ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A statistical model to predict one-year risk of death in patients with cystic fibrosis

Shawn D. Aaron[a,*], Anne L. Stephenson[b], Donald W. Cameron[a], George A. Whitmore[a,c]

[a]*Division Head of Respiratory Medicine, Ottawa Hospital Research Institute, University of Ottawa, The Ottawa Hospital, 501 Smyth Road, Ottawa, Ontario, K1H 8L6 Canada*

[b]*Department of Medicine, St Michael's Hospital, 30 Bond Street, 6th floor, Bond Wing, Toronto, Ontario, M5B 1W8, Canada*

[c]*McGill University, Faculty of Management, 845 Sherbrooke Street West, Montreal, Quebec H3A 0G4, Canada*

Accepted 23 December 2014; Published online xxxx

Abstract

Objectives: We constructed a statistical model to assess the risk of death for cystic fibrosis (CF) patients between scheduled annual clinical visits. Our model includes a CF health index that shows the influence of risk factors on CF chronic health and on the severity and frequency of CF exacerbations.

Study Design and Setting: Our study used Canadian CF registry data for 3,794 CF patients born after 1970. Data up to 2010 were analyzed, yielding 44,390 annual visit records. Our stochastic process model postulates that CF health between annual clinical visits is a superposition of chronic disease progression and an exacerbation shock stream. Death occurs when an exacerbation carries CF health across a critical threshold. The data constitute censored survival data, and hence, threshold regression was used to connect CF death to study covariates. Maximum likelihood estimates were used to determine which clinical covariates were included within the regression functions for both CF chronic health and CF exacerbations.

Results: Lung function, *Pseudomonas aeruginosa* infection, CF-related diabetes, weight deficiency, pancreatic insufficiency, and the deltaF508 homozygous mutation were significantly associated with CF chronic health status. Lung function, age, gender, age at CF diagnosis, *P aeruginosa* infection, body mass index <18.5, number of previous hospitalizations for CF exacerbations in the preceding year, and decline in forced expiratory volume in 1 second in the preceding year were significantly associated with CF exacerbations. When combined in one summative model, the regression functions for CF chronic health and CF exacerbation risk provided a simple clinical scoring tool for assessing 1-year risk of death for an individual CF patient. Goodness-of-fit tests of the model showed very encouraging results. We confirmed predictive validity of the model by comparing actual and estimated deaths in repeated hold-out samples from the data set and showed excellent agreement between estimated and actual mortality.

Conclusion: Our threshold regression model incorporates a composite CF chronic health status index and an exacerbation risk index to produce an accurate clinical scoring tool for prediction of 1-year survival of CF patients. Our tool can be used by clinicians to decide on optimal timing for lung transplant referral.

Keywords: Cystic fibrosis; Health index; Mortality; Registry data; Risk scoring tool; Threshold regression

1. Introduction

Understanding the natural history of cystic fibrosis (CF) requires knowledge of the statistical trajectory of lung function and health status in the disease. The statistical properties of the trajectory are the basis for judging whether CF treatments and clinical management are having significant effects on the time course of the disease and for assessing patient prognosis. Perhaps most importantly, understanding of a CF patient's health status and health trajectory is critical toward deciding when an individual should be referred for lung transplantation. Lung transplantation is a complex rescue therapy used in CF patients with terminal lung disease with frequently dramatic results. However, lung transplant is not a cure for CF lung disease; the median survival for adults after lung transplantation is only 6.4 years [1]. Therefore, choosing to list a patient for a lung transplant requires timing the transplant procedure to maximize survival benefits. Survival benefit can be maximized by restricting lung transplantation to patients who otherwise would

Conflict of interest: Research salary support for two authors (S.D.A. and D.W.C.) was provided by the Department of Medicine of the University of Ottawa, which also assisted with research expenses through a Development Research Award (S.D.A. and G.A.W.).

* Corresponding author. Tel.: 613-739-6636; fax: 613-739-6807.
E-mail address: saaron@ohri.ca (S.D. Aaron).

http://dx.doi.org/10.1016/j.jclinepi.2014.12.010
0895-4356/

What is new?

- We constructed a stochastic model based on threshold regression techniques to estimate risk of mortality in patients with cystic fibrosis.
- The model includes measurements of chronic disease progression and an exacerbation shock stream to model the overall course of CF lung disease.
- The model incorporates a scoring risk tool to predict one-year risk of death for patients with cystic fibrosis.
- After further validation studies, the model will be produced as a mobile application to predict one-year risk of death, and will be available to CF clinicians so that they can decide on optimal timing for lung transplant referral for individual patients.

Key findings

- The research builds a new statistical model for assessing 1-year survival of cystic fibrosis (CF) patients using data on annual clinical visits from the Canadian CF registry. The stochastic process model consists of an exacerbation shock stream superimposed on chronic disease progression. Death occurs when an exacerbation first takes CF health across a critical threshold. The formulation is a special implementation of threshold regression. The model produces a scoring tool for risk of CF death that can be used as a mobile application by CF clinicians. As a research by-product, the analysis constructs a CF health index that shows the influence of different risk factors on CF chronic health and on the severity and frequency of CF exacerbations.

have a high risk of mortality within the ensuing year. Conversely, transplant listing needs to occur early enough for the patient to survive the length of time needed to wait for organs to become available [2].

In efforts to optimize timing for lung transplant, investigators have attempted to design models to predict survival probability for individuals with CF. Liou et al. [3] used data from the Cystic Fibrosis Foundation Patient Registry and multivariate logistic regression methods to create a 5-year survivorship model. They suggested that patients with a 5-year predicted survival less than 30% be considered for transplantation, whereas patients with a predicted survival greater than 50% be excluded. Mayer-Hamblett et al. [4] identified similar risk factors to those of Liou et al. as predictors of 2-year mortality. However, they found that their model was no better than using ''forced expiratory volume in 1 second (FEV1) less than 30%'' criteria as a predictor, and both methods had poor positive predictive value. The authors concluded that ''transplant referral decisions based either on a multivariate logistic model or on the criterion of an FEV1 of less than 30% predicted are likely to result in high rates of premature referral. Better clinical predictors of short-term mortality among patients with CF are needed.'' To further complicate matters, a recent cohort study suggests a marked improvement has occurred in the median survival of those patients with cystic fibrosis with an FEV(1) less than 30% predicted, from 1.2 years in 1990 to 5.3 years in 2002. These evolving changes in survival in at-risk patients make transplant referral decisions even more difficult to operationalize for the CF clinician [5].

The Canadian CF Patient Data Registry (CPDR) is a comprehensive high-quality database that covers Canadian CF patients from the 1960s to the present. Our research uses CPDR data to develop a statistical model of CF health status for individual patients. Our focus in this report is on the experience of patients between annual visits recorded in the registry because it is between these visits that the major events of death or transplant occur. Our models are estimated from longitudinal data on lung function and other variables in the CPDR, together with transplant and mortality data. These records give useful but limited aggregate information about events occurring between annual visits (such as the number of hospital admissions) so our statistical methodology must consider the masking of the continuous disease process produced by irregularly spaced observation points.

FEV1, expressed as a percent of the normal value (FEV1% predicted), is the primary measure of lung function used clinically to monitor CF disease. A principal objective of this research was to create a composite index of respiratory measures including FEV1% predicted, comorbidity measures, and other risk factors that correlate well with CF mortality. We refer to this index as the CF health index. Our index can serve as a clinical scoring tool for CF mortality risk for individual patients, using only data from their most recent clinical visits. In our research, the individual time course of lung function is viewed as a stochastic process that has three components: (1) chronic disease progression, (2) exacerbations, and (3) minor transient elements produced by measurement variability and short-lived daily fluctuations in physiology. Chronic disease progression is postulated to be a Markov process in which lung function undergoes persistent physiological changes over time having stationary independent increments. The exacerbation component captures intermittent departures from the path of natural disease progression caused by infections or other aggravating agents, which vary in degree and duration. Exacerbation onsets are followed by recovery intervals in which time, natural healing processes, and clinical interventions bring the trajectory back to the natural path of disease progression. Finally, the transient component is modeled as a random disturbance term.