

# Individualized prediction of lung-function decline in chronic obstructive pulmonary disease

Zafar Zafari MSc, Don D. Sin MD MPH, Dirkje S. Postma MD, Claes-Göran Löfdahl MD, Judith Vonk PhD, Stirling Bryan PhD, Stephen Lam MD, C. Martin Tammemagi PhD, Rahman Khakban MSc, S.F. Paul Man MD, Donald Tashkin MD, Robert A. Wise MD, John E. Connett PhD, Bruce McManus MD PhD, Raymond Ng PhD, Zsuzsanna Hollander PhD, Mohsen Sadatsafavi MD PhD

See also [www.cmaj.ca/lookup/doi/10.1503/cmaj.160611](http://www.cmaj.ca/lookup/doi/10.1503/cmaj.160611)

**Competing interests:** See end of article.

This article has been peer reviewed.

**Accepted:** May 2, 2016

**Online:** Aug. 2, 2016

**Correspondence to:** Don Sin, Don.Sin@hli.ubc.ca

*CMAJ* 2016. DOI:10.1503/cmaj.151483

## ABSTRACT

**Background:** The rate of lung-function decline in chronic obstructive pulmonary disease (COPD) varies substantially among individuals. We sought to develop and validate an individualized prediction model for forced expiratory volume at 1 second (FEV<sub>1</sub>) in current smokers with mild-to-moderate COPD.

**Methods:** Using data from a large long-term clinical trial (the Lung Health Study), we derived mixed-effects regression models to predict future FEV<sub>1</sub> values over 11 years according to clinical traits. We modelled heterogeneity by allowing regression coefficients to vary across individuals. Two independent cohorts with COPD were used for validating the equations.

**Results:** We used data from 5594 patients (mean age 48.4 yr, 63% men, mean baseline FEV<sub>1</sub> 2.75 L) to create the individualized prediction equations. There was significant between-

individual variability in the rate of FEV<sub>1</sub> decline, with the interval for the annual rate of decline that contained 95% of individuals being -124 to -15 mL/yr for smokers and -83 to 15 mL/yr for sustained quitters. Clinical variables in the final model explained 88% of variation around follow-up FEV<sub>1</sub>. The C statistic for predicting severity grades was 0.90. Prediction equations performed robustly in the 2 external data sets.

**Interpretation:** A substantial part of individual variation in FEV<sub>1</sub> decline can be explained by easily measured clinical variables. The model developed in this work can be used for prediction of future lung health in patients with mild-to-moderate COPD. **Trial registration:** Lung Health Study — ClinicalTrials.gov, no. NCT00000568; Pan-Canadian Early Detection of Lung Cancer Study — ClinicalTrials.gov, no. NCT00751660

Chronic obstructive pulmonary disease (COPD) is a global health burden that affects 300 million people worldwide<sup>1</sup> resulting in more than 3 million deaths annually.<sup>1,2</sup> Although COPD is defined by airflow limitation, the rate of decline in lung function is extremely variable across patients.<sup>3,4</sup> Accordingly, data on the rate of change of forced expiratory volume in 1 second (FEV<sub>1</sub>), the most commonly used measure of lung function, can be very noisy (i.e., quite variable), often associated with a coefficient of variation that exceeds 1.50.<sup>3</sup> FEV<sub>1</sub> is directly related to severity classifications (i.e., Global Initiative for Chronic Obstructive Lung Disease [GOLD] grades) that determine treatment algorithms.<sup>1</sup> However, the relatively poor signal-to-noise ratio of this measurement has made it difficult to risk-stratify patients for disease progression, especially in mild COPD, where the difference between

patients with rapid and slow decline might be difficult to detect. Such risk stratification can help physicians personalize strategies for disease management and help researchers design more efficient therapeutic trials. For cardiovascular diseases, prediction tools (e.g., the Framingham Risk Score<sup>5</sup>) have been available for several decades and have played major roles in clinical, research and policy domains. A lack of equivalent risk-prediction tools and the reduced ability to individualize disease prevention and management might explain the lack of success in reducing the burden of COPD compared with cardiovascular diseases.<sup>6</sup>

The objective of this study was to create and externally validate a probabilistic model to predict the individualized rate of decline in FEV<sub>1</sub> over 11 years and the corresponding GOLD severity grades in current smokers with mild-to-moderate COPD.

## Methods

### Study population

To derive the prediction equations, we used data from the Lung Health Study (LHS). The details of the LHS design and its major findings have been published elsewhere.<sup>7</sup> In summary, LHS was a multicentre clinical trial, in which 5887 smokers were randomly assigned to 3 arms of usual care and special intervention (smoking cessation) with or without a bronchodilator (ipratropium). All patients had mild-to-moderate COPD and were between the ages of 35 and 60 years.<sup>7-9</sup> Patients were excluded if they had any other substantial respiratory diseases.<sup>7,8</sup> All patients were seen in person on an annual basis for 5 years, and spirometry was performed according to criteria of the American Thoracic Society.<sup>7</sup> The study was subsequently extended by the addition of an in-person visit at about the 11th year of follow-up.<sup>10</sup> In the current study, we included all individuals with no missing FEV<sub>1</sub> values and other independent variables at baseline and with at least 1 follow-up FEV<sub>1</sub>.

### Exposure and outcomes

The exposures (predictors) were clinical and demographic variables (e.g., sex, age, weight and height), treatment group assignment, methacholine responsiveness (by use of the O'Connor 2-point slope<sup>11</sup>), smoking history (in pack-years) and baseline FEV<sub>1</sub>. In line with the original analysis of LHS, individuals were considered continuous smokers if they smoked throughout the first 5 years of follow-up, sustained quitters if they did not smoke in this period and intermittent quitters if their smoking behaviour varied.<sup>7</sup> The outcome of interest was the individualized FEV<sub>1</sub> value after bronchodilator treatment for up to 11 years after the baseline visit.

### Validation cohort

We determined the external validity of the prediction equations using 2 independent data sets: the European Respiratory Society Study on Chronic Obstructive Pulmonary Disease (EUROSCOP) and the Pan-Canadian Early Detection of Lung Cancer Study (PanCan). EUROSCOP was a multicentre clinical trial that compared inhaled budesonide with placebo over 3 years in patients with mild-to-moderate COPD, with recruitment and follow-up between 1992 and 1996,<sup>12</sup> and with spirometry performed every 3 months. PanCan<sup>13</sup> recruited current or former smokers with or without COPD and performed spirometry yearly for up to 3 years, with recruitment and follow-up between 2008 and 2013.<sup>14</sup> Because different treatments were used in these

trials, external validation was determined only in the placebo arms of these studies. In addition, for external validation in PanCan, we included only patients with FEV<sub>1</sub> values between 55% and 90% of the predicted value to be in line with inclusion criteria of LHS and EUROSCOP.<sup>7-9,12</sup>

### Statistical analysis

The details of the statistical analysis are provided in Appendix 1, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1). Briefly, we used mixed linear regression to model the FEV<sub>1</sub> for each individual. A mixed-effects model enables explicit specification of heterogeneity by assigning random-effect terms to parameters whose effect is variable between individuals. The regression equation was of the form

$$FEV_{1t} = \beta_0 + \beta.X + \beta_0'.t + \beta'.X.t + \beta_0''.t^2 + (int_{smo} + int_{ipra}) + e$$

with  $FEV_{1t}$  representing FEV<sub>1</sub> at  $t^{\text{th}}$  follow-up year and  $e$  representing an independent normally distributed error term.  $X$ 's are the set of covariates (i.e., baseline age, sex, weight, height, height squared, smoking status, O'Connor slope, and interaction of baseline age and height squared) as described above. Intercept,  $\beta_0$ , and slope,  $\beta_0'$ , were modelled as random effect (to vary across individuals), and other coefficients were modelled as fixed effect.  $\beta$  coefficients predict the baseline FEV<sub>1</sub> and  $\beta'$  coefficients predict the slope of FEV<sub>1</sub> change over time,  $\beta_0''$  captures the potential nonlinear component of decline, and  $int_{smo}$  and  $int_{ipra}$  represent smoking-cessation and ipratropium interventions that model 1-time jump in FEV<sub>1</sub> after the baseline visit for those who quit smoking or received ipratropium (these 2 variables were set to 0 for all 3 arms for the baseline visit because baseline FEV<sub>1</sub> was measured before the initiation of interventions).<sup>8</sup> This model connects serial FEV<sub>1</sub> measurements for an individual through a multivariate normal distribution, enabling conditional prediction of future FEV<sub>1</sub> values based on observable characteristics, baseline FEV<sub>1</sub> and, potentially, previous FEV<sub>1</sub> values (details of calculations are provided in Appendix 2, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1)).

Using this framework, we constructed different models with different choices of predictors. We used the Akaike information criterion<sup>15</sup> to choose the best predictive model (the final model). Details of model selection are shown in Appendix 1. From the final model, we calculated the predicted individualized lung function, as well as the range of FEV<sub>1</sub> values around the esti-

mate at the individual level that covers 95% of individuals with similar characteristics (i.e., 95% prediction interval). We also predicted the probabilities of different GOLD grades over 11 years for an individual based on clinical traits and smoking status (continued smoker v. sustained quitter) during the follow-up period. In addition, we predicted future FEV<sub>1</sub> and GOLD grades by adding a 1-year prior FEV<sub>1</sub> value to other baseline clinical traits (analogous to knowing a previous history of exacerbation, which can enhance prediction of future COPD exacerbations). A variance component analysis was performed to determine the contribution of covariates in explaining the variation of the follow-up FEV<sub>1</sub> values. Finally, we evaluated the discriminatory power of the model in predicting future GOLD grades by calculating the C statistics. GOLD grading classifies lung-function decline into 4 categories: mild (FEV<sub>1</sub> ≥ 80% of predicted), moderate (FEV<sub>1</sub> 50%–79% of predicted), severe (30%–49% of predicted) and very severe (< 30% of predicted).<sup>1</sup> We used reference equations from the Third National Health and Nutrition Examination Survey to calculate the predicted FEV<sub>1</sub> values,<sup>16</sup> and we combined severe and very severe grades because not many predicted values fell into the very severe category.

We performed internal validation using LHS and external validation using EUROSCOP<sup>12</sup> and PanCan.<sup>13</sup> Because EUROSCOP and PanCan did

not measure bronchial responsiveness (i.e., O'Connor slope), we refitted the final model after removing this variable (Appendix 3, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-/DC1)). Validity was assessed in 3 ways: plotting the observed versus predicted mean FEV<sub>1</sub> at follow-up visits, calculating the root mean squared error (RMSE) of the predicted versus observed FEV<sub>1</sub> values (the smaller the RMSE, the better the prediction), and determining the coverage probability defined as the proportion of the observed FEV<sub>1</sub> values falling within the 95% prediction interval for that observation (the closer the coverage probability to 95%, the better the prediction).

To make the prediction equation accessible to the research and clinical communities, we developed a Web application. All analyses were performed using SAS Version 9.4.

## Results

We used data from 5594 individuals (mean age 48.4 yr; 63% men). The mean follow-up time was 9.2 years with a combined total of 35 046 visits. Details of the baseline characteristics can be found in Table 1.

The results of model comparison are provided in Appendix 1. The final model included the following variables: baseline age, follow-up time, sex, weight, height, height squared, smoking status during follow-up, the O'Connor slope, and smoking-cessation and ipratropium interventions. Regression coefficients from this model are presented in Table 2. Most of the included variables in the final model were significantly associated with the rate of FEV<sub>1</sub> decline ( $p < 0.05$ ). There was significant between-individual variability in the rate of FEV<sub>1</sub> decline, with the interval for the annual rate of decline that contained 95% of individuals being –124 to –15 mL/yr for smokers and –83 to 15 mL/yr for sustained quitters. In the final model, 88% of total variation around the follow-up FEV<sub>1</sub> values was explained by the included clinical covariates and baseline FEV<sub>1</sub> as well as random effects. The final model had a C statistic of 0.90 for follow-up GOLD grades. Within follow-up periods, the C statistics were 0.92 for years 1 and 2, 0.91 for year 3, 0.90 for year 4, 0.88 for year 5 and 0.85 for year 11.

Figure 1 shows an example of prediction of lung-function decline and the corresponding GOLD grades stratified by future smoking behaviour for a patient with a baseline FEV<sub>1</sub> of 2.75 L. Based on these figures, if the patient continues smoking, over 11 years the average

**Table 1:** Baseline characteristics of 5594 patients included in the final model

Characteristic	Mean ± SD*
Follow-up time, yr	9.2 ± 2.9
Baseline age, yr	48.4 ± 6.8
Baseline FEV <sub>1</sub> , Lt	2.75 ± 0.63
FEV <sub>1</sub> , % of the predicted value	78.47 ± 9.06
Weight, kg	75.9 ± 15.1
Height, m	1.72 ± 0.09
Methacholine responsiveness (O'Connor slope)‡	–12.73 (23.4)
Pack-years of smoking	40 ± 19
<b>Sex, no. (%)</b>	
Male	3524 (63.0)
Female	2070 (37.0)
<b>Smoking status by year 5, no. (%)</b>	
Sustained quitters	951 (17.0)
Intermittent quitters	1566 (28.0)
Continuous smokers	3077 (55.0)
Note: FEV <sub>1</sub> = forced expiratory volume in 1 second, SD = standard deviation.	
*Unless stated otherwise.	
†After bronchodilator treatment.	
‡Unit is change in FEV <sub>1</sub> per mg/mL change in methacholine concentration.	

rate of decline would be  $-70$  (95% prediction interval  $-128$  to  $-11$ ) mL/yr. If the patient stops smoking, the expected rate of decline would be  $-40$  (95% prediction interval  $-98$  to  $18$ ) mL/yr. In terms of GOLD grades, if the patient continues to smoke, there is a 9% chance that he will remain in mild COPD, an 84% chance that he will transition to moderate COPD and a 7% chance that he will transition to severe or very severe COPD. These transitions can be substantially improved if the patient quits smoking, with almost no chance of severe or very severe COPD developing, a 63% chance of moderate COPD developing and a 37% chance that he will remain in mild COPD. Incorporation of a previous FEV<sub>1</sub> value for this patient will reduce the width of the prediction interval. This effect is more pronounced in short-term prediction: the width of the prediction interval is reduced by 38% and 30% for the first- and second-year prediction, respectively.

## Validation

### Internal validation using LHS

Figure 2A–B presents the predicted-versus-observed mean FEV<sub>1</sub> at follow-up visits for continuous smokers and sustained quitters. The RMSE for both smokers and sustained quitters was 0.24, and the actual coverage probabilities of the 95%

prediction intervals were 94% and 93%, respectively (Appendix 4, e-Figure 1, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.151483/-DC1), provides further internal validation for other subgroups).

### External validation using EUROSCOP

There were 542 patients (72% men, baseline age 52.5 yr, baseline FEV<sub>1</sub> 2.56 L [73.23% of predicted]) in the placebo arm of EUROSCOP. Baseline characteristics of this cohort are summarized in the Appendix 4, e-Table 3. Figure 2C shows the replication data in this cohort. The RMSE in this cohort was 0.22, and the actual coverage probability of 95% prediction interval was 91% (Appendix 4, e-Figure 2 provides results for subgroups within EUROSCOP).

### External validation using PanCan

There were 940 patients with COPD in PanCan (59% men, baseline age 63 yr, baseline FEV<sub>1</sub> 2.21 L). Baseline characteristics of this cohort are presented in Appendix 4, e-Table 3. Figure 2D–E presents predicted-versus-observed values for FEV<sub>1</sub> decline for current smokers and former smokers, respectively. For current smokers, RMSE and the actual coverage probability of 95% prediction interval were 0.25 and 90%, respectively; whereas for sustained quitters, these values were 0.19 and 93%, respectively.

**Table 2:** Association of covariates with baseline FEV<sub>1</sub> and rate of FEV<sub>1</sub> change (millilitres per year)

Variable	Baseline FEV <sub>1</sub>		Rate of FEV <sub>1</sub> change	
	Effect (95% CI), mL	<i>p</i> value	Effect (95% CI), mL/yr	<i>p</i> value
Intercept	1421.2 (–1277.33 to 4119.73)	0.3	–177.9 (–456.42 to 100.62)	0.2
Baseline age, yr	–5.19 (–17.2 to 6.82)	0.4	2.31 (1.06 to 3.56)	< 0.001
Sex (male v. female)	462.5 (436.71 to 488.29)	< 0.001	–8.86 (–11.55 to –6.17)	< 0.001
Weight, kg	–0.11 (–0.86 to 0.64)	0.8	0.15 (0.07 to 0.23)	< 0.001
Height, m	–1760.3 (–4729.11 to 1208.51)	0.2	74.13 (–232.61 to 380.87)	0.6
Height squared, m <sup>2</sup>	1893.1 (1037.36 to 2748.84)	< 0.001	11.39 (–77.2 to 99.98)	0.8
Continuous smoker* (v. sustained quitters)	–77.22 (–88.73 to –65.71)	< 0.001	–25.79 (–28.29 to –23.29)	< 0.001
Intermittent quitter* (v. sustained quitters)	–41.31 (–53.92 to –28.7)	< 0.001	–10.02 (–12.79 to –7.25)	< 0.001
Methacholine responsiveness (O'Connor slope)†	2.61 (2.24 to 2.98)	< 0.001	0.2 (0.16 to 0.23)	< 0.001
Baseline age × height squared‡	–8.2 (–12.2 to –4.2)	< 0.001	–0.92 (–1.34 to –0.5)	< 0.001
Follow-up time from baseline, yr	—	—	–0.44 (–0.59 to –0.29)	< 0.001
<b>Parameters</b>	Effect on 1-time jump in FEV <sub>1</sub>			
Intervention: smoking cessation	27.35 (18.78 to 35.92)	< 0.001		
Intervention: ipratropium	33.71 (24.05 to 43.37)	< 0.001		

Note: CI = confidence interval, FEV<sub>1</sub> = forced expiratory volume in 1 second.  
 \*Dummy variable.  
 †Measure of hyperresponsiveness. This variable is log-transformed. Unit is per-unit log change in FEV<sub>1</sub>, per mg/mL change in methacholine concentration.  
 ‡This interaction term was chosen among different interaction terms based on their Akaike information criterion values.

### Web application

Using prediction equations, we developed a Web-based application (available at <http://resp.med.ubc.ca/software/ipress/epic/fev1pred>). This tool enables the prediction of future FEV<sub>1</sub> values and GOLD grades for up to 11 years using clinical variables that can be collected at the point of care. It also allows users to incorporate, from external sources, the effect of pharmacologic interventions such as bronchodilators in terms of 1-time increase in FEV<sub>1</sub>.

### Interpretation

Using data from LHS, we developed equations that enable individualized probabilistic prediction of FEV<sub>1</sub> decline for up to 11 years based on readily

available clinical features at the point of care in patients with mild-to-moderate COPD. Our data are consistent with the original data by Fletcher and Peto, which showed that continuous smokers on average experienced a faster decline than nonsmokers,<sup>17</sup> and with subsequent studies, which showed the tremendous heterogeneity in COPD.<sup>17,18</sup> We also validated the robustness of our equations in 2 independent data sets, EUROSCOP and PanCan. The latter study is more recent than LHS and EUROSCOP, which assures the relevance of our prediction for modern patients with COPD.

The present study can be seen as a step toward creating a quantitative framework for outcome predictions in COPD. The present work was focused on FEV<sub>1</sub> and did not incorporate other meaningful

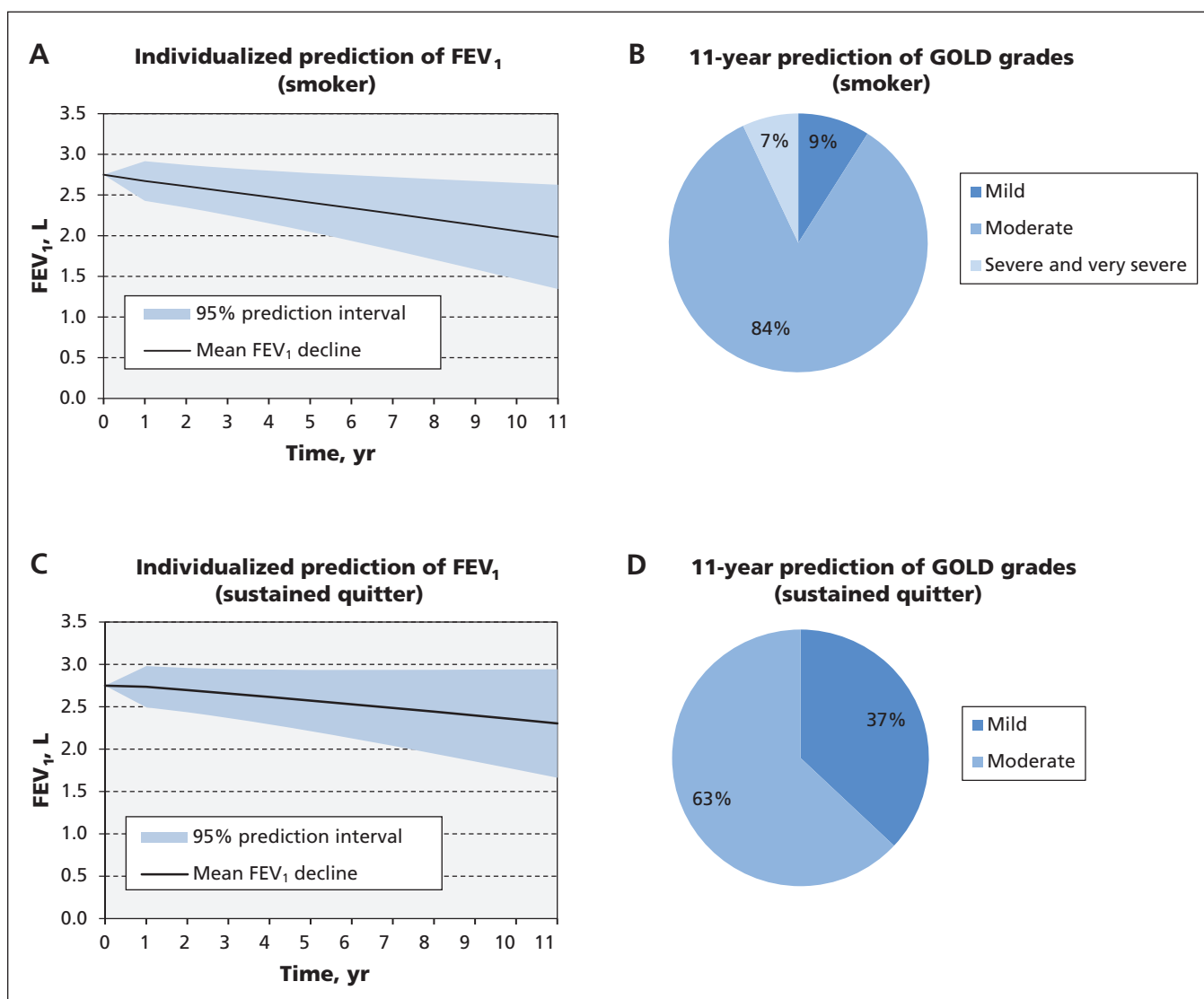


Figure 1: Prediction results based on baseline FEV<sub>1</sub> and clinical traits for an example patient (a 55-year-old man who is a continuous smoker, weight 75 kg, height 170 cm, baseline FEV<sub>1</sub> 2.75 L). A) Mean estimates and 95% prediction intervals for future FEV<sub>1</sub>, and B) 11-year prediction of GOLD grades if the patient continues smoking. C) Mean estimates and 95% prediction intervals for future FEV<sub>1</sub>, and D) 11-year prediction of GOLD grades if the patient stops smoking. This is an illustrative case only. The reader can use the online FEV<sub>1</sub> calculator (<http://resp.med.ubc.ca/software/ipress/epic/fev1pred>) to estimate future FEV<sub>1</sub> decline in patients with different clinical features. Note: COPD = chronic obstructive pulmonary disease, FEV<sub>1</sub> = forced expiratory volume in 1 second.

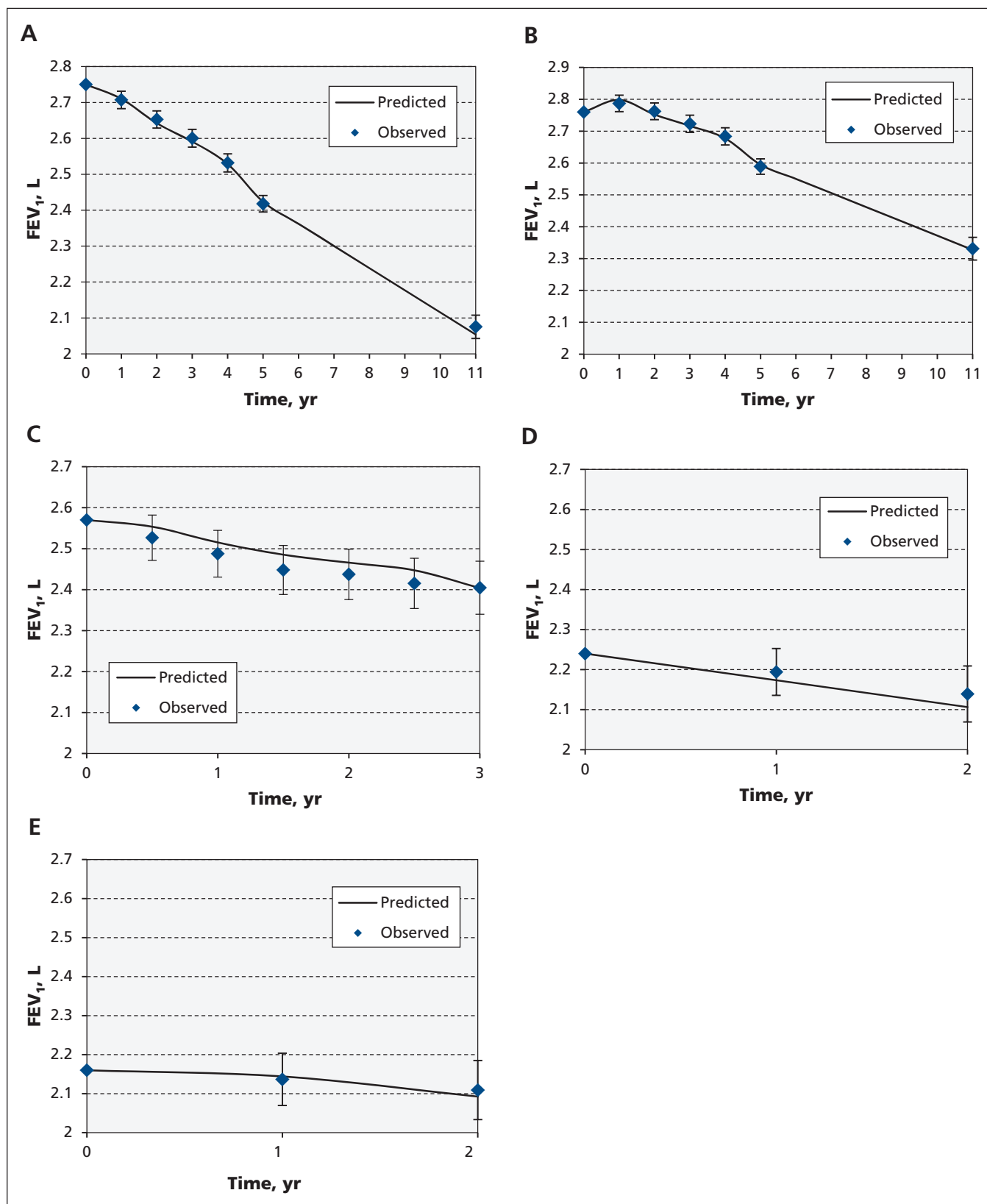


Figure 2: Internal validation of the model in (A) LHS smokers (RMSE 0.24, actual 95% coverage probability 94%) and (B) LHS sustained quitters (RMSE 0.24, actual 95% coverage probability 93%). External validation of the model in (C) EUROSCOP smokers (EUROSCOP included only smokers) (RMSE 0.22, actual 95% coverage probability 91%), (D) PanCan smokers (RMSE 0.25, actual 95% coverage probability 90%) and (E) PanCan sustained quitters (RMSE 0.19, actual 95% coverage probability 93%). Note: EUROSCOP = European Respiratory Society Study on Chronic Obstructive Pulmonary Disease, FEV<sub>1</sub> = forced expiratory volume in 1 second, LHS = Lung Health Study, PanCan = Pan-Canadian Early Detection of Lung Cancer Study, RMSE = root mean squared error.

end points in COPD. However, it has the potential to be expanded, incorporating exacerbations and mortality as watershed COPD events that are affected by the degree of lung-function impairment. For other conditions, such as cardiovascular diseases, such frameworks have been in place for decades, allowing for evidence-based decision-making at the clinical and policy levels, as well as more informed design of clinical trials.<sup>19</sup> Given the high and escalating burden of COPD, it is time to develop similar frameworks for this disease.

### Limitations

There were some limitations to our study. First, LHS did not include thoracic computed tomography of patients;<sup>20</sup> thus the impact of emphysema on the rate of FEV<sub>1</sub> decline could not be incorporated into our model. Second, the determinants of FEV<sub>1</sub> are likely to be very complex with multiple interactions, and, although we examined the performance of several models, they are inevitably simplified versions of the underlying disease process. Third, our model is applicable in patients with mild-to-moderate COPD, which has the greatest opportunity for disease modification. Our model may not be generalizable to patients with more severe disease. Moreover, our equations may not be generalizable to individuals with asthma-COPD overlap syndrome, patients with COPD who are lifetime nonsmokers, or patients whose predominant risk factor is biomass or other forms of indoor or outdoor pollution.

### Conclusion

This framework will allow clinicians to risk-stratify patients with mild-to-moderate COPD in terms of their future lung-function decline and to identify patients with rapid disease progression, who can be targeted for close follow-up and intervention (e.g., smoking cessation programs). Another potential application of such a prediction tool is to promote the design of efficient clinical trials of interventions to modify disease progression by improving the signal-to-noise ratio of the FEV<sub>1</sub> decline variable and reducing the sample size. The latter is achieved in 2 ways: by providing estimates of residual variance for sample size calculation that remove the effect of heterogeneity due to observable characteristics, and by enriching the recruitment by patients who are most likely to experience rapid decline in lung function. In addition, the development of the Web-based tool can enable rapid translation of the study's findings into clinical practice and research designs.

### References

1. *Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: clinical practice guidelines. Revised, 2014.* Global Initiative for Chronic Obstructive Lung Disease; 2014.

2. World Health Report. Geneva: World Health Organization; 2000. Available: [www.who.int/whr/2000/en/whr00\\_en.pdf](http://www.who.int/whr/2000/en/whr00_en.pdf) (accessed 2016 July 22).
3. Vestbo J, Edwards LD, Scanlon PD, et al.: ECLIPSE Investigators. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med* 2011;365:1184-92.
4. Lange P, Celli B, Agustí A, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015;373:111-22.
5. Anderson KM, Odell PM, Wilson PWF, et al. Cardiovascular disease risk profiles. *Am Heart J* 1991;121:293-8.
6. Ford ES. Hospital discharges, readmissions, and ED visits for chronic obstructive pulmonary disease or bronchiectasis among US adults: findings from the Nationwide Inpatient Sample 2001–2012 and Nationwide Emergency Department Sample 2006–2011. *Chest* 2015;147:989-98.
7. Connett JE, Kusek JW, Bailey WC, et al. Design of the Lung Health Study: a randomized clinical trial of early intervention for chronic obstructive pulmonary disease. *Control Clin Trials* 1993;14(Suppl 2):3S-19S.
8. Anthonisen NR, Connett JE, Kiley JP, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV<sub>1</sub>. The Lung Health Study. *JAMA* 1994;272:1497-505.
9. Anthonisen NR. Lung Health Study. *Am Rev Respir Dis* 1989;140:871-2.
10. Man SFP, Connett JE, Anthonisen NR, et al. C-reactive protein and mortality in mild to moderate chronic obstructive pulmonary disease. *Thorax* 2006;61:849-53.
11. O'Connor G, Sparrow D, Taylor D, et al. Analysis of dose-response curves to methacholine. An approach suitable for population studies. *Am Rev Respir Dis* 1987;136:1412-7.
12. Pauwels RA, Löfdahl CG, Laitinen LA, et al. Long-term treatment with inhaled budesonide in persons with mild chronic obstructive pulmonary disease who continue smoking. European Respiratory Society Study on Chronic Obstructive Pulmonary Disease. *N Engl J Med* 1999;340:1948-53.
13. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910-9.
14. Leung JM, Mayo J, Tan W, et al.; Pan-Canadian Early Lung Cancer Study Group. Plasma pro-surfactant protein B and lung function decline in smokers. *Eur Respir J* 2015;45:1037-45.
15. Myung JI, Cavagnaro DR, Pitt MA. Model evaluation and selection. In: Batchelder WH, Colonius H, Ossietzky CV, et al., editors. *New handbook of mathematical psychology. Vol. 1: Foundations and methodology.* Cambridge (UK): Cambridge University Press. In press.
16. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999;159:179-87.
17. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977;1:1645-8.
18. Kohansal R, Martínez-Cambor P, Agustí A, et al. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* 2009;180:3-10.
19. Sheridan S, Pignone M, Mulrow C. Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians. *J Gen Intern Med* 2003;18:1039-52.
20. Galbán CJ, Han MK, Boes JL, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711-5.

**Competing interests:** Don Sin reports grants and personal fees from AstraZeneca, and personal fees from Boehringer Ingelheim and Almirall. Dirkje Postma reports grants from AstraZeneca, Chiesi, Genentech, GlaxoSmithKline and Roche, and consultant fees from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Takeda and Teva. Robert Wise reports grants, personal fees and nonfinancial support from GlaxoSmithKline; grants and personal fees from AstraZeneca, Boehringer Ingelheim and Teva; personal fees from Novartis, Mylan, Sunovion, Takeda, Roche, Merck, Pfizer, Pulmonx, Spiration, Vertex, Verona, Bristol-Myers Squibb, Janssen, Theravance, Sarepta, Grifols, Sunovion and ContraFect. Mohsen Sadatsafavi reports grants from AstraZeneca Canada. No other competing interests were declared.

**Affiliations:** Centre for Clinical Epidemiology and Evaluation (Zafari, Bryan, Sadatsafavi), Vancouver Coastal Health

Research Institute, University of British Columbia; Centre for Heart and Lung Innovation (Sin, Man, McManus, Hollander), St. Paul's Hospital; Institute for Heart and Lung Health (Sin, Man, McManus, Hollander, Sadatsafavi), University of British Columbia; Division of Respiratory Medicine (Sin, Lam, Man, Sadatsafavi), Department of Medicine, University of British Columbia, Vancouver, BC; University Medical Center Groningen (Postma, Vonk), Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, Groningen; University Medical Center Groningen (Postma), Department of Pulmonary Diseases, University of Groningen, Groningen, the Netherlands; Department of Respiratory Medicine and Allergology (Löfdahl), Lund University, Lund, Sweden; Department of Epidemiology (Vonk), University of Groningen, University Medical Center Groningen, the Netherlands; School of Population and Public Health (Bryan), University of British Columbia, Vancouver, BC; Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences (Khakban), University of British Columbia, Vancouver, BC; David Geffen School of Medicine at UCLA (Tashkin), Los Angeles, Calif.; Johns Hopkins University School of Medicine (Wise), Baltimore, Md.; University of Minnesota School of Public Health (Connett), Minneapolis, Minn.; PROOF Centre for Excellence (McManus, Ng, Hollander), Vancouver, BC; Department of Community Health Sciences (Tammemagi), Brock University, St. Catharines, Ont.

**Contributors:** Zafar Zafari, Mohsen Sadatsafavi and Don Sin conceived the study idea. Zafar Zafari and Mohsen Sadatsafavi designed the experiments. Zafar Zafari was the main statistical analyst and developed an online calculator for the individualized prediction of lung-function decline. Don Sin provided clinical support and insights, and helped with data provision. Mohsen Sadatsafavi supervised data analysis and reporting. Stirling Bryan provided methodologic insight and helped with interpretation

of the results. Zsuzsanna Hollander helped with data provision. Rahman Khakban provided methodologic insights and helped with interpretation of the results. Donald Tashkin, Robert Wise, John Connett, Stephen Lam, C. Martin Tammemagi, S.F. Paul Man, Dirkje Postma, Claes-Göran Löfdahl and Judith Vonk collected the data and provided intellectual input to the study. Bruce McManus and Raymond Ng obtained funding for the study and provided intellectual input to the study. Zafar Zafari drafted the manuscript, which all of the authors revised. All of the authors gave final approval of the version to be published and agreed to act as guarantors of the work.

**Funding:** This study was funded by Genome Canada, Genome British Columbia, Génome Québec, Canadian Institutes of Health Research, the Providence Health Care Research Institute, St. Paul's Hospital Foundation, the PROOF Centre for Excellence and the Canadian Respiratory Research Network (grant nos. 144COP and IEF132213). Zafar Zafari received the Four Year Doctoral Fellowship award from the University of British Columbia. Mohsen Sadatsafavi receives salary support from the National Sanatorium Association. Don Sin is a Tier 1 Canada Research Chair in COPD. The European Respiratory Society Study on Chronic Obstructive Pulmonary Disease (EUROSCOP) study was supported by AstraZeneca, and the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) was supported by the Terry Fox Research Institute and the Canadian Partnership Against Cancer. The funders had no role in design, execution or reporting of the study.

**Acknowledgement:** The authors acknowledge the contributions of the investigators of the Lung Health Study (LHS), European Respiratory Society Study on Chronic Obstructive Pulmonary Disease (EUROSCOP) and Pan-Canadian Early Detection of Lung Cancer Study (PanCan). An earlier version of this work was presented in American Thoracic Society 2015 Conference.