questionnaire appears in Figure 2 of the main article. The AUC of this curve is 0.74. Table S.1 here gives the sensitivity and specificity values of this questionnaire for obstructive lung disease at a selection of cutoff values for the probability of OLD.

Probability of OLD (%)	Sensitivity (%)	Specificity (%)
5	99	7
10	89	30
15	79	54
20	67	69
25	58	79

Table S.1: Sensitivity and specificity values of the 8-item questionnaire for obstructive lung disease at a selection of cutoff values for the probability of OLD.

S.2 Illustration of Risk Scoring of OLD for a Demonstration Subject

Question	Response	Regression Coefficient	Score Contribution
What is your age?	36	0.0201	0.722
Are you male?	Yes (value=1)	0.395	0.395
What is your BMI?	33.52 kg/m ²	-0.0441	-1.478
How many pack years have you smoked?	15	0.0236	0.353
Do you experience worsening of wheezing following physical activity?	No (value=0)	0.604	0.000
Do you experience shortness of breath following physical activity?	Yes (value=1)	0.526	0.526
I have no phlegm in my chest at all (0) – my chest is completely full of phlegm (5)	“3” (value= -1.419)	0.802	-1.138
I sleep soundly (0) – I don’t sleep soundly because of my chest condition (5)	“2” (value=-1.229)	1.045	-1.284
Constant			-0.054
Total Logistic Score			-1.958

Table S.2: Risk scoring for a demonstration subject using the 8-item questionnaire for obstructive lung disease.

Table S.2 illustrates the risk scoring calculations for a demonstration subject using the 8-item questionnaire for obstructive lung disease. By summing the score contributions of the subject’s

responses to the eight questions and then adding the regression constant term, one obtains a total logistic score of -1.958. The risk is calculated from the total score using the following logistic formula:

$$\exp(-1.958) / [1 + \exp(-1.958)] = 0.124.$$

This score corresponds to a probability of 0.124 or a 12% risk that the subject has OLD.

We note that the last two questions are taken from the CAT questionnaire and have respective rating responses of “3” and “2” for this demonstration subject. These two ratings are assigned scores of -1.419 and -1.229, respectively, for the risk score calculation. Refer to Section S.3 below for an explanation of how values are assigned to the CAT responses for risk assessment with the 8-item questionnaire. Table S.4 is a reference table for these assigned values.

S.3 Scoring of Questions in the CAT Questionnaire

The eight symptom questions in the COPD Assessment Test (CAT) questionnaire are each structured as a six-point ordinal scale. The scale is a progression of symptom severity from none (0) to extreme (5). In spite of the ordinal construction, however, the relationships of the ratings to the outcome probabilities for OLD, asthma, or COPD are not monotonic. Subjects in our study who have the disease tend to respond at intermediate levels of severity, avoiding the most extreme response. The CAT question about ‘phlegm in my chest’ provides a convincing example of this pattern. Phlegm symptoms are strongly indicative of OLD. The relative frequency of OLD among subjects tends to rise through ratings 0 to 4 on the phlegm question of CAT. Then, the relative frequency drops significantly at a rating of 5. This point is illustrated in Table S.3.

CAT phlegm rating	All Subjects	Subjects with OLD	Percentage with OLD
0	86	14	16.3
1	177	24	13.6
2	181	37	20.4
3	260	51	19.6
4	136	46	33.8
5	53	10	18.9
Total	893	182	20.4

Table S.3: Percentage of subjects with OLD among subjects giving different ratings to the CAT phlegm question. The table only includes subjects with complete records who were used to estimate the 8-item risk model (n=893).

The non-monotonic relationship between outcome risk for OLD and the rating response for each CAT question led us to score the CAT question using a fitted logistic function. In particular, logistic regression was used to relate disease outcome to the question response, treating the latter as a categorical variable. The fitted value of the logistic regression function for each subject is the subject’s risk score for that question. Mathematically this score is the estimated log-odds for the disease outcome based on the subject’s individual response to the CAT question. As an illustration, Table S.4 gives the predicted scores from these fitted logistic regression functions for the CAT questions for phlegm and sleep, both of which

feature in the 8-item risk model for OLD presented in Table 4 of the main article. These scores define predictor variables *zphlegm* and *zsleep* in the 8-item model. For example, a subject reporting ratings of “3” for phlegm and “2” for sleep would have scores of *zphlegm* = -1.419 and *zsleep* = -1.229 for these two symptoms.

CAT rating	Phlegm		Sleep	
	All Subjects	Logistic Score	All Subjects	Logistic Score
0	86	-1.651	171	-1.798
1	177	-1.872	176	-1.379
2	181	-1.386	185	-1.229
3	260	-1.419	197	-1.325
4	136	-0.650	112	-1.147
5	53	-1.459	52	-1.316
Total	893		893	

Table S.4: Logistic scores for phlegm and sleep computed from logistic regression functions that relate the OLD outcome to CAT question ratings for the 893 subjects whose records entered the 8-item risk model.