Exp Biomed Res 2023; 6(1): 24-37 **DOI: 10.30714/j-ebr.2022.167**

Experimental Biomedical Research *Original article*

Joint modeling of survival and longitudinal data: Carrico index data example

Ebru Turgal [1*], Beyza Doganay Erdogan [1], Fatma Arslan [2]

[1]*Department of Biostatistics, Ankara University, Faculty of Medicine, Ankara, Türkiye*

[2]*Department of Chest Diseases, Ankara University, Faculty of Medicine, Ankara, Türkiye*

ABSTRACT

Aim: When the respiratory system is unable to adequately absorb oxygen or excrete carbon dioxide, acute respiratory failure (ARF) develops. A current area of study is the survival analysis of patients with acute hypercapnic respiratory failure (AHRF) in the field of pulmonary diseases. In the follow-up period, several biochemical markers are repeatedly measured, such as respiration rate and Carrico Index; however, baseline or averaged values are mostly related to treatment failure. Although this approach is not inaccurate, it causes information loss, which leads to biased estimates. This prospective cohort study primarily looked at the relationship between changes in Carrico Index and failure of treatment in AHRF patients.

Methods: We included 86 patients from Ankara University School of Medicine Pulmonary Diseases Department. The association between the trajectory of the Carrico Index and failure in AHRF patients was examined using a joint modeling approach for longitudinal and survival data.

Results: Results showed that averaged Carrico Index change was inversely and significantly associated with failure (HR: 0.97, 95% CI: -0.05 to 1.97). With hazard ratios of 1.43 and 1.4, chronic health evaluation II (Apache II), and COPD Assesment test (CAT) were positively correlated with failure risk.

Conclusions: The present study demonstrate that applying the risk predictors' trajectory through an appropriate statistical method improved accuracy and avoid biased results.

Key words: Joint modeling, non-invasive ventilation, carrico index, survival.

✉ * *Ebru Turgal,*
Department of Biostatistics, Ankara University, Faculty of Medicine, Ankara, Türkiye
E-mail: ebruturgal@gmail.com
Received: 2022-09-22 / Revisisons: 2022-10-23
Accepted: 2022-11-16 / Published: 2023-01-01

Introduction

ARF is defined as the inadequacy of respiratory function or oxygen / carbon dioxide gas exchange in the lung. When acute respiratory failure develops, non-invasive positive pressure mechanical ventilation (NPPV) is applied to the patient as the first option after optimal medical treatment. NPPV is a mechanical ventilation without any invasive artificial respiratory route. ARF was first discussed by Tschirgi [1]. AHRF was accepted as type II respiratory failure. In type II respiratory failure, hypercapnia predominates. $PaCO_2$ level is above 45 mmHg and respiratory acidosis is present.

Data collection in medical research can be done in a single period of time or in a method that collects data periodically at different time intervals. Repeated measurements occur when observations are taken at different time points or under different conditions from the same subject and are called "longitudinal data". On the other hand, by accounting for potential confounders such comorbidities, biochemical/chemical

variables, and demographic factors, the Cox proportional hazards regression model enables revising predicted survival probability. The Kaplan-Meier analysis and/or Cox proportional hazards regression models were widely used to predict the survival of a research group. The vast majority of studies ignored repeated measurements and used either the Kaplan-Meier or the Cox proportional hazard regression model based on a single measurement (i.e. baseline or average of multiple records) of related risk factors [2,3,4,5].

During the follow-up period, AHRF patients are monitored for PaO_2 / FiO_2 (Carrico Index), respiratory rate (RR), Glasgow Coma Scale and COPD Assesment test (CAT). These variables are measured repeatedly in AHRF patients; although averaged values of continuous risk factors related with AHRF can be used to analyze survival probabilities, repeated measurements may offer an in-depth insight for predicting them. The purpose of studies involving longitudinal data is usually to examine how the mean response profiles differ among groups as well as the time course of responses. Longitudinal data analysis is carried out by examining the variations both within the same individual and between individuals. Thus, the trajectories throughout the data collection range are investigated considering inter-individual biological fluctuations. On the other hand, a survival data comprise of event of interest (death, recurrence of a disease, transplantation, etc.) and time of the event. Conventionally, these two types of data collected under the same study are analyzed separately using different statistical methods. However, in some cases there is a relationship between longitudinal data and the survival process of the individuals (e.g. PSA antigen and prostate cancer [6], systolic and diastolic blood pressure measures and time to coronary artery disease [7]). When this is the case, a joint model should be fitted to obtain unbiased and more efficient results [8]. Recently, studies on modeling of these two processes together have been increasing, when the longitudinal and the survival processes are related [9, 10]. In order to obtain accurate results from these collected data, the necessity of using the joint modeling method has been showed [11]. Joint modeling is a model that links longitudinal and survival data, and consists of two linked sub-models. One of them corresponds to the measurement model for the longitudinal process and the other corresponds to the density model for the survival process.

Association between Carrico Index and mortality is well known [12]; besides, the relationship between the time-dependent Carrico Index levels and failure of treatment or mortality is a recent and long-standing area of research. Moreover, the majority of current research focus on population-based risk estimates and do not account for within-patient heterogeneity. Furthermore, personalized medicine and risk forecasts have gained popularity in recent years, particularly in the treatment and following-up of chronic diseases [13, 14] at the individual patient level. It is now possible to predict the hazard of treatment of AHRF patients individually using patient-level data and changes in biological markers over time, thanks to recent advances in statistical modeling. In addition, patient-specific risk predictions for future time points can be updated dynamically as new information becomes available known [5]. It is possible to predict the results of AHRF patient survival using the Carrico Index levels at the start of the study or averaged values while the follow-up period. However, long-term variation and trend in Carrico Index levels will present more accurate risk evaluations. We used a unique and recently proposed method called joint modeling to extend the survival model (usually the Cox

proportional hazards regression model) to repeated measurements (i.e. time-dependent coefficients) in order to investigate the relationship between a longitudinal biomarker and survival outcome for AHRF patients in this study.

To the best of the authors' knowledge, present study is the first application which investigates the joint modeling approach on AHRF patients in Turkiye and worldwide.

Materials and metods

Study Design and Participants

This section offers an overview of the overall research design and the plans for involving participants. Eighty-six patients, who had AHRF related to chronic obstructive pulmonary disease (COPD), acute cardiogenic pulmonary edema (ACPE), community-acquired pneumonia, bronchiectasis, or kyphoscoliosis and who were received NPPV as an initial ventilatory support strategy at Ankara University School of Medicine Pulmonary Diseases Clinic, were prospectively included in the study.

Patients aged ≥65 years who were hospitalized to the intensive care unit between January 2012 and September 2016 were followed up until treatment, lost to follow up owing to withdrawal or unknown reasons, or the trial ended, whichever took place first. Patients who were followed up for AHRF in the intensive care unit were included in the study. Carrico Index values were taken before treatment (0 hour), and 1, 2, 4, 12, 24, 48, 72, 96 and 120 hours after treatment. Measurements of PaO_2 / FiO_2 were taken as dependent variable in the longitudinal data model. Acute respiratory failure and time to failure were used in the survival model. Effect of time-dependent PaO_2 / FiO_2 measurements to ARF is analyzed with joint modeling of longitudinal data model and survival model. In this study we aim to apply joint modelling approach to predict the hazard of treatment of ARF and its relation to time dependent Carrico Index measurements in the intensive care unit data.

Clinical Outcome Assessments

The initial clinical outcome was treatment owing to AHRF. In this prospective cohort study, patients' data, including the demographic (e.g. gender, age, body mass index, smoking status) and clinical/biochemical measurements (e.g. leukocyte count, Charlson Comorbidity Index, serum C-reactive protein (CRP), Glasgow coma score, chronic health evaluation II (Apache II), COPD assessment test (CAT), hemoglobin levels, respiration rate, current or previous comorbid diseases (coronary heart disease (CHD), congenital adrenal hyperplasia (CAH), chronic renal failure (CRF), cerebrovascular disease (CVD), haematocrit (%), arrhythmia, NPPV history, intubation, diabetes) were collected from medical history. During the follow-up period, repeated measures of biochemical and clinical outcomes were recorded. We examined the longitudinal associations between change in biological markers and treatment. Apache II, Glasgow, and CAT were identified as independent variables in patients with AHRF brought on by COPD exacerbations that resulted NIV failure [15].

Statistical analysis

The R programming environment (version 3.4.3, URL: https:// cran.r-project.org) was used to perform the statistical analyses. For survival data, Cox proportional hazard regression analysis "survival" library [16], for linear mixed-effects model "nlme" library [17] for joint model, "JM" library was used [18].

Quantitative data are stated as percentage, mean and standard deviation. Categorical variables were summarized using frequencies and percentages. A joint model, consisting of two

submodels: i) a linear mixed effect (LME) model for assessing the longitudinal biological marker and (ii) a Cox proportional hazard model for treatment, was used to observe the link between the trajectory of longitudinal biological markers and treatment. The goal of the joint modeling technique is to predict the effect of longitudinal biological markers on treatment while adjusting for potential confounders in both the longitudinal and time-to-event outcomes. Data were analyzed using three approaches. In the first approach, longitudinal process and time-to-event process of the study (extended Cox regression model) were analyzed separately. In the second approach two stage modeling was applied. In the third approach the longitudinal and the survival process were modeled jointly. So the likelihood of two processes of the study was evaluated jointly. The details of these approaches are given below.

Time-Dependent Explanatory Variables in the Survival Process: Extended Cox Regression Model

The model is based on a proportional hazard assumption but does not assume a particular probability distribution for survival times. The Cox regression model is considered as a semi-parametric model since the basic hazard function is not specified, i.e. failure time, no probability distribution for the random variable *T* and the most important problem in the model is the estimation of the parameter β [19].

$h(x) = h_0(t)exp\left[\sum_{i=1}^{p}\beta_i x_i\right]$ (Equation 3)

In the equation, x is the vector of explanatory variables in the form $(x_1, x_2, \ldots, x_p)$. An important feature of this formula is the baseline hazard function ($h_0(t)$), which includes the survival time (t) but not the explanatory variables, in relation to the proportional hazard assumption. The part where the explanatory variables are included is independent of the survival time. The explanatory variables in this case are sometimes called time-independent. $h_0(t)$ shows how the risk of failure changes over time, whereas the exponential part of the explanatory variables has an effect on the hazard function. Time independent variables are variables that do not change in value over time, like gender, smoking status.

The Cox proportional hazard model can be extended as follows to deal with time dependent explanatory variables.

$h_i(t\,|\gamma_i(t), w_i) = h_0(t)R_i(t)exp\{\gamma^T w_i + \alpha y_i(t)\}$ (Equation 4)

In counting process notation, the event process for subject *i* is written as $\{N_i(t), R_i(t)\}$ with $N_i(t)$ denoting the number of events for subject *i* by time *t*, and $R_i(t)$ is a left continuous at risk process with $R_i(t)$ = 1 if subject *i* is at risk at time *t*, and $R_i(t)$= 0 otherwise.

w_i: a vector of baseline covariates, such as sex or randomized treatment,

$y_i(t)$: a vector of time-varying covariates.

The regression coefficients vector α is interpreted in the same way as γ.

The model shown above with Equation 4 is known as the extended Cox hazard model [19]. In this model, instead of explanatory variables in the traditional Cox model (Equation 3), time-varying and time-independent explanatory variables and required regression coefficients are separately found. In this model, w_i shows time-independent explanatory variables (e.g. gender, treatment groups, etc.), y_i represents time-dependent explanatory variables. In the model, α and γ are regression coefficients of time-independent and time-varying explanatory variables, respectively. The interpretation of the regression coefficients of time-dependent common variables can be made as follows.

A unit increase in $y_i(t)$ measured at time *t* can be said to increase the relative risk at time *t* of

interest by *exp (α)*. For example, in a study of the risk of death in cirrhosis patients and the effect of prothrombin index values repeatedly measured weekly during the time they were hospitalized, *exp (α)* = 3 was found in the prothrombin index. According to this, a unit increase in the measure of prothrombin index in the second week when patients are in the hospital increases the risk of death in the second week by 3 times.

However, in the extended Cox model, since the inter-individual biological heterogeneity is not considered, when the time-varying variables are endogenous, the endogenous variation effect of the individual biological variation is ignored. In joint modeling, the longitudinal sub-model of time-varying explanatory variables and the survival sub-model are connected by random effects that show inter-individual heterogeneity. In this way the effect of individual biological variations to the time-dependent variables are also taken into account [14].

Joint Modeling

Joint modeling is a model linking longitudinal and survival data and consists of two connected sub-models. One of them corresponds to the mixed effect model for the longitudinal process and the other corresponds to the proportional hazard model for the survival process. These two sub-models are connected to each other by random effects. Joint modeling of common random effects and longitudinal and survival processes allows simultaneous estimation of these two sub-models. In joint modeling, the survival process may be solved with the joint likelihood of the longitudinal process.

Sub-Models to be used in Joint Model

Longitudinal Submodel

As a longitudinal sub-model, a mixed effects model with random and fixed effects can be used. The mixed effect sub-model shown in Equation 5 for use in the joint model could be rewritten as follows:

$M_i(t) = \{m_i(s), 0 \le s < t\}$;
$y_i(t) = m_i(t) + \varepsilon_i(t)$,
$m_i(t) = x_i{}^T(t)\beta + z_i{}^T(\mathrm{t})\, b_i$,
$b_i \sim N(O, D)$, $\varepsilon_i(t) \sim N(O, \sigma^2)$, (Equation 5)

where, $M_i(t)$ represents the expected values of the longitudinal process until *t* time. That is, the actual values of the measurements taken in the longitudinal process, adjusted from measurement errors. $m_i(t)$ is the expected value of the longitudinal explanatory variable at time *t*. $m_i(t)$ is different from $y_i(t)$ because it does not include the measurement error for the longitudinal outcome variable at time *t*.

$x_i(t)$ is the design vector of fixed effects at time *t* for individual *i*, $z_i(t)$ is the design vector of random effects at time *t* for individual *i*, whereas β and b_i are the corresponding regression coefficients. *D* is the (q × q) dimensional general covariance matrix of random effects. In the above model $x_i(t)$, $z_i(t)$ and $\varepsilon_i(t)$ terms are all time-dependent.

Survival Sub-Model

Before specifying the survival submodel used in the joint model, the cumulative hazard function could be defined as given in Equation 6.

$h_i(\mathrm{t}\,|M_i(t), w_i) = Pr\ Pr\ \{t \le T_i^* < t + dt \mid T_i^* \ge t, M_i(t), w_i\}/\ \mathrm{dt} = h_0(t) exp\ \{\gamma^T w_i + \alpha\, m_i(\mathrm{t})\}$, t>0, (Equation 6)

In this model, $M_i(t)$ is the measurement error-free true values measured up to time *t* in the longitudinal process, and w_i shows the baseline explanatory variable vector (e.g. treatment indicator, disease history). γ is the parameter vector containing the regression coefficients for the explanatory variables. α parameter is the regression coefficient indicating the effect of the longitudinal process on survival. $exp(\gamma_j)$ specifies the hazard ratio for a unit change in w_{ij} at any time during at *t*. *exp (α)* is the relative increase of an event at time *t*, and this is the result of a unit increase in $m_i(t)$ at the same time point.

Notably, the survival function to be utilized in the joint model may therefore be defined as follow using the known relationship between the survival and the cumulative hazard function.

$S_i(t|M_i(t), w_i) = Pr\ Pr\{t > T_i^* \mid M_i(t), w_i\}$

$= exp\ (-\int_0^t h_0(s) exp\{\gamma^T w_i + \alpha m_i(s)\}ds)$, (Equation 7)

Survival function also depends on the values of the baseline explanatory variables $M_i(t)$ and the longitudinal explanatory variable values. Here, there are various options to determine the structure of the baseline risk function $h_0(\cdot)$. The classical option is to use the known risk function with known parametric distribution. The distributions commonly used for the baseline risk function within the scope of survival analysis are Weibull, log-normal and Gamma. The second option is to use a risk function which is also parametric but more flexible. In literature, several approaches have also been proposed to model the flexible basic risk function, for example B-splines or cubic splines.

When selecting variables for the Cox regression model; variables with significance level above 0.20 in the univariate Cox analysis were not included into the analyses [20]. Then, Model 1.a was obtained by considering the clinical significance with the backward elimination method.

Primarily, to identify a subset of significant independent variables from among biochemical, clinical, and demographic factors, univariate Cox proportional hazard modeling and an LME model were used. The survival and longitudinal submodels were then linked to a joint model, and model parameters were determined simultaneously [5]. The longitudinal nature of the dependent variable in joint modeling is represented using an LME model described in Equation 8.

$y_{it} = m_{it} + \varepsilon_{it}$, (Equation 8)

$y_{it} = m_{it} = \beta^T X + b^T Z + \varepsilon_{it}$, (Equation 9)

where X and Z are the vectors of fixed and random effects with β and b, respectively, the vectors of regression parameters, and ε_i is the random error term of the *i*th patient. With an association parameter, the fitted trajectories from the longitudinal model were included as a time-dependent covariate in the survival analysis section. As in Equation 10, the survival submodel can be defined,

$h_i(t) = h_0(t)\exp(\gamma_i{}^T w_i + \alpha m_i(t))$, (Equation 10)

where $\gamma_i{}^T$ is the vector of model parameters, $m_i(t)$ is the fitted curves of trajectory of the Carrico Index generated using a LME model for the *i*th patient at time *t*, and w_i is the vector of baseline covariates of the *i*th patient related with failure or mortality [11]. The association parameter between the longitudinal and survival sub-models called as α.

After taking into account potential confounders, an LME model is used to incorporate the impact of the longitudinal biomarkers into the survival model. There is no relationship between the treatment and the longitudinal biological marker, if the parameter α is statistically insignificant. The JM package, which was developed especially for joint modeling of longitudinal and survival data, was used to carry out the investigations in R [18].

Several biomarkers, including the Carrico Index, might be associated with AHRF-related failure of treatment. Carrico Index as a longitudinal response was the primary focus of this work. p value was set 0.05 significance level in all analyses.

Results

Clinical characteristics of patients with successful treatment and of those with unsuccessful treatment were given in Table 1.

Table 1. Demographic, clinical/biochemical characteristics of the patients.

Parameters	**Total**	**Patients with successful treatment (n=73)**	**Patients with unsuccessful treatment (n=13)**
Gender (male)	44 (0.51)	33 (0.45)	11 (0.85)
Age	71.62±10.98	70.75±11.25	76.46±8.08
BMI(kg/m^2)	28.54±7.95	29.78±7.88	21.57±3.51
CAT	31.21±4.71	30.19±4.23	36.92±2.84
Respiratory rate	23.45±4.74	22.19±3.56	30.54±4.35
Diabetes	30 (0.35)	24 (0.33)	6 (0.46)
CRF	26 (0.30)	19 (0.26)	7 (0.54)
CAH	9 (0.10)	8 (0.11)	1 (0.08)
Aritmia	19 (0.22)	14 (0.19)	5 (0.38)
Haematocrit	63 (0.73)	54 (0.74)	9 (0.69)
CVD	1 (0.01)	1 (0.01)	0 (0.00)
Intubation	11 (0.13)	10 (0.14)	1 (0.08)
Glasgow	14.72±0.63	14.92±0.32	13.62±0.77
Apache	17.99±3.88	16.88±2.96	24.23±2.13
PaO_2/FiO_2			
0. hour	207.80±27.21	214.92±22.95	168.39±8.99
1. hour	215.64±25.27	222.80±20.04	176.54±10.68
2. hour	220.75±25.67	228.38±19.36	179.08±12.01
4. hour	224.40±25.57	231.67±19.04	180.82±13.25
6. hour	227.33±25.85	234.92±18.82	183.18±14.18
12. hour	230.91± 27.63	238.51±21.11	186.00±16.67
24. hour	233.19±29.28	241.11±22.90	187.09±17.33
48. hour	237.16±30.21	245.37±22.16	183.8± 18.94
72. hour	241.55±29.05	248.27±21.70	188.63±25.70
96. hour	246.56±24.91	249.70± 22.29	201.75± 16.5
120.hour	246.00±24.38	250.00±20.02	193.00±10.89

Mean±standard deviation was given for quantitative data. Percentage was given for categorical data.

These data were used to model longitudinal process and survival process for each of the three approaches.

The results of the extended Cox regression model, two stage approach and joint model are showed in Table 2. The extended Cox regression model, two stage approach and joint model results showed that Apache II, CAT and Carrico Index were associated with failure of treatment. In parallel, the joint model's survival component was used to calculate the final risk estimates and the results were compared with the time-dependent Cox model and two-stage model (Table 2).

Modeling of two processes separately

First, we modeled longitudinal and the survival data part of the study separately thus each model was interpreted independently. A linear mixed effect model was fitted to longitudinal data, while an extended Cox regression model is used for the survival process by taking the PaO_2 / FiO_2 biomarker values as a time dependent covariate.

Model 1.a.

$$PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 time + \beta_2 time * time + \beta_3 Apache\ II + \beta_4 RR + \beta_5 Glasgow + b_{i0} + b_{i1} * time + b_{i2} * time * time + \varepsilon_i(t)$$

Model 1.b.

$$h(t) = h_0(t) exp(\beta_1 Apache\ II + \beta_2 Cat + \beta_3 PaO_2/FiO_2(t))$$

Two-stage modelling

The idea behind this model is to take the dependent variable (PaO_2 / FiO_2) estimates from linear mixed effects model (Model 1.a) as a time-varying explanatory variable in the extended Cox model (Model 1.b). When building a linear mixed effects model, repeated measures of PaO_2 / FiO_2 are modeled depending on time, Apache II, RR, and Glasgow score. Random intercept and random slope over time were also included as random effects. In the Cox regression model part of the joint model, we included CAT and Apache II score as time-independent variables and true estimates of PaO_2 / FiO_2 at time *t*, shown as $m_i(t)$.

Model 2.a.

$$PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 time + \beta_2 time * time + \beta_3 Apache + \beta_4 RR + \beta_5 Glasgow + b_{i0} + b_{i1} * time + b_{i2} * time * time + \varepsilon_i(t)$$

Survival submodel that splits the time effect into step functions was used for estimating the hazard on each cut interval.

Model 2.b.

$$h(t) = h_0(t) exp\left(\beta_1 Apache + \beta_2 Cat + \beta_3 \widehat{PaO_2/FiO_2}(t)\right)$$

The relative risk model with a Weibull baseline risk function ($h_0(t)$) was used for estimating the hazard. Weibull baseline risk function is a flexible model for survival data and has a hazard rate either monotone increasing or decreasing or constant.

$$f(t) = \lambda\, p\, (\lambda)^{p-1} e^{-(\lambda t)^p}, \quad t \geq 0,\ p,\ \lambda > 0,$$
$$S(t) = exp[-(\lambda t)^p],$$
$$h(t) = \lambda\, p (\lambda t)^{p-1} \qquad \text{(Equation 11)}$$

As seen from the above formula, it is defined by a shape (p) and a scale (λ) parameter [21]. After we described two submodels of longitudinal and survival processes, the joint model can be written as below.

Joint Modeling

Model 3.a.

$$PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 time + \beta_2 time * time + \beta_3 Apache + \beta_4 RR + \beta_5 Glasgow + b_{i0} + b_{i1} * time + b_{i2} * time * time + \varepsilon_i(t)$$

Table 2. Parameter estimates from different modeling strategies.

Parameters		Separate Model			Two Stage Model			Joint Model		
		$coefficient \pm SE$		*p-value*	$coefficient \pm SE$		*p-value*	$coefficient \pm SE$		*p-value*
Longitudinal Process	$Intercept$	77.56±66.49		*0.244*	77.56±66.49		*0.244*	78.43±67.56		*0.245*
	$Time$	4.12±0.29		*<0.001*	4.12±0.29		*<0.001*	4.11±0.29		*<0.001*
	$Time^2$	0.23±0.06		*0.0004*	0.23±0.06		*0.0004*	0.22±0.06		*0.0005*
	APACHE II	-1.49±0.63		*0.021*	-1.49±0.63		*0.021*	-1.49±0.63		*0.018*
	RR	-1.61±0.49		*0.0016*	-1.61±0.49		*0.0016*	-1.61±0.49		*0.011*
	$Glasgow$	19.91±3.92		*0.0007*	19.91±3.92		*0.0007*	13.87±4.00		*0.0005*
	σ_{b_0}	18.89			18.89			18.86		
	σ_{b_1}	2.38			2.38			2.39		
	σ_{b_2}	0.46			0.46			0.46		
	σ_{ε}	5.59			5.59			5.59		
		HR	**%95 CI**	*p-value*	**HR**	**%95 CI**	*p-value*	**HR**	**%95 CI**	*p-value*
Survival Process	$Apache\ II$	1.32	(0,28;2.36)	*<0.001*	1.16	(0.11;2.21)	*0.0047*	1.43	(0.31;2.55)	*0.001*
	CAT	1.15	(0.11;2.19)	*<0.001*	1.17	(0.13;2.21)	*<0.001*	1.4	(0.27;2.53)	*0.009*
	$PaO_2/FiO_2(t)$	0.97	(-0.04;1.98)	*<0.001*	0.96	(-0.05;1.97)	*<0.001*	0.96	(-0.05;1.97)	*0.004*
	$log(p)$	NA		NA	NA		NA	3.01		*<0.001*
	$log(\lambda)$	NA		NA	NA		NA	-44.97		*<0.001*

SE: Standard error, RR: Respiratory Rate, HR: hazard rate, CI: confidence interval.

Model 3.b.

$$h(t) = h_0(t)exp(\gamma_1 Apache + \gamma_2 Cat + \alpha m_i(t))$$

When joint model was built, likelihoods of model 3.a for linear mixed model and model 3.b. for Cox regression model were optimized together.

Figure 1 shows the subject-specific PaO_2 / FiO_2 index profiles and the change in PaO_2 / FiO_2 levels in time, where the bold lines are for average profiles for patients with successful NNPV and unsuccessful NNPV. As seen from the figure, the individual trends deviate from the average trends for both successful and unsuccessful treatment groups. PaO_2 / FiO_2 levels were higher in successfully treated patients and changed a little over time. However, in failure situation, descents and ascents in PaO_2 / FiO_2 levels were observed. Moreover, there was significant variability in PaO_2 / FiO_2 levels between patients. Since the average trends for both groups seem to curvilinear, we tested both quadratic and linear trend over time in modeling the longitudinal process.

Two different baseline hazard functions (Weibull and unspecified) used for the Cox regression model part. The model with Weibull

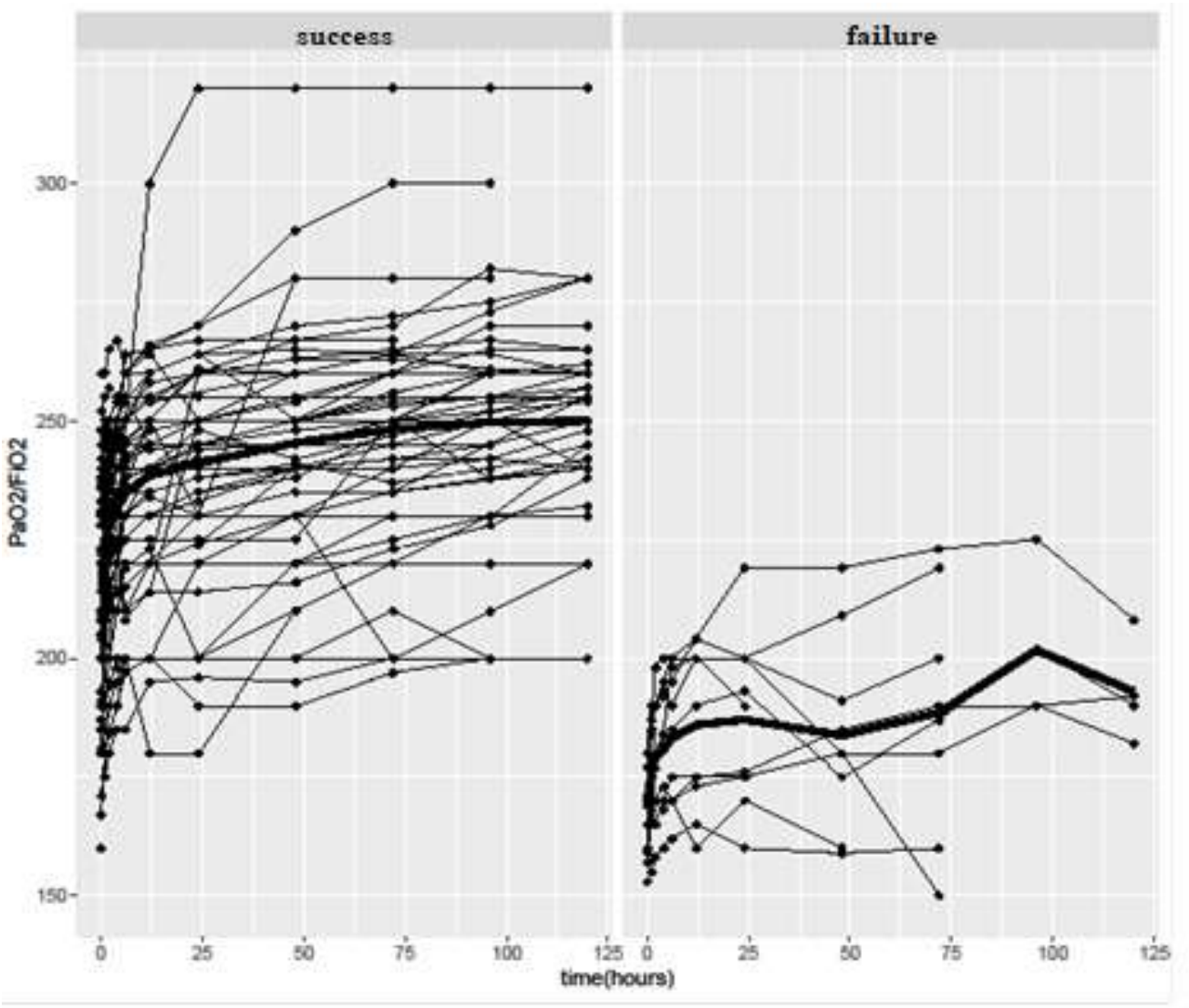

Figure 1. Subject-specific longitudinal profiles of the PaO_2 / FiO_2 index for the Carrico-Index data per group. Bold line shows the average trend over time for the group.

baseline hazard was selected for the final joint model according to the AIC and BIC criteria (AIC: 5865.09, 5912.725 and BIC: 5909.227, 5951.995, for Weibull and unspecified baseline hazard functions respectively).

As seen from Table 2, parameter estimates of linear mixed effects model are the same with separate and two stage approaches as we used exactly same models and estimation procedures for the longitudinal process. When we look at survival process in separate and two stage approaches, the effect of Apache II was diminished while the effects of CAT and Carrico Index were slightly increased. Carrico Index entered to the model as a time-varying variable in separate analysis of survival process, while estimated Carrico Index values were put in survival process of the two stage model (In two stage model, by fitting a longitudinal model the random effects are estimated in the first stage, and in the second stage these random effects are put in the extended Cox proportional hazards model).

In the joint modeling part, both longitudinal process and survival process models are fitted simultaneously using shared parameters. We see

that the effects of Apache II and CAT were increased, whereas the effect of Carrico Index was the same with two stage approach (Table 2). In the joint modeling, the patient-specific trajectory of Carrico Index levels was fitted to an LME model, and were associated with treatment through the time-to event submodel. Based on the joint modeling results, Apache II, CAT and PaO_2 / FiO_2 were found statistically significant. The estimated PaO_2 / FiO_2 increased 1.49 mm Hg and 1.61 mm Hg with every 1 score decrease in Apache II and RR, respectively. Moreover, there was 13.87 mm Hg increase in PaO_2 / FiO_2 with every 1 score increased in Glasgow score. There was a significant and inverse relationship between treatment success and the trajectory of PaO_2 / FiO_2 levels. 1 mm Hg decrease in the PaO_2 / FiO_2 at a time point *t* resulted in 1.041 (1/0.96) times higher risk of failure. The risk of failure increased 1.43 times with one unit increase in Apache II score, and also the risk of failure increased 1.4 times with one unit increase in CAT score.

Discussion

Using a joint modeling method, we investigated at the association between PaO_2 / FiO_2 levels and treatment in AHRF patients. Three types of modeling approaches were established for the analysis of changes in PaO_2 / FiO_2 measurements over time and the factors affecting survival (failure of treatment). Our study revealed that changes (ascents and descents) in PaO_2 / FiO_2 in time were strongly and significantly related with failure. In AHRF patients, the joint modeling approach offered more accurate survival predictions than the extended Cox regression and two-stage model. The following are possible explanations for why the joint modeling approach was more accurate: (i) it used the cumulative and historical information of PaO_2 / FiO_2, (ii) the true and unobserved value of the longitudinal outcome, here PaO_2 / FiO_2, was estimated by the linear mixed model, (iii) model parameters were computed simultaneously by taking into account the relationship between the longitudinal and survival processes and (iv) the patient-specific random effects were used to estimate the trend of PaO_2 / FiO_2 levels. According to the findings, the risk of failure increased by 1.041 times with every 1 mm Hg decrease in PaO_2 / FiO_2 at any time point. The decrease in the PaO_2 / FiO_2 ratio, which is frequently used in the evaluation of ventilated patients, is a sign of the presence of abnormal gas exchanges. Detection of less than 200 mm Hg indicates "Severe Hypoxemia". PaO_2 / FiO_2 is an indicator of the lung function in mechanically ventilated critically ill patients. Although PaO_2 / FiO_2 ratio is easy to calculate and correlates with the severity of respiratory failure, it is an imperfect measure as it varies with different positive end-expiratory pressure (PEEP) levels and tidal volume [22].

In prior studies, low PaO_2 / FiO_2 at baseline AHRF, one time point after NPPV initiation, or averaged in time was related with failure and mortality [23]. Nevertheless, few research, including the current one, have concentrated on PaO_2 / FiO_2 trajectory [24, 25]. Zhang [25] reported every PaO_2 / FiO_2 measurement was observed for each patient during the follow-up period. In the current research, it is observed that different patterns among the overall trajectory of PaO_2 / FiO_2 measurements for unsuccessfully treated group. PaO_2 / FiO_2 levels were higher in successfully treated patients and changed a little in time and were shown in Figure 1.

To sum up, this study evaluated the effect of mainly scores (i.e. Glasgow, CAT, Apache II) on PaO_2 / FiO_2 levels in AHRF patients. The model parameters were returned in two divisions after building the joint model: the main effects and the

random effects, which may be used to assess population-based and patient-specific risk predictions, respectively.

Also, the present study concentrated on the trajectory of PaO_2 / FiO_2 and analyzed its relationship with treatment failure. Results of the joint modeling indicated that the Apache II and CAT levels were associated with treatment failure, similar to the results of the extended Cox proportional hazard model.

Briefly, rather of using a single biological marker value to make a treatment, physicians prefer to use repeated marker values to explore the patients' longitudinal trajectory in detail on either the best treatment option or the diagnosis. Present study was aimed to identify the best joint model combination for obtaining the optimal model for AHRF patients. Within the context of this analysis, extended Cox regression model and LME separately, two-stage model and joint including two different parameterization tecniques to link between longitudinal and survival sub-models are included. To the best of the authors' knowledge, it's the most comprehensive study that analyzes different combinations such in dept in this area, compares the criterias under these combinations.

Model Selection Criteria

After the evaluation of model fit, several models fitted can be compared with the help of certain information criteria. Commonly used ones are the Akanke Information Criteria (AIC) and the Bayesian Information Criteria (BIC). Also, it is known that the BIC method in the large sample gives better results than the AIC [21].

Limitations, Possible Sources of Bias, and Generalization Issues

There are several constraints to this study. This is a single-center prospective cohort research, and the findings may reveal center-specific effects. For this reason, the validity of the present study's findings should be evaluated in the presence of nonmeasured confounding variables. This research has potential for improvement. The multivariate extension of joint modeling has the potential to increase model performance. Because a larger sample size is required, multivariate joint modeling was not discussed in this research. This is left as a research subject for a larger study group.

Appendix

R code to fit Cox models and joint models

Description: The R script illustrates the use of the R package JM for fitting joint models for longitudinal and survival data.

Pulmonary Diseases 86 patients dataset

```
# Carrico Index (CI) = longitudinal marker
# ST = Survival times
# SS = Survival status
# CITIMES = Carrico Index measurement times
#Baseline covariates: Apache II Score, Glagow Score, respiratory rate (RR), The COPD Assessment Test (CAT)
R script
#load first package JM
library JM
#linear mixed effects model with random intercept and slope
lme.fit <-lme(CI~ CITIMES+CITIMES^2 +Apache II Score, +Glasgow+RR, ~ CITIMES+CITIMES^2 | id, data = lmedat)
#basic Cox PH model;
cox.fit<-coxph(Surv(ST, SS)~ CAT+Apache II Score, data=coxdat, x=TRUE)
#joint model with a relative risk submodel for the event time outcome, the baseline risk function with Weibull and Cox.
joint1<- jointModel(lme.fit, cox.fit, timeVar = "CITIMES", method ="weibull-PH-aGH")
joint2<- jointModel(lme.fit, cox.fit, timeVar = "CITIMES", method ="Cox-PH-aGH")
```

Funding: *The authors received no financial support for the research, authorship, and/or publication of this article.*

Conflict of Interest: *The authors declare that they have no conflict of interest.*

Ethical statement*: Ethics committee approval was received for the study (date: 01.04.2022, meeting number: 151, decision number: 2022/3732 :(9405)).*

References

[1]Tschirgi RD, Gerard RW. The carotid-mandibular reflex in acute respiratory failure. Am J Physiol. 1947; 150(2): 358-64.

[2]Ahmed Y, Adam M, Bakkar LM. Effectiveness of APACHE II and SAPS II scoring models in foreseeing the outcome of critically ill COPD patients. Egypt J Bronchol. 2019; 13(5): 654-59.

[3]Alaithan AM, Memon JI, Rehmani RS, et al. Chronic obstructive pulmonary disease: hospital and intensive care unit outcomes in the Kingdom of Saudi Arabia. Int J Chron Obstruct Pulmon Dis. 2012; 7:819-23.

[4]Zheng DW, Wang CZ, Liu RS, et al. The application of improved Glasgow coma scale score of 15 as switching point for invasive noninvasive mechanical ventilation in treatment of severe respiratory failure in chronic obstructive pulmonary disease]. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue. 2011;23(4):224-7.

[5]Basol M, Goksuluk D, Sipahioglu MH, et al. Effect of serum albumin changes on mortality in patients with peritoneal dialysis: a joint modeling approach and personalized dynamic risk predictions. Biomed Res. Int. 2021; 2021, 6612464.

[6]Yu M, Taylor JMG, Sandler HM. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. J Am Stat Assoc. 2008; 103(481): 178-87.

[7]Yang L, Yu M, Gao S. Joint models for multiple longitudinal processes and time-to-event outcome. J Stat Comput Simulat. 2016; 86(18): 3682-3700.

[8]Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. J Clin Oncol. 2010; 28(16): 2796.

[9]Ediebah DE, Galindo-Garre F, Uitdehaag BM, et al. Joint modeling of longitudinal health-related quality of life data and survival. Qual Lif Res. 2015; 24(4), 795-804.

[10] Lim HJ, Mondal P, Skinner S. Joint modeling of longitudinal and event time data: application to HIV study. J Med Stat Informat. 2013; 1(1): 1.

[11] Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R. CRC press; 2012.

[12] Fagyas M, Fejes Z, Sütő R, et al. Circulating ACE2 activity predicts mortality and disease severity in hospitalized COVID-19 patients. Int J Infect Dis. 2022; 115: 8-16.

[13] Konar NM, Karaismailoglu E, Karaagaoglu E. The Use of Joint Modeling Approach in Personalized Medicine, In: Abstracts of the 9th Conference of the Eastern Mediterranean Region and the Italian Region of the

International Biometric Society. Thessalonoki, Greece, 8-12 May 2017, pp. 91.
[14] Rizopoulos D, Taylor JM, Van Rosmalen J, et al. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. Biostatistics. 2016; 17(1): 149-64.
[15] Çiftci F, Çiledağ A, Erol S, et al. Non-invasive ventilation for acute hypercapnic respiratory failure in older patients. Wiener klinische Wochenschrift. 2017; 129(19): 680-86.
[16] Therneau T. A package for survival analysis in S. R package version, 2015. 2(7). Available from: https://CRAN.R-project.org/package=survival. 2022/
[17] Pinheiro J, Bates D, DebRoy S, et al. Linear and nonlinear mixed effects models. R package version, 2007; 3(57): 1-89. R package version 3. pp. 1-157. Available from: https://CRAN.R-project.org/package=nlme/
[18] Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. J Stat Soft. 2010; 35: 1-33.
[19] Kleinbaum DG, Klein M. Survival analysis: a self-learning text (Vol. 3). New York: Springer; 2012.
[20] Auton T Applied survival analysis: regression modelling of time to event data. J Appl Stat. 2000; 27(1): 132-133.
[21] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data (Vol. 1230). New York: Springer; 2003.
[22] Seeley E, McAuley DF, Eisner M, et al. Predictors of mortality in acute lung injury during the era of lung protective ventilation. Thorax. 2008; 63(11): 994-98.
[23] Satou T, Imamura H, Mochiduki K, et al. Efficacy of protocol-based non-invasive positive pressure ventilation for acute respiratory distress syndrome: a retrospective observational study. Acute Med Surg. 2020; 7(1): e465.
[24] Antonelly M, Conti, G, Esquinas A. A multiple-centre survey on the use in clinical practice of noninvasive ventilation as a first-line intervention for acute respiratory distress syndrome. Crit Care Med. 2007; 35(1): 18-25.
[25] Zhang L, Wang Z, Xu F, et al. The Role of Glucocorticoids in the Treatment of ARDS: A Multicenter Retrospective Study Based on the eICU Collaborative Research Database. Front Med. 2021;8: 678260.