# COPD: still an unpredictable journey

Alex Bottle[1] and Jenni Quint[2]

**Affiliations**: [1]Dept of Primary Care and Public Health, Imperial College London, London, UK. [2]National Heart and Lung Institute, Imperial College London, London, UK.

**Correspondence**: Alex Bottle, Dr Foster Unit, Imperial College London, 3 Dorset Rise, London EC4Y 8EN, UK. E-mail robert.bottle@imperial.ac.uk

@ERSpublications
**New study of hospital-recruited COPD patients finds that three biomarkers explain 13% of the total explained 8-year mortality, but still long way to go for risk prediction** https://bit.ly/3h4jKyV

**Cite this article as:** Bottle A, Quint J. COPD: still an unpredictable journey. *Eur Respir J* 2021; 57: 2002933 [https://doi.org/10.1183/13993003.02933-2020].

## Introduction

COPD is the fourth leading cause of death and affects nearly 400 million worldwide, causing much disability as well as early mortality [1]. Management should be tailored to each patient to improve outcomes, for instance by stratifying patients according to the risk of acute exacerbations of COPD, in order to prescribe treatments including inhaled corticosteroids or phosphodiesterase-4 inhibitors earlier. Treatment is particularly important before the patient's first acute exacerbation, as each subsequent exacerbation damages the lungs and treatment is less effective thereafter [2–4]. Prior acute COPD exacerbation is well known to be a strong predictor of later events [5]. In this issue of the *European Respiratory Journal*, Celli *et al.* [6] aim to tackle the more general problem of the lack of disease activity measures, such as biomarkers. The concept of "active" disease in COPD has been around for a while, but it is not known whether this should be defined by frequent acute exacerbations, inflammation by one or other biomarker, or "fast" decline of forced expiratory volume in 1 s ($FEV_1$). Rather than answer this directly, Celli *et al.* [6] explored the relationship between disease severity measures "as surrogate markers of disease activity" (and changes in those measures over time) and mortality. We now discuss some key issues regarding the study's dataset, analysis, contribution to collective knowledge, and implications for clinical practice and research.

## Dataset

ECLIPSE was a multicentre observational study in which hospitalised COPD patients with Global Initiative for Chronic Obstructive Lung Disease grades II–IV, plus smoking and nonsmoking controls, were evaluated regularly for 3 years. It was set up to identify clinically relevant COPD subtypes and find predictors, including biomarkers, of disease progression [7]. While $FEV_1$ is considered the standard measure of this, its limitations include the fact that people with the same $FEV_1$ can have different functional status and that it can take quite a long time for decline to be usable as a measure of progression: a point to which we will return.

The 89% of ECLIPSE enrolees with COPD who were alive after 3 years were eligible for 8-year follow-up, of whom 80% consented. While the supplementary analysis gives some reassurance that the final sample reflected the wider initial cohort, all the participants were recruited from hospital, with 20% having had frequent acute COPD exacerbation at baseline, and presumably many more with at least one (figures were not given). It would have been useful to know more about the predictors in people who had not yet had

Received: 28 July 2020 | Accepted: 11 Aug 2020

an exacerbation. As Celli *et al.* [6] acknowledge, a community-based cohort would be needed for this and would have improved the generalisability.

## Ranking predictors in importance

With any risk model, a key question is: how much better is the new set of predictors than the existing ones? Celli *et al.* [6] compared a baseline model containing the BODE index, age and prior COPD hospitalisation with a model containing three systemic biomarkers in addition, using the adequacy index to show that the three biomarkers explain 13% of the total mortality explained by the model: not a huge amount, but a useful addition nonetheless. Many studies that have a binary outcome such as death mistakenly use the c statistic (area under the receiver operating characteristic curve) for this, leaving authors disappointed in the often tiny rise. This reflects a limitation of the c statistic, not the model, and it should not be used to compare and choose between nested models [8]. Other studies use p-values to rank predictors, which is also a poor choice, as p-values are heavily dependent on the prevalence of the risk factor in the dataset.

Another way to think about predictor importance is patient *versus* system. For an individual patient with a given risk factor, *e.g.* COPD hospitalisation, the hazard ratio (or other similar regression coefficient) is important. For instance, for frequent exacerbations in year 1, the hazard ratio was 1.30 (95% CI 1.09–1.54). At the system or population level, however, we need to consider not just the hazard ratio but the prevalence. This brings in established epidemiological measures, such as the population attributable risk or fraction (PAF; terminology varies). This estimates the proportion of outcomes accounted for by one or more given risk factors if a causal relation can be assumed (a big "if"). This is how we know that smoking is the main but not sole important explanation of COPD cases, with suspicion falling on outdoor pollution, occupational exposure, *etc*., for the remainder [9]. For mortality in cohort studies such as this one, hazard ratios can be used to estimate the strength of the association. The PAF tells us what proportion of mortality can be delayed during the given follow-up time; confidence intervals are available (see Laaksonen *et al.* [10] for a summary). For frequent exacerbations in year 1, the hazard ratio of 1.30 and a baseline prevalence of 20.1% overall (a weighted average for survivors and deceased from table 2), the PAF would be approximately 6%. A similar calculation for frequent exacerbations in all 3 years of the main study, with an overall prevalence of only 11.7% but a higher hazard ratio of 1.63, yields a PAF of approximately 7%. These figures suggest that reducing the frequency of exacerbations is important (not surprisingly), but that delaying mortality requires much else besides. PAFs have their limitations, but they offer a vital population or system perspective.

## Implications for research and practice

The authors make three main claims to novelty and importance. The first is that $FEV_1$ decline takes between 1 and 3 years to affect mortality, though they note that other studies have used a window of between 2 and 6 years. However, the ECLIPSE protocol [7] itself noted that, "a substantial length of observation time is required for using $FEV_1$ decline as a measure of progression, and biomarkers that could evaluate progression over a shorter period of time would be useful." The other two claims relate to the lack of association between most of their variables, including the cross-sectional measurements made at baseline, and 8-year mortality. Given the heterogeneity of COPD and, as the authors acknowledge, of the measurements in their model, perhaps we should not be too surprised by this.

There are other statistical options for this kind of dataset beyond what was performed in this study. One is the inclusion of time-varying covariates rather than the use of changes from baseline. Others concern phenotyping, which in COPD has continued to develop since the concept was included in the 1995 American Thoracic Society guidelines with the classic three subgroups of chronic bronchitis, emphysema, and asthma (see Corlateanu *et al.* [11] for a review of COPD phenotyping). Cluster analysis is common, but group-based trajectory modelling could also be used for risk stratification, as we have done for admissions during a 5-year period in patients with heart failure [12], a condition with several elements in common with COPD.

How might these results or any subsequent risk score be used in clinical practice? There is a brief mention of electronic health records (EHRs) in the paper, when it is suggested that information such as $FEV_1$ decline and previous hospitalisation could be readily retrieved to aid the identification of high-risk patients and the tailoring of treatment. An albeit still limited number of risk prediction algorithms that rely on EHRs are in use in clinical practice for other conditions, particularly coronary heart disease [13]. For COPD, the study's model could readily incorporate comorbidities, for example, and we have found clinic appointment non-attendance to predict mortality, which is captured in electronic administrative data in England [14]. Much has been written about the potential for machine learning methods to improve prediction, whilst noting some major challenges such as data biases and inequalities. Ultimately, for a risk

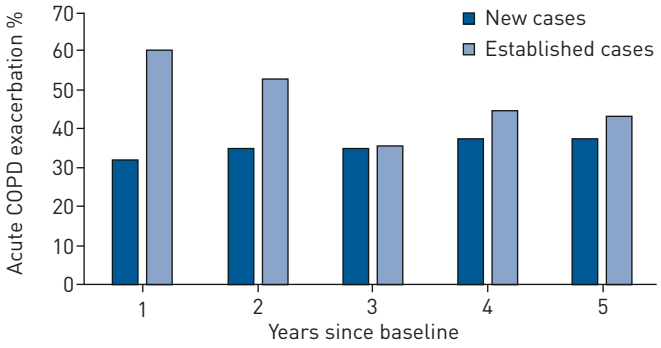


FIGURE 1 Percentage of newly diagnosed and established COPD cases having an acute COPD exacerbation by year of follow-up in a UK population-based linked database.

model to be useful, it must pass many tests, not merely those of statistical performance. It needs to be based on information available to the clinician making decisions with the patient, which EHRs can help with, but at a time when the clinician and/or the patient can take risk-mitigating action. When the $FEV_1$ has declined appreciably from multiple exacerbations, it is too late. Acute exacerbations of COPD are common in both incident and established cases, with around a third having one in the first year after diagnosis (figure 1, with data from our study using the Clinical Practice Research Datalink, a representative linked primary care-based UK database [15]).

If people with (or indeed without) COPD were predictable, data analysis would be a very simple task. This study reminds us that, although our understanding of mortality in COPD has advanced, we still have a long way to go.

Conflict of interest: A. Bottle reports grants from Dr Foster and Medtronic, outside the submitted work. J. Quint reports personal fees for consultancy from GSK, AZ, Insmed and Vertex, grants from MRC, the Health Foundation, AZ, GSK, Bayer, Asthma UK, Chiesi and BI, outside the submitted work.

## References

1 GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med* 2020; 8: 585–596.
2 Seemungal TA, Donaldson GC, Paul EA, *et al.* Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 157: 1418–1422.
3 Soler-Cataluña JJ, Martínez-García MA, Sánchez PR, *et al.* Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005; 60: 925–931.
4 Hurst JR, Vestbo J, Anzueto A, *et al.* Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; 363: 1128–1138.
5 Guerra B, Gaveikaite V, Bianchi C, *et al.* Prediction models for exacerbations in patients with COPD. *Eur Respir Rev* 2017; 26: 160061.
6 Celli B, Locantore N, Yates JC, *et al.* Markers of disease activity in COPD: an 8-year mortality study in the ECLIPSE cohort. *Eur Respir J* 2021; 57: 2001339.
7 Vestbo J, Anderson W, Coxson HO, *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* 2008; 31: 869–873.
8 Cook N. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928–935.
9 Eisner MD, Anthonisen N, Coultas D, *et al.* An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 182: 693–718.
10 Laaksonen MA, Knekt P, Härkänen T, *et al.* Estimation of the population attributable fraction for mortality in a cohort study using a piecewise constant hazards model. *Am J Epidemiol* 2010; 171: 837–847.
11 Corlateanu A, Mendez Y, Wang Y, *et al.* Chronic obstructive pulmonary disease and phenotypes: a state-of-the-art. *Pulmonology* 2020; 26: 95–100.
12 Rao A, Kim D, Darzi A, *et al.* Long-term trends of use of health service among heart failure patients. *Eur Heart J Qual Care Clin Outcomes* 2081; 4: 220–231.
13 Hippisley-Cox J, Coupland C, Brindle O. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; 357: j2099.
14 Bottle A, Honeyford K, Chowdhury F, *et al.* Factors associated with hospital emergency readmission and mortality rates in patients with heart failure or chronic obstructive pulmonary disease: a national observational study. Southampton, NIHR Journals Library, 2018.
15 Rothnie KJ, Mullerova H, Smeeth L, *et al.* Natural history of chronic obstructive pulmonary disease exacerbations in a general practice-based population with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 198: 464–471.