Accuracy of History, Wheezing, and Forced Expiratory Time in the Diagnosis of Chronic Obstructive Pulmonary Disease

Sharon E. Straus, MD, Finlay A. McAlister, MD, David L. Sackett, MD, Jonathan J. Deeks, MSc,
On Behalf of the CARE-COAD2 Group

OBJECTIVE: To determine the accuracy of the history and selected elements of the physical examination in the diagnosis of chronic obstructive pulmonary disease (COPD).

DESIGN: Independent blind comparison of the standard clinical examination (evaluating the accuracy of history, wheezing, and forced expiratory time [FET]) with spirometry. The gold standard for diagnosis of COPD was a forced expiratory volume at 1 second (FEV₁) below the fifth percentile (adjusted for patient height and age).

SETTING: Seven sites in 6 countries, including investigators from primary care and secondary care settings.

PARTICIPANTS: One hundred sixty-one consecutive patients with varying severity of disease (known COPD, suspected COPD, or no COPD) participated in the study.

MAIN RESULTS: One hundred sixty-one patients (mean age 65 years, 39% female, 41% with known COPD, 27% with suspected COPD, and 32% normal) were recruited. Mean ($\pm$ SD) FEV₁ and forced vital capacity were 1,720 ($\pm$ 830) mL and 2,520 ($\pm$ 970) mL. The likelihood ratios (LR) for the tested elements of the clinical examination (and their *P* values on χ^2 testing) were: self-reported history of COPD, 5.6 (*P* < .001); FET greater than 9 seconds, 6.7 (*P* < 0.01); smoked longer than 40 pack years, 3.3 (*P* = .001); wheezing, 4.0 (*P* < .001); male gender, 1.6 (*P* < .001); and age over 65 years, 1.6 (*P* = .025). The accuracy of these elements was not appreciably different when reference standards other than FEV₁ below the 5th percentile were applied. Only 3 elements of the clinical examination were significantly associated with the diagnosis of COPD on multivariate analysis: self-reported history of COPD (adjusted LR 4.4), wheezing (adjusted LR 2.9), and FET greater than 9 seconds (adjusted LR 4.6). Area under the receiver operating characteristic curve for the model incorporating these 3 factors was 0.86.

CONCLUSIONS: Less emphasis should be placed on the presence of isolated symptoms or signs in the diagnosis of COPD. While numerous elements of the clinical examination are associated with the diagnosis of COPD, only 3 are significant on multivariate analysis. Patients having all 3 of

these findings have an LR of 59 (ruling in COPD); those with none have an LR of 0.3 (ruling out COPD).

KEY WORDS: chronic obstructive pulmonary disease; clinical examination; spirometry; diagnosis.

J GEN INTERN MED 2002;17:684-688.

A consistent theme emerging from the clinical examination literature is its scarcity and generally poor quality. Due to the paucity of methodologically rigorous studies, the reported accuracy of symptoms and signs vary greatly between these studies, impeding their usefulness for the practicing clinician.¹ For example, the accuracy of the history and physical examination in detecting chronic obstructive pulmonary disease (COPD) has been infrequently evaluated. Systematic review of the COPD literature revealed a preponderance of signs (32) that had been tested in a number of small studies (median 1 sign, 2 clinicians, and 93 patients per study) of poor methodological rigor.¹ Indeed, only 1 study identified in the review reported scientifically robust methods (an independent, blind comparison with a reference standard among an appropriate spectrum of consecutive patients).² Moreover, the reported accuracy of symptoms or signs varies sharply among studies of patients with COPD. For example, in a recent systematic review, it was observed that the detection of wheezing on auscultation has been evaluated in 7 studies and that sensitivity ranged from 9% to 100%, specificity from 37% to 100%, and positive likelihood ratios (LRs) varied from 0.9 to infinity.³

In an effort to address this shortcoming, the multinational Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease (CARE-COAD) group was formed with the mandate to conduct large, simple studies to investigate the precision and accuracy of the clinical examination for the diagnosis of COPD.¹ The first study completed by this group (investigating the accuracy of history, wheezing, laryngeal height, and laryngeal descent in 309 patients) has been reported elsewhere.⁴ In this paper, we describe a second study, which was designed to replicate and extend the findings of COAD-1 by evaluating the accuracy of self-reported history of COPD from the patient, smoking history (yes/no, number of pack years), wheezing on auscultation, and forced expiratory time (FET) in the diagnosis of COPD. This study represents an ongoing effort to whittle down the 32 signs and symptoms that have been described for the diagnosis of COPD to those that are most useful.

Received from The Division of General Internal Medicine, University Health Network-Mount Sinai Hospital, University of Toronto, Toronto, Canada (SES); The Division of General Internal Medicine, University of Alberta, Edmonton, Canada (FAM); The Trout Research and Education Centre at Irish Lake (DLS); and The ICRF/NHS Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, United Kingdom (JJD).

Address correspondence and requests for reprints to Dr Straus: Department of Medicine, Toronto General Hospital, 200 Elizabeth St., ENG-248 Toronto, ON M5G 2C4, Canada (e-mail: sstraus@mtsinai.on.ca).

METHODS

Investigators were recruited from various centers around the world via the Internet (www.carestudy.com) and the evidence-based health care e-mail discussion group. All investigators had to have access to the Internet and joined the study in groups of at least 2 (1 clinician and 1 spirometrist). Each group of investigators took responsibility for obtaining local ethics approval for the study. Investigator enrollment and data entry were done via a secure Internet-based data entry system (hosted in Canada); data collation and analysis were done at the Centre for Evidence-Based Medicine at the University of Oxford.

Consecutive patients at each site were enrolled in the study from February, 1999 to March, 1999 and included: adult patients (older than 50 years) who were known to have COPD, suspected of having COPD, or neither known nor suspected of having COPD. Known COPD was defined as: prior pulmonary function tests demonstrating forced expiratory volume at 1 second (FEV_1) below the fifth percentile and/or FEV_1 :forced vital capacity (FVC) ratio below the fifth percentile (or <0.7); patient self-report of a prior diagnosis of COPD, emphysema, or chronic bronchitis; or patient chronically taking inhaled bronchodilators and/or inhaled steroids. Suspected COPD was defined as: none of the criteria for known COPD fulfilled, but patient referred to participating clinician for suspected COPD, or participating clinician thought that COPD was a diagnostic possibility before the structured clinical exam. Patients with known or suspected COPD were eligible for enrollment during exacerbations of their disease provided that no bronchodilator treatment was given between the clinical exam and spirometry. Excluded were: patients with purely reversible airway obstruction (i.e., asthma); patients with a terminal illness whose goals of therapy were confined to comfort and dignity; patients under 50 years of age; patients with respiratory distress so severe that bronchodilators could not safely be withheld until after spirometry; patients who were medically unstable from other causes (e.g., acute myocardial infarction, drug overdose); patients who were enrolled in COAD-1⁴; and, patients who were unable to cooperate in either the clinical exam or spirometry (due to impaired cognition, level of consciousness, or language).

All patients underwent clinical examination and independent, blinded spirometry. The items chosen for the clinical examination were based on literature review and consensus among the investigators, with an emphasis on simplicity and brevity of exam. The items assessed included: self-reported history of COPD, smoking history, wheezing, and FET. To evaluate the presence or absence of wheezing, investigators listened for wheezes during exhalation over 4 standardized areas (bilateral upper and lower back).⁵ Forced expiratory time was evaluated by asking the patient to sit upright, to maximally inhale and then to exhale through an open mouth as forcefully as possible.

The clinician listened with a stethoscope positioned over the trachea in the suprasternal notch, and the duration of audible exhalation was timed to the nearest second.⁶⁻⁸ Standard instructions for offering encouragement to the patient during this process were provided. Practice runs were done until the patient learned the protocol, and subsequently the maneuver was repeated twice and the fastest result recorded. Each patient also underwent spirometry within 30 minutes of the clinical exam (without intercurrent bronchodilator use) to assess FEV_1 and FVC. A standard protocol for spirometry was used and the better of 2 attempts was recorded. The spirometrists and clinicians were blind to the results of the others' investigations.

Sensitivity, specificity, and likelihood ratios for each element of the clinical examination were calculated using spirometry as the gold standard (COPD was defined objectively as an FEV_1 below the fifth percentile as per the recommendations of the American Thoracic Society and the Rational Clinical Exam review on this topic^{6,9}). Percentile flow rates, adjusted for age and height, were calculated using the regression equation of Crapo et al.¹⁰ Continuous measurements (age and pack years of smoking) were categorized either according to cut points previously used in the literature or values based on receiver operating characteristic (ROC) curve analysis. The relationships between each diagnostic element and COPD were tested using χ^2 tests and Fisher's Exact test for dichotomous features, the χ^2 test for trend for categorical variables, and Student's *t* test for continuous variables.

Multivariate analyses were carried out using the method of Spiegelhalter and Knill-Jones (which adjusts for confounding from related diagnostic elements),^{11,12} and a reduced multivariate model was produced by grouping categories with similar likelihood ratios in each element and only selecting diagnostic elements with adjusted likelihood ratios greater than 2 or less than 0.5. All statistical analyses were done using the STATA statistical package (Release 5.0; Stata, College Station, Tex).

RESULTS

A total of 177 patients were recruited by 7 investigator groups from 6 countries. Of these, 16 patients had a primary diagnosis of asthma and were excluded from the analysis; no other protocol violations were identified. None of the patients enrolled in this study were in our previously reported study.⁴

Patient demographics are outlined in Table 1. Twelve percent of the patients were 75 years of age or older, and 56% of patients had an FEV_1 below the fifth percentile. The clinical exam took longer than spirometry (mean time 4.9 minutes, [95% confidence interval (95% CI), 4.4 to 5.4] vs 4.3 minutes [95% CI, 4.0 to 4.6]).

The accuracy of the various elements of the clinical exam assessed in this study is outlined in Table 2. While we used FEV_1 below the fifth percentile as the reference

Table 1. Patient Demographics

Variable	N = 161
Mean age (SD)	64.3 (9.0)
Female gender, n (%)	62 (39)
White, n (%)	120 (75)
Smoking status*	
Never smoked, n (%)	54 (34)
Smoked <20 pack years, n (%)	27 (17)
Smoked 20–40 pack years, n (%)	43 (27)
Smoked >40 pack years, n (%)	37 (23)
COAD status	
Known, n (%)	66 (41)
Suspected, n (%)	43 (27)
Neither known nor suspected, n (%)	52 (32)
Point of contact	
Primary, n (%)	74 (46)
Secondary/tertiary, n (%)	87 (54)
Mean forced expiratory time, s (SD)	5.7 (3.8)
Mean FEV ₁ , mL/s (SD)	1,720 (830)
Mean FVC, mL (SD)	2,520 (970)

FEV₁, forced/expiratory volume in 1 second; FVC, forced vital capacity.

standard, the likelihood ratios for each element of the clinical examination were also calculated using other reference standards (FEV₁:FVC <0.7) given the controversy over spirometric definitions for COPD.⁶ The accuracies reported in Table 2 did not change appreciably (data available from corresponding author on request).

As in our previous study, the ROC curve for smoking demonstrated that the most appropriate cut point for smoking was at 40 pack years (ROC curve available from the corresponding author). The ROC curve for FET demonstrated that cut points at 6 seconds and 9 seconds most accurately represented the data (ROC curve available from the corresponding author).

The reduced multivariate model included 3 items (Table 3). In patients with all 3 items present (self-reported history of COPD, wheezing audible on auscultation, and FET ≥9 seconds), the LR is 59. In patients without any of these items, the LR for having COPD is 0.3.

DISCUSSION

We evaluated the accuracy of several elements of the clinical examination in diagnosing COPD. On history the most useful items to assist in ruling in a diagnosis of COPD are self-reported history of COPD and smoking in excess of 40 pack years. On physical examination, an FET of >9 seconds and auscultated wheezing increases the probability of a diagnosis of COPD. None of these features on history or physical examination were sufficient individually to either rule in or rule out the diagnosis of COPD. For example, if a patient had a 10% pretest probability of having COPD (the prevalence of COPD among smokers⁷), an FET >9 seconds would only increase his post-test probability of COPD to 40%. Similarly, the presence of

Table 2. Accuracy of Elements of the Clinical Examination in Diagnosing COAD (Univariate Analysis)

Diagnostic Element	Sensitivity	Specificity	Crude Likelihood Ratio (95% CI)	P Value
COPD history*				<.001
Known			5.6 (2.9 to 10.9)	
Not known, but suspected			0.8 (0.5 to 1.3)	
Not known or suspected			0.2 (0.1 to 0.4)	
Smoking				.001
Smoked >1 y	0.74	0.52	1.5 (1.2 to 2.0)	
Never smoked, or smoked <1 y	0.26	0.48	0.5 (0.3 to 0.7)	
Amount smoked				
>40 Pack years			3.3 (1.5 to 7.1)	
20 to 40 Pack years			2.0 (1.0 to 4.0)	
<20 Pack years			0.7 (0.4 to 1.2)	
Age, y				.025
>65	0.54	0.66	1.6 (1.1 to 2.3)	
≤65	0.47	0.36	0.7 (0.5 to 0.9)	
Gender				<.001
Male	0.73	0.53	1.6 (1.2 to 2.1)	
Female	0.27	0.47	0.7 (0.3 to 0.7)	
Wheezing				<.001
Present	0.35	0.89	4.0 (1.6 to 9.9)	
Absent	0.65	0.11	0.8 (0.7 to 0.9)	
Forced expiratory time, s				<.001
>9			6.7 (2.1 to 21.1)	
6–9			1.8 (0.77 to 4.0)	
<6			0.6 (0.5 to 0.8)	

* Likelihood ratios and sensitivity/specificity for categorical variables do not correspond exactly since likelihood ratios are calculated using only the number of subjects within each category while sensitivity and specificity are calculated by collating all subjects above or below each cutpoint.

COAD, chronic obstructive airways disease; COPD, chronic obstructive pulmonary disease.

Table 3. Diagnostic Elements Included in the Multivariate Model for the Diagnosis of COPD

Diagnostic Element	Adjusted Likelihood Ratio	
	Factor Present	Factor Absent
Self-reported history of COPD	4.4	0.5
Wheezing	2.9	0.8
FET ≥ 9 s	4.6	0.8

COPD, chronic obstructive pulmonary disease; FET, forced expiratory time.

wheezing or a history of smoking >40 pack years in this same patient would only increase the post-test probability of COPD to approximately 30%.

Using multivariate analysis, we developed a 3-variable model for diagnosing COPD. The likelihood ratios for each of these variables can be multiplied together (as they are adjusted to account for their nonindependence) to generate a likelihood ratio for an individual patient.¹¹ For example, in a patient with self-reported history of COPD, audible wheezing on auscultation, and an FET of 9 seconds, the likelihood ratio is 59. If his pre-test probability was only 10%, his constellation of symptoms and signs increases his post-test probability to approximately 85%.

Our study adds substantially to the literature on the rigorous evaluation of the clinical examination for COPD in that it was a large study with a variety of clinicians involved. Furthermore, our findings are generally consistent with the literature and serve to refine estimates of the accuracy of the clinical examination for the diagnosis of COPD. For example, in a previous study⁴ we found that age, self-reported history of COPD, smoking more than 40 pack years, and maximum laryngeal height were useful in diagnosing COPD. In our earlier study, we also observed that the LR for wheezing was only 2.7 (consistent with the results of this study). In a study of 164 patients, Holleman and Simel² reported an LR of 12 for wheezing. However, the 95% CI on this estimate was quite large and ranged from 1.7 to 98. The results of our study are consistent with this estimate, and given our larger sample size, serve to refine the previously published estimates. Combining the results of 2 previous studies, Holleman and Simel found that an FET >9 seconds had a likelihood ratio of 4.9, which is consistent with the results of our study.⁶ Similarly, Schapira et al.⁸ reported LR (and 95% CI) of 0.4 (0.2 to 0.7) for FET <6 seconds, 2.2 (1.0 to 4.8) for FET from 6 to 8 seconds, and 4.1 (2.5 to 6.8) for FET >8 seconds in patients older than 60 years. The results from our study serve to confirm and refine these estimates.

The ability of our model to predict COPD is similar to previously published models (one incorporated years of smoking exposure, patient-reported wheezing and auscultated wheezing,² and the other incorporated pack year smoking history, self-reported history of COAD, and decreased breath sounds⁷); however, our model is better

at ruling out COPD than those other models. For example, in patients with no previous report of COPD, FET <9 seconds and no wheezing, the likelihood ratio (0.3) virtually rules out the diagnosis of COPD.

In our previous study,⁴ wheezing (adjusted LR [ALR] 1.5) did not enter the multivariate model, in contrast to the present study where it did contribute, although the adjusted LR (ALR 2.4) was not appreciably different. Also, a smoking history of more than 40 pack years (ALR 1.7) did not enter the model developed in this current study, although it did in the first study (ALR 6.4). A greater proportion of heavy smokers were included in this current study (50% of patients smoked more than 20 pack years vs 30% in our previous study), which may indicate that in settings where heavy smoking is more common, it is less useful in helping to rule in the diagnosis of COPD.

This study confirmed the feasibility of using the Internet to recruit investigators and conduct studies of the clinical examination. It has resulted in the development of a practice-based network of clinicians interested in performing other studies of the clinical examination.¹³

Although some may argue that it is tautologous to include "self-reported history of COPD" in a prediction rule for COPD, we believe it is important to include this item in the prediction rule for 2 reasons. First, this replicates clinical practice: clinicians take a history before proceeding to physical examination and laboratory tests and routinely incorporate patient symptoms or self-reported history when making diagnoses or decisions about further testing. Second, given the possible variation with which disease labels are applied to patients, we felt it important to determine whether a prior label of COPD added to diagnostic accuracy or muddied the waters. Data from patients with a self-reported history of COPD were excluded and the analysis was repeated. No additional items entered into the model, although the overall accuracy of the model was poorer (AUC = 0.70).

However, there were some limitations to our study. First, as with our first study, we did not assess inter-rater reliability. This was a deliberate exclusion since the primary focus of this study was to evaluate the accuracy of these elements of the clinical examination and to prove the feasibility of the study design. In order to achieve these objectives, the study was designed such that data collection would be brief. We decided a priori to defer assessment of interobserver variation for a future study in which we will only evaluate those signs that have been shown to be accurate. Second, in order to participate, investigators had to have access to the Internet. Critics might be concerned that this may affect the applicability of our results to patients or clinicians in other settings. However, our patients and results are similar to those in other studies,^{2,4,6-8} suggesting that our findings are generalizable. Third, the applicability of our model for clinical practice has yet to be determined, although the preliminary observations from our dataset suggest it holds significant promise. Finally, while this study serves to

validate a number of the clinical examination elements we found to be accurate in predicting COPD in our previous study,⁴ a complete model incorporating the elements found to be accurate in this study and those from our previous study⁴ must be tested in an independent sample of patients. Such a study is being planned.

In summary, our results suggest that less emphasis should be placed on the individual findings of the presence of wheezing or prolonged FET in making a diagnosis of COPD. We found that a combination of 3 symptoms/signs (self-reported history of COPD, presence of wheezing on auscultation, and FET ≥ 9 seconds) can be used to predict airway obstruction. In those settings where spirometry is readily available, it should be used because it takes less time to do than the clinical examination, definitively establishes the diagnosis of airway obstruction, and provides prognostic information. However, in those settings where spirometry is unavailable, our model provides useful diagnostic support for the clinician. Future studies are under way to evaluate other signs and symptoms that have been described for the diagnosis of COPD and to test our model in independent samples of patients.

The COAD2 Investigators, their affiliations, and the number of patients entered by each investigator are: C. Baicus, M. Georgescu, E. Parvanescu, N. Gh, Lupu Hospital, Bucharest, Romania (60); M. Kamei, N. Yamada, Shinsakae-machi, Naka-ku, Nagoya, Japan (45); P. Sestini, R. Refini, R. Amerini, M. Vichi, Institute of Respiratory Diseases, University of Siena, Italy (20); K. Sharma, C.M. Parker, Ottawa, Canada (15); M. Urtasun, M. Molinari, Hospital Regional Ushuaia, Ushuaia, Argentina (15); D. Ross, C. Nurse, Chorley, United Kingdom (12); S. Straus, F. McAlister, D. Sackett, John Radcliffe Hospital, Oxford, United Kingdom (10).

The authors thank Dr. Malcolm Benson for his assistance with patient recruitment and Mr. Richard Straus for his technical assistance and support.

SES, DLS, and JJD were supported by the NHS Research and Development Programme, United Kingdom; SES is supported by a Career Scientist Award from the Ontario Ministry

of Health and by the Knowledge Translation Program at the University of Toronto; FAM is a Population Health Investigator of the Alberta Heritage Foundation for Medical Research and was supported by the Medical Research Council of Canada.

REFERENCES

1. McAlister FA, Straus SE, Sackett DL, on behalf of the CARE-COAD1 Investigators. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet*. 1999;354:1853-7.
2. Holleman DR, Simel DL, Goldberg JS. Diagnosis of obstructive airways disease from the clinical examination. *J Gen Intern Med*. 1993;8:63-8.
3. Holleman DR Jr., Simel DL. Does the clinical examination predict airflow limitation. *JAMA*. 1995;273:313-9.
4. Straus SE, McAlister FA, Sackett DL, Deeks J, on behalf of the CARE-COAD1 Investigators. The accuracy of the patient history, wheezing, and laryngeal measurements in diagnosing obstructive airways disease. *JAMA*. 2000;283:1853-7.
5. Sapira JD. *The Art and Science of Bedside Diagnosis*. Munich: Urban and Schwarzenberg; 1990.
6. Holleman DR, Simel DL. Does the clinical examination predict airflow limitation. *JAMA*. 1995;273:313-9.
7. Badgett RG, Tanaka DJ, Hunt DK, et al. Can moderate chronic obstructive pulmonary disease be diagnosed by historical and physical findings alone. *Am J Med*. 1993;94:188-96.
8. Schapira RM, Schapira MM, Funahashi A, McAuliffe TL, Varkey B. The value of the forced expiratory time in the physical diagnosis of obstructive airways disease. *JAMA*. 1993;270:731-6.
9. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis*. 1991;144:1202-18.
10. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis*. 1981;23:659-64.
11. Feinstein AR. Clinical biostatistics XXXIX. The haze of Bayes, the aerial palaces of decision analysis, and the computerized Ouija board. *Clin Pharmacol Ther*. 1977;21:482-96.
12. Spiegelhalter DJ, Knill-Jones RP. Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology. *J R Statist Soc*. 1984;147:35-77.
13. Straus SE, McAlister FA, Sackett DL, on behalf of the CARE group. Clinical assessment of the reliability of the examination (CARE). *ACP J C*. 2000;133:A11-12.