ELSEVIER
Journal of Cystic Fibrosis xx (2017) xxx – xxx
Journal of Cystic Fibrosis

Original Article

# A 3-year prognostic score for adults with cystic fibrosis

L. Nkam [a,*], J. Lambert [b], A. Latouche [c], G. Bellis [d], PR. Burgel [e,f,1], M.N. Hocine [a,1]

[a] Laboratoire Modélisation, Epidémiologie et Surveillance des Risques Sanitaires, Conservatoire National des Arts et Métiers, Paris, France
[b] Hôpital Saint-Louis, Service de Biostatistique et Information Médicale, Assistance Publique-Hôpitaux de Paris, Paris, France
[c] Centre d'Etudes et de Recherche en Informatique et Communications, Conservatoire National des Arts et Métiers, Paris, France
[d] Institut National d'Etudes Démographiques, Paris, France
[e] Assistance Publique-Hôpitaux de Paris, Hôpital Cochin, Service de Pneumologie, Paris, France
[f] Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Received 20 September 2016; revised 2 March 2017; accepted 2 March 2017
Available online xxxx

**Abstract**

*Background:* Therapeutic progress in patients with cystic fibrosis (CF) has resulted in improved prognosis over the past decades. We aim to reevaluate prognostic factors of CF and provide a prognostic score to predict the risk of death or lung transplantation (LT) within a 3-year period in adult patients.
*Methods:* We developed a logistic model using data from the French CF Registry and combined the coefficients into a prognostic score. The discriminative abilities of the model and the prognostic score were assessed by c-statistic. The prognostic score was validated using a 10-fold cross-validation.
*Results:* The risk of death or LT within 3 years was related to eight characteristics. The development and the validation provided excellent results for the prognostic score; the c-statistic was 0.91 and 0.90 respectively.
*Conclusion:* The score developed to predict 3-year death or LT in adults with CF might be useful for clinicians to identify patients requiring specialized evaluation for LT.


*Keywords:* Cystic fibrosis; Prognostic factors; Registry data; Logistic model; Prognostic score

## 1. Introduction

Cystic fibrosis (CF) is a multiorgan disease that affects primarily the lungs, causing diffuse bronchiectasis which often leads to progressive respiratory insufficiency and premature death [9]. Lung transplantation (LT) is proposed to CF patients with terminal respiratory failure with the aim of improving life expectancy and quality of life [25]. Although criteria for referring patients for LT have been proposed [14], the optimal timing for referring CF patients for transplantation remains difficult to establish in an individual patient. A recent study in France has shown that respiratory death in CF patients often occurs due to late or no referral for LT [19], suggesting the need to develop novel strategies for referring patients at high risk of death for transplant evaluation.

In the past 25 years, several statistical models have been developed to identify prognostic factors in CF patients. In their seminal study, Kerem et al. identified forced expiratory volume in 1 s ($FEV_1$) as the main prognostic factor and suggested that patients with an $FEV_1$ value less than 30% should be considered for LT [16]. Subsequent studies identified multiple other factors related to death in patients with CF including older age [17,21],

* Corresponding author at: Laboratory 'Modélisation, Epidémiologie et Surveillance des Risques Sanitaires', Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France.
*E-mail address:* lionelle.nkam@cnam.fr (L. Nkam).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jcf.2017.03.004

female gender [16,17], lower body mass index (BMI) [21], pancreatic insufficiency [1,17], diabetes [1,17], *Pseudomonas aeruginosa* colonization [1,6], *Burkholderia cepacia* colonization [4,6,17], *Staphylococcus aureus* colonization [7], massive hemoptysis [11], pneumothorax [10] and high number of pulmonary exacerbations [4,17,21]. Although several attempts at developing prognostic scores have been performed in CF patients [1,4,13,17,20,21], it has proven difficult to develop a score that better predicts death than $FEV_1 < 30\%$ predicted [20]. Further, prognosis has dramatically improved over the past decades due to advance in integrated care provided by multidisciplinary teams in CF centers [2,8,12,18,23]. As a result, prognostic factors have changed over time and studies performed using data obtained in previous decades may not be appropriate for current evaluation of CF patients. For example, George et al. showed an important improvement in the survival of patients whose $FEV_1$ has fallen below 30% of predicted value. Consequently, they suggested that the threshold of 30% predicted for $FEV_1$ should no longer be considered in isolation as an indication for LT [12]. Moreover, pediatric mortality in patients with CF has almost disappeared in developed countries due to improvement in patients care by multidisciplinary teams [26].

The aim of the present study was to develop a 3-year predictive model that provides a prognostic score to better predict the risk of death or LT in adult patients with CF.

## 2. Materials and methods

### 2.1. The French Cystic Fibrosis Registry

We used data from the French Cystic Fibrosis Registry (French CF Registry). This registry contains longitudinal data on more than 8000 patients since 1992, which represents approximately 90% of all CF patients in France [3]. Each patient is assigned to a center specialized in CF, where his/her health status is regularly monitored. A numeric code is assigned to each patient to link information between specialized centers and the French CF Registry. This registry records annual health-check data for each subject including vital status, therapeutic management, anthropometry, spirometry, morbidity factors, consultations and hospitalizations, arterial blood gas, microbiological tests, pregnancy and paternity, and transplantations and sociodemographic data [3].

### 2.2. Patients and data collection

The period of the study was 2010–2013. Patients alive and aged 18 years or older on 31st December 2010 and for whom vital status was known on 31st December 2013 were included in the study. Patients who received a lung transplant before 2010 and patients lost to follow-up between 2011 and 2013 were excluded from the study. Forty-two covariates (listed in Table 1 and Supplementary Table 1) considered as potential predictors and records of the year 2010 were extracted to predict the outcome, defined as death or LT before the end of 2013. Fig. 1 presents the selection scheme of patients who were included in the study.

### 2.3. Missing data imputation

In 2010, only 12% of patients had complete information for all the 42 potential predictors. However, the percentage of missing data represented only 4%, as illustrated in Table 1. To deal with missing data in the covariates, a multiple imputation by chained equations was used [27]. We assumed that data were missing at random that is, the probability of missingness depends on the values of the observed covariates.

### 2.4. Model development

The characteristics of patients according to the outcome were compared using chi-square test or Fisher's exact test for categorical variables, and the Mann–Whitney test for continuous variables. We developed a multivariable logistic regression model to predict the outcome of interest, defined as death or LT by the end of 2013. Covariates that were significantly associated to the outcome in 2013 (p value $<0.25$) with univariate analysis were considered for the multivariable logistic regression model. A forward stepwise selection process was used to select the subset of variables independently associated with the outcome. The predictors retained in the final model were combined into a prognostic score to easily estimate the individual risk of death or LT within 3 years. To this end, continuous predictors were transformed into categorical variables according to clinically relevant thresholds. The contribution of each predictor to the prognostic score was proportional to its regression coefficient. To help the clinician to easily obtain the prognostic score and the risk of death or LT in a 3-year period using the patient characteristics, a nomogram was provided.

### 2.5. Model performances

Performances of the developed model and the prognostic score were investigated in terms of discrimination and calibration. Discrimination assesses how well the model can distinguish patients with the outcome of interest and patients without. This was evaluated using the c-statistic, also known as the area under the receiver operating characteristic curve [5]. The calibration compares the observed proportion of events against the predicted probabilities. It was evaluated using the Hosmer-Lemeshow test [15]. These performances were tested for both the developed model and the prognostic score, on each imputed dataset.

### 2.6. Model internal validation

To avoid overestimation of the model performances, we performed an internal validation using a 10-fold cross-validation. Overestimation happens when the model performs well on the data used for development but not on test data. Cross-validation can help detect overestimate models and helps to assess how well the model fits new observations. We randomly partitioned the initial dataset into 10 subsamples, fitted the model on nine of the subsamples and evaluated its performances on the other. We repeated this ten times, leaving out each subsample once. The performance of the prognostic score was evaluated using the

Please cite this article as: Nkam L, et al, A 3-year prognostic score for adults with cystic fibrosis, J Cyst Fibros (2017), http://dx.doi.org/10.1016/j.jcf.2017.03.004