Development and External Validation of 1- and 2- year Mortality Prediction Models in Cystic Fibrosis

Sanja Stanojevic[1,2], Jenna Sykes[3], Anne L. Stephenson[2,4], Shawn D. Aaron[5], G. A. Whitmore[6]

SUPPLEMENT

Study Population

Patient data are entered into the CCFR once a diagnosis of CF has been confirmed based on diagnostic guidelines (1) and after patient consent has been obtained. All 42 Canadian CF clinics submit data annually to the Registry and each CF clinic receives financial support from Cystic Fibrosis Canada dependent on data submission to the Registry. It is estimated that less than 1% of the Canadian CF population has declined consent to have their data captured in the Registry (personal communication with CF Canada). The frequency of CF in Canada as captured by national statistics was virtually the same as found in the Registry suggesting that the Registry captured most CF patients in the country. (12)

Description of Clinical Variables

Date of CF diagnosis, if missing, was imputed by assuming the date of diagnosis occurred 30 days after the date of birth.(2) Pancreatic status was defined by whether the patient had ever used pancreatic enzymes and was used as a proxy for functional impairment of the CFTR protein since many older patients were missing genotype classification. Lung function measurements FEV_1 and forced vital capacity (FVC) were converted to percent predicted values using the Global Lung Function Initiative (GLI) reference equations.(3) Extreme values of FEV_1 or FVC percent predicted that were clinically implausible (<8% or >150% predicted) were set to missing. The change in lung function was calculated as the log-relative decline in lung function since the preceding year if a decline has occurred and 0 otherwise. The log-relative decline is the natural log of the ratio of the current FEV_1 percent predicted to the previous measurement. If the preceding year's lung function was missing, the most recent lung function (up to a maximum of three years earlier) was used. Nutritional status was categorized as underweight (Body Mass

Index (BMI) < 18 kg/m^2), overweight (BMI ≥ 25 kg/m^2) and normal weight (BMI ≥ 18 and BMI < 25 kg/m^2) according to World Health Organization (WHO) cut-offs for adults (4). For children, Centers for Disease Control and Prevention growth charts were used, as follows: underweight (BMI centile <12), overweight (BMI centile ≥85), normal weight (BMI centile ≥12 & BMI centile <85).(5) CF-related diabetes (CFRD) was defined by each clinic based on published consensus guidelines.(6) History of microbiological infections was categorized using a hierarchical approach: patients with a history of (1) *Burkholderia cepacia* complex, (2) *Pseudomonas aeruginosa*, and (3) neither *B. cepacia* complex *nor P. aeruginosa* infection. Once a patient was infected with *B. cepacia* complex the patient was assumed to be positive from that point forward, whereas infection with *P. aeruginosa* was categorized as positive/negative based on the culture results from the report year. Hospitalizations treated with intravenous (IV) antibiotics and home IV antibiotic courses were treated as separate variables. **The vast majority of IV courses were administered in hospital, and it is likely that courses administered at home represent 'milder' events. Due to the retrospective nature of the analysis it was not possible to clearly distinguish these events. There also may be some overlap of these variables.** The number of outpatient clinic visits per patient was recorded annually.

Table S1. Model Coefficients for the 1-year and 2-year models.

Model	Variable	One-year Coefficients	Two-year Coefficients
Chronic Health Index	Intercept	5.702963	4.55962
	Male	-0.0162938	.3189947
	Log_e(FVC % Predicted/100)	0.7360137	.5809873
	Log_e(FEV_1 % Predicted/100)	0.7899955	.8404154
	Underweight	-0.7302478	-.4187824
	B. cepacia complex	-0.4588687	-.9285728
	Age (in years)	-0.0398486	N/A
	# Hospitalizations in preceding year	-0.2818584	N/A
Shock Index	Intercept	0.1146547	.1863934
	# Hospitalizations in preceding year	-0.0792965	-.1263516
	1 year decline in lung function	-0.5616525	.1858131
	Log_e(FEV_1 % Predicted/100)	0.2554754	.4353779
	Pancreatic Insufficient	0.6058589	.1927758
	CFRD	0.2340407	-.172767
	Age at CF diagnosis (in years)	0.0079757	.0012487

How to calculate the probability of survival at 1-year and 2-years.

Step 1: Sum the intercept and the products of the coefficients and values of the variables for each patient for the chronic health index. This sum is denoted by *lnY*.

Step 2: Sum the intercept and the products of the coefficients and values of the variables for each patient for the shock index. This sum is denoted by *lnβ.*

Step 3: Exponentiate both *lnY* and *lnβ*, i.e. Y = exp(*lnY*) and β=exp(*lnβ*).

Step 4: $lnS = \frac{1}{(-1)*\beta}\left(\frac{1}{Y}\right)^{\beta}[\exp(\beta) - 1]$

Step 5: Calculate S_1 = exp(*lnS*). This is the probability of survival at one year.

Step 6: To calculate the probability of survival at two years, we repeat Steps 1-5 to calculate S_2 using the two-year coefficients. Then, the overall probability of survival at two years is S_1*S_2.

Example Calculations:

Table S2. Baseline values for two patients – one at low risk of death in one year, and one at high risk of death in one year.

Model	Variable	Low Risk Patient	High Risk Patient
Chronic Health Index	Age (years)	16	40.4
	Gender	Female	Male
	FEV1 % Predicted	47.4%	19.2%
	FVC % Predicted	66.7%	25.7%
	# Hospitalizations in preceding year	0	6
	Underweight	No	Yes
	B. cepacia Complex	No	Yes
Shock Index	Age at CF Diagnosis	0.9 yrs	27.2 yrs
	CFRD	No	No
	Pancreatic Status	Insufficient	Sufficient
	FEV1 % Predicted	47.4%	19.2%
	FEV1 % Predicted in preceding year	80.5%	20.0%
	1-year decline in FEV1	33.1%	0.8%
	# Hospitalizations in preceding year	0	6
Outcome	Status in one year	Alive	Deceased
	Status at two years	Alive	N/A
One-year survival	ln Y	4.18	-1.11
	Y	65.37	0.33
	ln β	0.24	-0.59
	β	1.27	0.55
	ln S	-0.01	-2.45
	S	0.990	0.086

Two-year survival	ln Y	3.70	N/A
	Y	40.45	
	ln β	0.15	
	β	1.16	
	ln S	-0.026	
	S	0.975	
	Overall 2-year survival	0.965	

To determine the probability of survival at one-year for the low-risk patient:

Step 1:

$$lnY = 5.702963 - 0.0162938 * (Male) + 0.7360137 * ln(FVC\ \%\ Predicted/100) + 0.7899955 * ln(FEV1\ \%\ Predicted/100) - 0.7302478 * (B.cepacia\ Complex) - 0.4588687 * (Underweight) - 0.0398486 * (Age\ in\ years) - 0.2818584 * (\#\ Hospitalizations\ in\ preceding\ year)$$

$$lnY = 5.702963 - 0.0162938 * (0) + 0.7360137 * ln(66.7/100) + 0.7899955 * ln(47.4/100) - 0.7302478 * (0) - 0.4588687 * (0) - 0.0398486 * (16) - 0.2818584 * (0)$$

$$lnY = 4.18$$

Step 2:

$$ln\beta = 0.1146547 - 0.0792965 * (\#\ Hospitalizations\ in\ preceding\ year) - 0.5616525[ln(FEV1\ \%\ Predicted\ in\ Preceding\ Year\ /100) - ln(FEV1\ \%\ Predicted\ in\ current\ year/100)] + 0.2554754 * (ln(FEV1\ \%\ Predicted\ in\ current\ year/100)) + 0.6058589 * (Pancreatic\ Insufficient) + 0.2340407 * (CFRD) + 0.0079757 * (Age\ at\ CF\ diagnosis\ in\ years)$$

$$ln\beta = 0.1146547 - 0.0792965 * (0) - 0.5616525 * [ln(80.5/100) - ln(47.4/100)] + 0.2554754 * (ln(47.4/100)) + 0.6058589 * (1) + 0.2340407 * (0) + 0.0079757 * (0.9)$$

$$ln\beta = 0.24$$

Step 3:

$$Y = exp(lnY) = exp(4.18) = 65.4$$
$$\beta = exp(ln\beta) = exp(0.23) = 1.27$$

Step 4:

$$lnS = \frac{1}{(-1) * \beta}\left(\frac{1}{Y}\right)^{\beta}[exp(\beta) - 1]$$
$$lnS = \frac{1}{(-1.27)}\left(\frac{1}{65.4}\right)^{1.27}[exp(1.27) - 1]$$
$$lnS = -0.00997$$

Step 5:
$$S = exp(lnS)$$
$$S = exp(-0.00997)$$
$$S = 99.0\%$$

Therefore, the probability of survival at one-year is 99.0%.

Step 6:

Repeat Steps 1-5 to calculate S_2:

Step 6_1:

$$lnY = 4.55962 + .3189947 * (Male) + .5809873 * (ln(FVC\ \%\ Predicted/100)) + .8404154 * (ln(FEV1\ \%\ Predicted/100)) - 0.4187824 * (Underweight) - 0.9285728 * (B.\ cepacia\ Complex)$$
$$lnY = 4.55962 + .3189947 * (0) + .5809873 * (ln(66.7/100)) + .8404154 * (ln(47.4/100)) - 0.4187824 * (0) - 0.9285728 * (0)$$
$$lnY = 3.70$$

Step 6_2 :

$$ln\beta = .1863934 - .1263516 * (\#\ Hospitalizations\ in\ preceding\ year) + .1858131 * [ln(FEV1\ \%\ Predicted\ in\ Preceding\ Year/100) - ln(FEV1\ \%\ Predicted\ in\ current\ year/100)] + .4353779 * (ln(FEV1\ \%\ Predicted/100) + .1927758 * (Pancreatic\ Insufficient) - 0.172767 * (CFRD) + .0012487 * (Age\ at\ CF\ Diagnosis)$$
$$ln\beta = .1863934 - .1263516 * (0) + .1858131 * [ln(80.5/100) - ln(47.4/100)] + .4353779 * (ln(47.4/100) + .1927758 * (1) - 0.172767 * (0) + .0012487 * (0.9)$$
$$ln\beta = 0.15$$

Step 6_3:
$$Y = exp(lnY)$$
$$Y = exp(3.70)$$
$$Y = 40.45$$

$$\beta = exp(ln\beta)$$
$$\beta = exp(0.15)$$
$$\beta = 1.16$$

Step 6_4:

$$lnS = \frac{1}{(-1) * \beta}\left(\frac{1}{\overline{Y}}\right)^{\beta}[exp(\beta) - 1]$$

$$lnS = \frac{1}{(-1.16)}\left(\frac{1}{40.45}\right)^{1.16}[exp(1.16) - 1]$$

$$lnS = -0.0258$$

Step 6_5:

$$S_2 = exp(lnS)$$

$$S_2 = exp(-0.0258)$$

$$S_2 = 97.4\%$$

Step 6:

$S_2 * S_1 = 0.974 * 0.990 = 0.964$

Therefore, the overall probability of survival at 2 years is 96.4%.

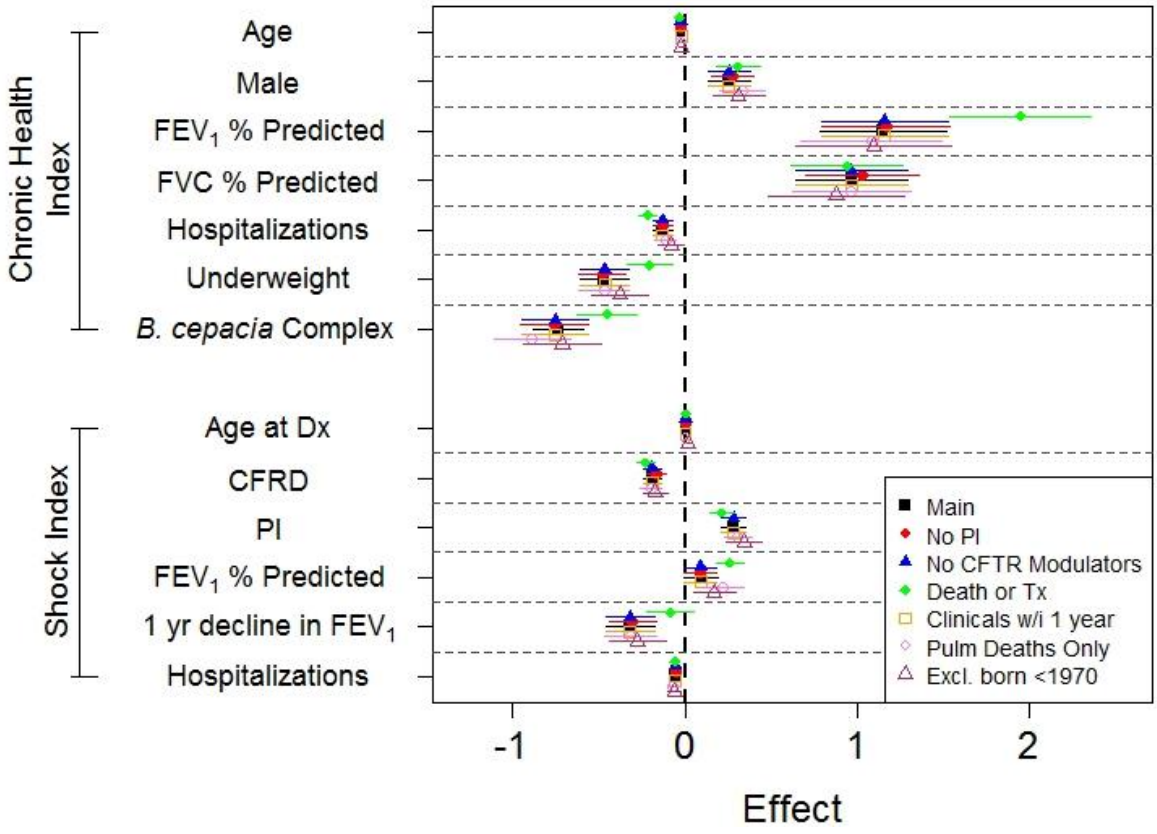
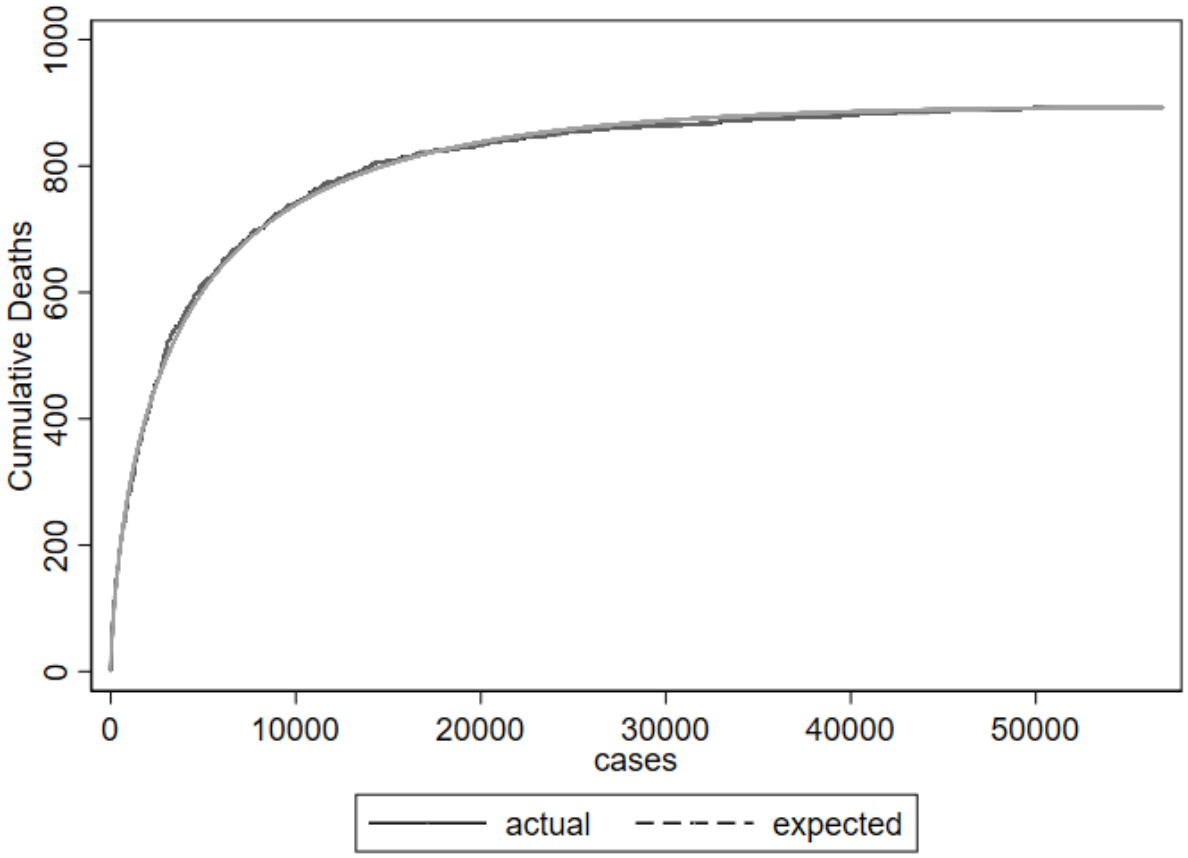


Figure S1: Summary of Sensitivity Analyses

a)


Cumulative Deaths
0
200
400
600
800
1000
0
10000
20000
30000
40000
50000
cases
actual
expected

b)

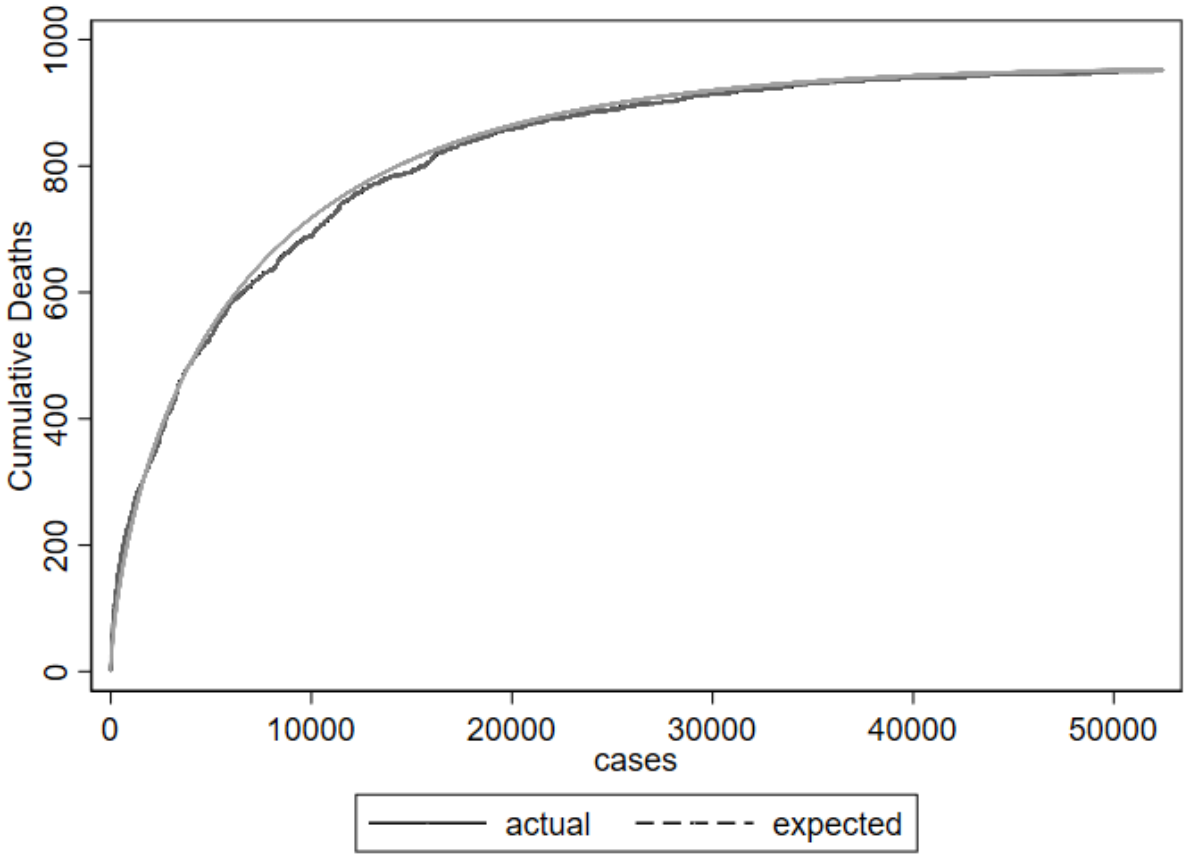
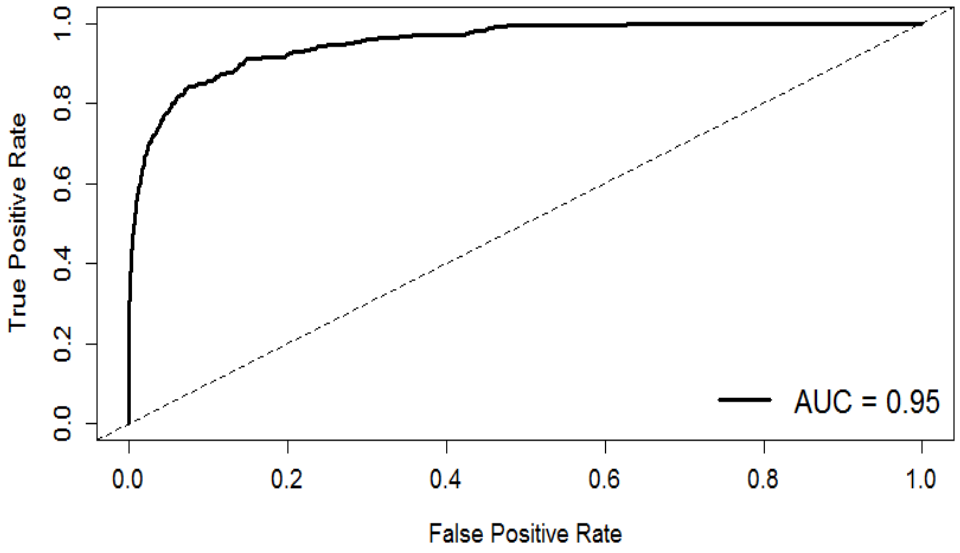


Figure S2: Goodness of fit for the un-weighted a) 1-year model and b) 2-year model for the combined health index and shock index. Both figures indicate good model fit as judged by comparison of the estimated probability of death and the actual deaths that were observed. The tracking of the two lines (actual and expected deaths) indicate good calibration; whereas the shape of the curve, the degree the curves bend towards the upper-left corner, indicates good predictive power.



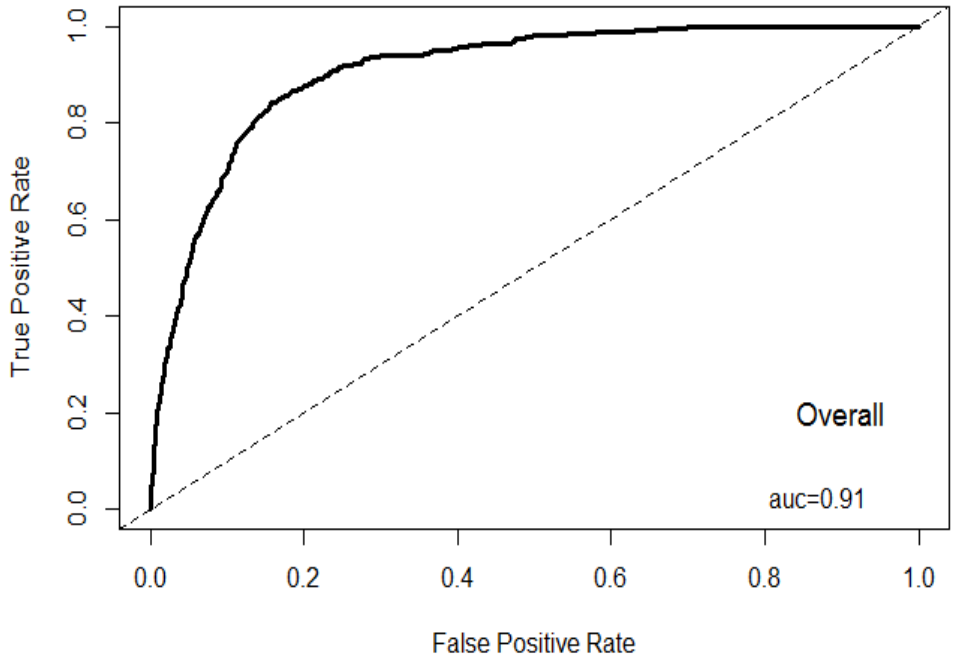


Figure S3: Goodness of fit for UK dataset validation for the a) 1-year model and b) 2-year model.

1. Konstan MW, Butler SM, Schidlow DV, Morgan WJ, Julius JR, Johnson CA. Patterns of medical practice in cystic fibrosis: part I. Evaluation and monitoring of health status of patients. Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. Pediatric pulmonology. 1999;28(4):242-7.
2. Stephenson AL, Tom M, Berthiaume Y, Singer LG, Aaron SD, Whitmore GA, et al. A contemporary survival analysis of individuals with cystic fibrosis: a cohort study. Eur Respir J. 2015;45(3):670-9.
3. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.
4. Organization WH. Body mass index classification Geneva, Switzerland [
5. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 2000;320(7244):1240-3.
6. Moran A, Hardin D, Rodman D, Allen HF, Beall RJ, Borowitz D, et al. Diagnosis, screening and management of cystic fibrosis related diabetes mellitus: a consensus conference report. Diabetes research and clinical practice. 1999;45(1):61-73.