Modeling of Repeated Measures for Time-to-event Prediction

Jianfei Zhang[1], Lifei Chen[2], and Shengrui Wang[1,2(✉)]

[1] Université de Sherbrooke, Sherbrooke, Québec, Canada
jianfei.zhang@usherbrooke.ca
[2] Fujian Normal University, Fuzhou, Fujian, China

Abstract. Predicting the time for an event to occur while simultaneously exploring the coexisting effects of various risk factors has captivated considerable research interest. However, the profusion of repeated measurements involving a diverse array of risk factors has outpaced the capabilities of current methods for analyzing time-to-event data. In this paper, we propose a novel approach that entails the conversion of the time-to-event analysis conundrum into a sequence of discrete survival learning and prediction tasks, each approached autonomously. Our innovative strategy for modeling repeated measures facilitates the quantification of measurement impacts on projected outcomes at distinct junctures. When extrapolating the trajectory of health status over time, our method harnesses both censored and uncensored data to refine logistic regression parameters. Through a series of comparative experiments and meticulous ablation studies conducted on two real-life health datasets, we underscore the intrinsic practical promise of our method. Notably, our approach showcases its efficacy in prognosticating the temporal aspects of breast cancer patient mortality and the onset of disabilities among the elderly.

1 Introduction

Time-to-event data materializes when attention zeroes in on the passage of time (measured in years, months, weeks, or days) from the inception of a follow-up study until the occurrence of a specific event. In medical research, this event of significance is typically an adverse incident encompassing facets like injury, the inception or resurgence of an ailment, (re)hospitalization, and even demise. The pursuit of time-to-event prediction seeks to tackle inquiries such as *"How much time remains before the event takes place?"* - a question that has held enduring practical import. Foreseen time-to-event insights empower clinicians to promptly address patients' queries about potential outcomes, while decision-makers glean information about the probable timing of disease-related rehospitalizations. To illustrate, envision prognosticating outcomes for individuals diagnosed with breast cancer. One could discern that a 60-year-old patient boasts a 70% chance of surviving one year and a 40% probability of reaching the three-year milestone. These prognostic revelations can exert a pivotal influence on

X. Yang et al. (Eds.): ADMA 2023, LNAI 14176, pp. 48–63, 2023.
https://doi.org/10.1007/978-3-031-46661-8_4

treatment choices, lifestyle adaptations, and, on occasion, considerations concerning end-of-life care.

Within time-to-event data, the outcome encompasses not only the occurrence or non-occurrence of an event but also the specific timing of that event. The events sometimes cannot be observed throughout the follow-up period due to various censorships [10]. Hence, predicting time-to-event accurately is challenging and even impossible. Standard regression models fall short in their capacity to incorporate both the event occurrence and the timing aspects as outcomes within the same model. In the recent past, a significant multitude of methods has been devised to confront this quandary, facilitating predictions that effectively consider both dimensions concurrently, e.g., statistical models [16] and machine-learning-based methods [25]. To make the problem simple, we can answer the question "*When and how probably will be free of failure?*"

Investigating the simultaneous impacts of risk factors on the outcome holds significant relevance for both researchers and clinicians, and has been a focal point of extensive research endeavors. However, the measurement of risk factors often occurs over time or through real-time devices during the follow-up period, particularly in the context of extended-term follow-ups. Consequently, this results in an abundance of recurrent measurements, leading to fluctuations in the corresponding risk factors over time. Consider breast cancer patients who are monitored over a span of time, undergoing monthly assessments like white blood cell counts. Similarly, in studies focusing on cognitive changes related to aging, researchers gather data on participants' cognitive capabilities along with details about their date of death or diagnosis of dementia. These instances involve time-dependent risk factors like mental health, injuries caused by falls, and smoking. It is a great challenge for current time-to-event prediction methods to take care of these complex repeated measures, although they can use a single measure; of course, this will ignore the effect caused by historical measures.

Analyzing repeated measures and inference needs special techniques. In this context, approaches such as time-varying analysis [30] can be employed, where a single measurement evolves over time and is subsequently substituted by successive measurements to explore short-term relationships. Alternatively, measure-averaged analysis [29] can be employed, wherein a participant's initial measurement is solely represented by risk factor values at a specific point in time or a single estimate derived from the mean of measurements over a defined period. It's noteworthy that within the literature [21,28], the impact of repeated measurements on time-to-event outcomes has been demonstrated. Consider an observational study focusing on the effects of a drug on specific health indicators. In this scenario, a patient's present health status could influence their future drug exposure or dosage. Consequently, establishing a model capable of automatically accommodating repeated measurements is of paramount importance.

In this paper, we undertake a transformation of the time-to-event prediction task, reconfiguring it into a sequence of challenges involving survival probability estimates (SSP) at distinct time junctures. For each of these estimation challenges, we methodically gauge the influence of recurrent measurements on

the survival probability. The recently devised comprehensive data-driven learning approach leverages all available data points to compute the cumulative event risk, thereby optimizing data utilization. To ascertain the parameters, we engage in the optimization of an objective function, which involves the regularization of model parameters and the facilitation of their gradual transitions. Empirical assessments are conducted using real-world time-to-event data featuring repeated measurements. The new technique exhibits superior performance when compared to other cutting-edge survival models. We summarize the contributions as follows:

- Our method is characterized by its simplicity, both computationally and conceptually. It entails a series of logistic regression fits, and the operations can be readily comprehended within the framework of regression modeling.
- Empirical evidence attests to the effectiveness of our approach. We applied it to predict survival rates in breast cancer patients and a diverse Canadian cohort, encompassing individuals without disabilities.

2 Related Work

The essence of this paper revolves around the realm of time-to-event data coupled with the inclusion of repeated measurements. Consequently, our focus in this review shall encompass endeavors related to time-to-event prediction models and the analysis of time-varying data. Broadly delineated, methods for analyzing time-to-event data can be categorized into two primary groups: statistical methods and machine-learning-based approaches.

2.1 Statistical Methods

Statistical methods concentrate on the intricacies of event time distributions and parameter estimation. These methods can be categorized into parametric, nonparametric, and semi-parametric approaches. Nonparametric methods empirically estimate the survival probability through Nelson-Aalen estimator [1] and Kaplan-Meier estimator [11]. While they yield unbiased descriptive insights, they lack the capability to evaluate the impact of multiple risk factors on outcomes. In contrast, parametric methods posit that the time-to-event adheres to a specific distribution, such as exponential, Weibull, (log-)logistic, etc. For example, the accelerated failure time model [26] operates under the premise that the underlying time-to-event distribution has been accurately specified. To circumvent this assumption, the semi-parametric Cox model [2] has gained broader traction in time-to-event data analysis due to its user-friendliness, established efficacy, and interpretability of results [28]. This model assumes a specific influence of risk factors on events, an assumption that is frequently violated in practical scenarios.

2.2 Machine-Learning-Based Approaches

As the utilization of time-to-event data, particularly those featuring time-varying risk factors, continues to rise, machine-learning-based methods have increasingly supplanted traditional statistical models within the realm of survival analysis. This transition is particularly evident in the context of time-to-event studies, where machine learning finds synergy with survival settings. Feed-forward neural networks [5], recurrent neural networks [8,27], and deep neural networks [12,15] have all been employed for time-to-event analysis. Additionally, multi-task learning has emerged as a common strategy for time-to-event prediction. For instance, in [18], a pioneering approach involved the creation of a multi-task logistic regression model to learn patient-specific survival distributions. Gaussian processes have also made their mark in survival analysis tasks. For instance, [13] introduced scalable variational inference algorithms for a Gaussian-process-based survival model, accounting for uncertainty in hazard function modeling. Furthermore, [6] proposed a semi-parametric Bayesian model for survival analysis employing a Gaussian process. This model modulates the hazard function through the multiplication of a parametric baseline hazard and a nonparametric component.

2.3 Time-Varying Models

Recognizing the necessity to accommodate the evolving impact of factors on survival outcomes as time progresses, especially in the context of extended follow-up periods [7], there has been a notable surge of interest in the pursuit of learning time-varying coefficients rather than adhering to fixed ones. These models, characterized by varying coefficients, allow for the exploration of dynamic patterns and naturally extend the scope of classical parametric models. Their growing popularity in data analysis [4] is attributed to their robust interpretability. To estimate the time-varying coefficients within the Cox model framework, diverse strategies have emerged. [23] maximized a kernel-weighted partial likelihood, while [22] employed local empirical partial likelihood smoothing. Another approach, as seen in [19], leveraged time-varying coefficients to elucidate the evolving effects of risk factors on breast cancer failure. However, the practical applicability of the proportionality assumption may be limited, particularly when the effects of risk factors undergo temporal shifts. To address this, [27] introduced a survival recurrent neural network, which yielded superior predictions compared to other state-of-the-art survival models. This advancement underscores the efficacy of modern machine learning techniques in capturing the intricate dynamics of time-varying factors affecting survival outcomes.

3 Problem Statement

We identify each individual using two variables, that is, the response 'censor' $C \in \{0, 1\}$ indicating whether or not the individual time is censored and the

response 'stamp' $S \in \mathbb{R}^+$ showing the time when the event happened or the individual is lost to follow-up. If $C = 0$, we have the time of event occurrence, say T, which is uncensored. The event occurred right at the last recorded time S, we thus have $T = S$. When T is censored, $C = 1$. For this situation, S underestimates the true but unknown T, i.e., $T > S$. Table 1 gives an example: three old Canadians aged over 65. We have $T = 7$ months for the 77-year-old Québecois. $C = 1$ indicates that T is censored due to dropout or early end of follow-up. Here, every individual can be presented by D factors that are time-invariant (e.g., 'province' and 'age') or time-varying (e.g., 'depression', 'sleep', and 'fall'). Generally, these measures can be written as $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_V) \in \mathbb{R}^{D \times V}$, where, at the vth time of the V different times $\tau_1 < \cdots < \tau_V$, we have $\mathbf{x}_v = (x_{v1}, \ldots, x_{vD})^\top$. The time-invariant factor d is $x_{1d} = x_{2d} = \cdots = x_{Vd}$. We aim to identify a mapping function $f_{\mathbf{W}}$ ($\mathbf{W}$ is the vector of parameters to learn) for probability prediction at $t_1 < t_2 \cdots < t_K$.

Table 1. An example: time-to-event data and repeated measures for three aging Canadians living in Québec, Ontario, and Alberta, respectively.

Risk Factors							Response	
time-invariant			time-varying (Repeated Measures)				Censor	Stamp
Province	Age	··	Depression	··	Fall	Smoking	(C)	(S)
Québec	77	··	mild	··	no	sometimes	0	7
			mild	··	yes	sometimes		
			..	..	..	..		
			no	··	no	seldom		
Ontario	69	··	severe	··	no	often	1	11
			moderate	··	no	seldom		
			..	..	..	..		
			moderate	··	no	sometimes		
Alberta	86	··	no	··	yes	never	1	18
			no	··	no	never		
			..	..	..	..		
			mild	··	no	never		

4 Our Approach

4.1 Health Status

The initial step involves encoding the response variables (C and S) into health statuses represented by Y. The value at a specific time point t can be expressed as $y[t] = (-\mathbb{1}(C = 1))^{\mathbb{1}(S<t)}$. This function assumes the value of 1 if the event

occurred on or before time t, which can be represented as $S \geq t$. Otherwise, it takes the value of 0 when $C = 0$, or -1 if $C = 1$. Here, $\mathbb{1}(judgment)$ denotes an indicator function that equals 1 when the $judgment$ is true, and 0 otherwise.

Example: The health statuses for the three individuals are illustrated in Table 2. Each health status $y[t_k]$ denotes whether a specific event (such as disability) has occurred by time t_k: it takes on a value of 1 if the event has occurred, 0 if it hasn't, and -1 if the information is unknown. Once $y[t_k]$ transitions to "0", it remains so and does not revert back to "1". Consequently, there exist $K+1$ valid sequences following the pattern $(1, 1, \ldots, 0, 0, \ldots)$, which encompasses sequences consisting entirely of "1" s and sequences entirely of "0"s. For instance, let's consider the case of the 77-year-old Québecois: his health status remains "1" until $S = 7$, at which point it transitions to "0". In contrast, for the other two, the exact onset times are censored. Consequently, their health statuses remain "1" until the corresponding time stamp, and subsequently shift to "-1" beyond that point.

Table 2. An example of health status for three Canadian seniors aged over 60.

Health Status											Response	
$y[1]$	$\cdot\cdot$	$y[7]$	$y[8]$	$\cdot\cdot$	$y[11]$	$y[12]$	$\cdot\cdot$	$y[18]$	$y[19]$	$\cdot\cdot$	C	S
1	1	1	0	0	0	0	0	0	0	0	0	7
1	1	1	1	1	1	–1	–1	–1	–1	–1	1	11
1	1	1	1	1	1	1	1	1	–1	–1	1	18

4.2 Training

For an individual indexed as $i \in \mathcal{G}_0 = \{i | \forall i : C_i = 0\}$, possessing known health statuses denoted as $Y = (y[t_1], \ldots, y[t_K])$ at time instances $t_1 < \ldots < t_K$, along with associated measurements $\mathbf{X}$ $(\tau_V \leq t_K)$, we estimate the probability of observing Y through the utilization of generalized logistic regression.

$$\begin{aligned} \Pr\left(Y \mid \mathbf{X}; \mathbf{W}\right)_0 &= \frac{\exp\left(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} \cdot \mathbb{1}(\mathbf{y} \leq 0)\right)}{\mathbf{exp}\left(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A}\right) \cdot \mathbf{1}} \\ \mathbf{W} &= (\mathbf{w}_1, \ldots, \mathbf{w}_V) \in \mathbb{R}^{D \times V} \\ \mathbf{w}_v &= (w_{v1}, \ldots, w_{vD})^\top \in \mathbb{R}^{D \times 1}, \forall v = 1, 2, \ldots, V \\ \mathbf{W} * \mathbf{X} &= (\mathbf{w}_1 \cdot \mathbf{x}_1, \ldots, \mathbf{w}_V \cdot \mathbf{x}_V) \in \mathbb{R}^{1 \times V}. \end{aligned} \tag{1}$$

The regression coefficient $\mathbf{W}$ serves to quantify the influence of factors and their recurrent measurements on the probability of an individual's event-free status. In this context, the coefficients $\mathbf{w}_v$ signify the contribution of measurements

denoted by D at the specific time τ_v. The summation of transformed measurements over V time instances is computed through a column-wise Hadamard product [20]. To assess the impact of the D factors and their V repeated measurements on the manifestation of Y, we introduce the concept of the time-decay ratio. This ratio is shaped by the passage of time and can be expressed as follows:

$$\boldsymbol{\Delta} = \mathbf{exp}(\delta(k, v)) \in \mathbb{R}^{K \times V} \tag{2}$$

$$\delta(k, v) = -(t_k - \tau_v) \times \mathbb{1}(t_k \geq \tau_v). \tag{3}$$

The symbol **exp** denotes the element-wise exponential operation applied to a matrix. The decay ratio, represented as $\exp(\sigma(k, v))$, captures the exponential of the time difference between t_k and τ_v, reflecting how the impact of a measurement at time τ_v on the probability at time t_k diminishes over time. As time passes, this impact reduces gradually. For instance, the effect of a fall-caused injury on the onset of disability would attenuate as time progresses. To delve into event-free probabilities, we employ the lower triangular identity matrix denoted as $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K) \in \mathbb{R}^{K \times K}$, where $\alpha_{ij} = 1$ if $i \geq j$ and 0 otherwise. This matrix aids us in investigating probabilities of remaining event-free. For individuals belonging to $\mathcal{G}_1 = \{i | \text{for all } i : C_i = 1\}$, who possess an unknown event time, their health statuses remain consistent prior to the given time stamp. Consequently,

$$\Pr\left(Y \mid \mathbf{X}; \mathbf{W}\right)_1 = \frac{\mathbf{exp}(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A} \cdot \mathbb{1}(\mathbf{y} \leq 0)) \cdot \mathbf{1}}{\mathbf{exp}\left(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A}\right) \cdot \mathbf{1}}. \tag{4}$$

The numerator represents the accumulation of the risks associated with the occurrence of target responses. To learn the matrix $\mathbf{W}$, we undertake minimization of the negative logarithm of the likelihood across all individuals through an expectation-maximization process [3]. This involves suitable initialization and can be outlined as follows:

$$\min_{\mathbf{W}} \; P(\mathbf{W}) - \sum\nolimits_{i \in \mathcal{G}_0} \log(\Pr(Y_i | \mathbf{X}_i; \mathbf{W})_0) - \sum\nolimits_{i \in \mathcal{G}_1} \log(\Pr(Y_i | \mathbf{X}_i; \mathbf{W})_1), \tag{5}$$

The incorporation of the elastic-net penalty $P(\mathbf{W}) = \lambda_1 |\mathbf{W}|_1 + \lambda_2 |\mathbf{W}|_2^2$ into the loss function enhances its strong convexity, resulting in a unique minimum. To establish suitable values for the hyperparameters λ_1 and λ_2, we rely on an independent validation set. The learning objective function exhibits strict convexity within the feasible domain $\mathbf{w}_k \in \mathbb{R}^D, \forall k$, ensuring a unique globally optimal solution.

4.3 Prediction

Assuming the availability of known measurements $\mathbf{X}'$ for a new individual at time instances $\tau_1 < \cdots < \tau_{V'}$ before time t_k, and utilizing the learned parameters $\widehat{\mathbf{W}} \in \mathbb{R}^{D \times V}$, the prediction of the event-free probability at time t_k effectively amounts to predicting the health statuses $Y'[t_k] = (1, 1, \ldots, 1) \in \mathbb{R}^{1 \times k}$. This

probability, denoted as $\Pr\left(Y'[t_k] \mid \mathbf{X}'; \widehat{\mathbf{W}}\right)$, is amenable to estimation through the regression model depicted in Eq. 1. (In the subsequent discussion, the notation involving a '′' signifies that it is redefined for the test set.)

4.4 Estimate of Time-to-Event

Upon obtaining the predicted survival probabilities for different time points from our model, the subsequent task entails estimating the time-to-event for the new individual. To initiate this process, we establish the relative error at time t using the following definition: $E(t) = \sum_{k=1}^{K} (|\log t - \log t_k|) \Pr\left(Y'[t_k] \mid \mathbf{X}'; \widehat{\mathbf{W}}\right)$, thereby seeking the time at which E is the lowest, i.e., the time-to-event is $\widehat{T} = \text{argmin}_{t \in \{t_1, \ldots, t_K\}} E(t)$.

5 Experiments

5.1 Data

The dataset comprises the Breast Cancer dataset sourced from the prognostic data repository provided by the PCoE of NASA Ames, in addition to the Aging Canada dataset extracted from the Canadian Community Health Survey (CCHS) statistical surveys. In the context of the Breast Cancer dataset, the times-to-event were computed by subtracting the date of diagnosis from the date of last contact, which corresponds to the study cutoff. As for the Aging Canada dataset, our focus was on respondents surveyed between 2009 and 2010. This dataset centers on the health of individuals aged 45 and above, probing various factors influencing healthy aging. Over 2,000 valid interviews covering the population residing in all ten provinces were included. Regarding the factors under consideration, we engaged in pairwise correlation analysis between independent factors via a correlation matrix. If the Pearson correlation coefficient [14] exceeded 0.75 for two factors, we made the decision to retain the more pertinent one and discard the other. This process yielded 16 factors for the Breast Cancer dataset and 24 for the Aging Canada dataset. For the categorization of factors into numerical representations (with the exception of the two existing numerical factors: age and BMI), we employed Softmax normalization. For a comprehensive overview of the data statistics, please refer to Table 3.

Table 3. Statistics of the two lifetime datasets, such as data size (number of individuals) N, dimensionality (number of risk factors) V, censoring rate (number of individuals with censored time-to-event) C, missing-value percentage M, and event of interest

Data	N	V	C	M	Event
Breast Cancer	2,165	12	23.3%	12.7%	death
Aging Canada	4,470	24	42.6%	27.5%	disability

5.2 Baselines

During the experimental phase, we executed a comparative analysis involving our method and nine distinct time-to-event prediction models. Cox model with elastic-net regularization (Cox-EN) [31]: This model optimizes regression coefficients through the incorporation of elastic-net regularization. Cox model with neural network (CoxNN) [5]: CoxNN departs from the traditional linear exponent of the Cox hazard by introducing a nonlinear artificial neural network output. Accelerated failure time (AFT) model [26]: AFT assumes a Weibull survival distribution to model survival times. Random survival forest (RSF) [9]: RSF aggregates tree-based Nelson-Aalen estimators to derive estimates for the conditional cumulative failure hazard. Multi-task logistic regression (MTLR) [18]: MTLR models the survival distribution by employing multi-task logistic regression in a dependent manner, with the regularization parameter determined via an additional 10-fold cross-validation (10CV). Accumulative-hazard-based joint likelihood (AHJ) model [28]: AHJ captures the connection between survival probability and time-varying factors of longitudinal data in a succinct yet potent manner. DeepSurv [12]: DeepSurv is a deep learning extension of the Cox proportional hazards model. DeepHit [15]: DeepHit employs a deep neural network to directly learn the distribution of survival times. It avoids presumptions about the underlying stochastic process, allowing for the possibility of evolving relationships between risk factors and risks over time. Survival neural network (SNN) [27]: SNN computes binary classification scores at fixed time intervals to estimate survival outcomes. For all models, the hyperparameters were optimized using a validation dataset. The chosen settings are those that yield optimal outcomes as per the respective work's findings.

5.3 Performance Evaluation Metrics

To assess the predictive efficacy, we employed three distinct evaluation metrics: the area under the ROC curve (AUC), the concordance index (C-index), and the Brier score (BS), which we adapted to align with our specific context.

AUC. AUC measures the predictive ability of the model at a particular time. It qualifies the ability of a model to address questions such as "Whether individual i would be likely to die one year after index discharge?" It can be defined as

$$\text{AUC} = \frac{1}{|\mathcal{G}_0'| \times |\mathcal{G}_1'|} \sum_{i \in \mathcal{G}_0'} \sum_{j \in \mathcal{G}_1'} \mathbb{1}\left(\Pr(Y_i'[\mathcal{T}^*]|\mathbf{X}_i) < \Pr(Y_j'[\mathcal{T}^*]|\mathbf{X}_j)\right).$$

C-Index. C-index is a generalization of the concept of AUC. It measures how accurately a model can answer the questions such as "Which one of the two patients i and j is more likely to die or be rehospitalized?" We define C-index as

$$\text{C-index} = \frac{1}{|\mathcal{P}|} \sum_{i,j \in \mathcal{P}} \mathbb{1}\left(\Pr(Y_i'[T_i]|\mathbf{X}_i) \leq \Pr(Y_j'[S_j]|\mathbf{X}_j)\right),$$

where $\mathcal{P} = \{(i,j)|\forall i,j : T_i \leq S_j\}$ is the number of comparable pairs of patients. (Refer to [28] for more details regarding this metric.) Similar to AUC, C-index takes values from 0.5 (completely random) to 1.0 (perfect prediction).

BS. Precisely, the Brier score (BS) operates as a mean squared error for time-to-event predictions, serving as an indicator of the caliber of survival probability predictions – in essence, quantifying prediction accuracy. The Brier score allows for an assessment of the model's capability to address queries such as "How precise is the prognosis regarding the recurrence of the disease for patient i?" This measure is computed as an overall error gauge spanning all the time points, enabling a comprehensive evaluation of predictive accuracy, as follows:

$$\mathrm{BS} = \frac{1}{|\mathcal{G}_0' \cup \mathcal{G}_1'|} \sum\nolimits_{i \in \mathcal{G}_0' \cup \mathcal{G}_1'} \left(C_i - \Pr(Y_i'[S_i]|\mathbf{X}_i)\right)^2,$$

where ϵ_i represents the event indicator for individual i, taking the value of 1 when an event occurs and 0 otherwise. Notably, the Brier score (BS) is constrained to fall within the interval $[0, 1]$. A smaller BS value corresponds to a heightened precision in prognostication, indicating greater accuracy in the predictions.

RE for Survival Time. In order to assess the accuracy of the predicted onset time, we introduce the concept of relative error (RE), which quantifies the disparity between the estimated time $\widehat{T}$ and the actual ground truth T for all individuals in the test set who have experienced an event:

$$\mathrm{RE} = \frac{1}{|\mathcal{G}_0'|} \sum\nolimits_{\forall i \in \mathcal{G}_0'} \min\left\{\left|(\widehat{T}_i - T_i)/T_i\right|, 1\right\}.$$

Here, a smaller RE means a more accurate estimate of time-to-event.

5.4 Results and Discussion

Comparison of Time-to-Event Prediction. Table 4 presents the 10 cross-validation results on the two datasets. Our model outperforms all the other models but the relative error of the predicted time-to-event on the Aging Canada test set. SSP achieves not only accurate predictions of survival probability but accurate estimates of time-to-event, where the RE of the survival estimates for cancer patients is the lowest and for aging people is the second lowest (only inferior to AHJ). In most cases, SNN is the second best, e.g., it achieves the second-best C-index, BS, and RE for the time-to-event estimate on the Breast Cancer dataset, and the second-best AUC and BS on the Aging Canada dataset. Compared to SNN, we can see that SSP achieves an over 2% AUC improvement on Breast Cancer data and 1% on Aging Canada data. Meantime, it yields accurate results with an over 2% C-index improvement, while achieving a lower BS and more accurate time-to-event prediction. The comparison between SSP and MTLR demonstrates that our approach designed for repeated measures

performs more effectively. ML-based models perform better than the statistic models, such as Cox-EN and AFT, revealing the strong processing capabilities and applicability of ML-based models to deal with complex time-to-event data. The performance of Cox-EN, CoxNN, and AHJ, is not as good as SSN, MTLR, and SSP; the most possible reason is that these three models assume that Cox's proportional risk meets. Although AHJ considers the change in the risk of the event, it does not perform better than SSP. This is mainly because of its proportional hazard assumptions (N.b.: AHJ is a typical Cox-based model). The performance of the parametric model AFT on two data sets is the worst since the Weibull distribution assumption does not fit most of the time distribution well. The models (e.g., SSP, SNN, and AHJ) that are specifically designed for handling time-varying risk factors perform better, in comparison with those non-time-varying models.

Table 4. Comparison of the 10 cross-validation (10CV) results, and RE of the estimated survival times, on the test data, in the form of the mean (standard deviation). The best results are in **bold** and the second-best performances are underlined.

Dataset	Model	AUC	C-index	BS	RE
Breast Cancer	Cox-EN	.676(.047)	.695(.027)	.283(.033)	.473(.068)
	CoxNN	.693(.034)	.673(.028)	.313(.025)	.542(.073)
	AFT	.670(.016)	.651(.056)	.260(.038)	.450(.062)
	RSF	.687(.021)	.682(.034)	.274(.036)	.410(.094)
	MTLR	735(.041)	.701(.029)	.264(.020)	.392(.085)
	AHJ	.712(.024)	.682(.017)	.302(.012)	.325(.073)
	DeepSurv	.673(.031)	.669(.028)	.372(.025)	.583(.059)
	DeepHit	.715(.014)	.677(.035)	.278(.029)	.339(.061)
	SNN	.724(.035)	.739(.030)	.196(.021)	.311(.053)
	SSP	**.757**(.022)	**.763**(.029)	**.192**(.019)	**.308**(.058)
Aging Canada	Cox-EN	.671(.023)	.682(.020)	.372(.023)	.420(.053)
	CoxNN	.674(.049)	.707(.013)	.341(.023)	.413(.084)
	AFT	.632(.027)	.665(.022)	.282(.019)	.439(.082)
	RSF	.634(.033)	.643(.044)	.302(.033)	.475(.075)
	MTLR	.742(.027)	.733(.019)	.298(.043)	.346(.082)
	AHJ	.729(.025)	.738(.023)	.251(.015)	**.334**(.064)
	DeepSurv	.703(.031)	.695(.027)	.299(.022)	384(.079)
	DeepHit	.698(.032)	.725(.031)	.291(.033)	.429(.058)
	SNN	.772(.018)	.723(.024)	.234(.041)	.369(.064)
	SSP	**.781**(.019)	**.743**(.018)	**.223**(.028)	.341(.073)

Predicted Survival Probability. Figure 1 illustrates the average predicted survival probabilities for all individuals across a 12-month period. Observing the graph, it becomes evident that the survival probabilities generated by our model (SSP) markedly diverge from the predictions of the other models. Notably, our model consistently yields lower probabilities at each time point (with one time point corresponding to each month). The survival curves produced by SNN, AHJ, and MTLR appear closely aligned, with their probabilities demonstrating significant discrepancies from the other six models (Cox-EN, CoxNN, AFT, RSF, DeepSurv, and DeepHit), which yield comparatively higher probabilities. While these curves don't conclusively indicate the superior model, the predictions provided by SSP, SNN, AHJ, and MTLR effectively facilitate the differentiation between individuals at high risk and those at low risk. Regarding the Aging Canada dataset, most models generate closely aligned monthly average survival probabilities. Notably, both SSP and SNN yield notably lower survival probabilities (dipping below 0.7) for the last three months of the follow-up period.

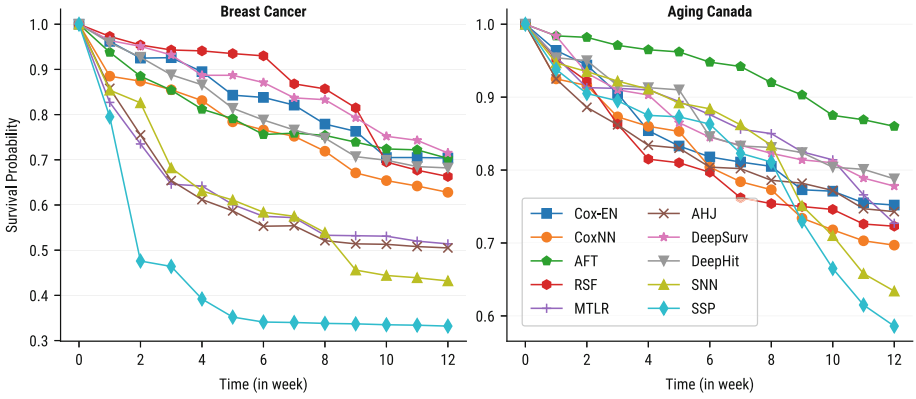

Fig. 1. Comparison of the average of predicted survival probability

Case Study. Figure 2 showcases the predicted survival probabilities for two distinct patients: one categorized as high risk and the other as low risk. (Note: In this context, high-risk patients are those who succumbed during the follow-up period, while low-risk patients remained unaffected.) Here are the particulars of the patients: 1) The high-risk is a 58-year-old female who was discharged from the hospital on July 10, 2010, following a 13-day hospital stay. Tragically, she passed away due to breast cancer 185 d after her discharge. Notably, there were no instances of rehospitalization prior to her demise; 2) The low-risk is a 75-year-old male discharged from the hospital on April 2, 2012, after 23 d of stay. He was still alive by the end of the 1-year follow-up period and had never been rehospitalized before his death. Only the SNN, MTLR, and SSP models are able

to distinctly differentiate between these two patients. Notably, the SSP model manifests the most pronounced distinction between the two cases. Notably, the SSP model is capable of projecting an exceedingly low survival probability for the high-risk patient at the 185-day mark. This functionality holds immense value as it can enable timely warnings and provision of advice regarding early interventions for high-risk patients.

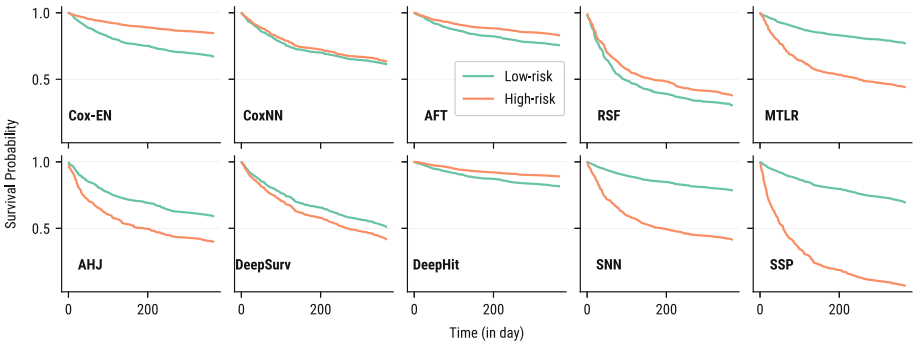

Fig. 2. Comparison of the predicted survival probability for two patients

Ablation Study. Given that SSP amalgamates several crucial components, including censoring likelihood (see Eq. 4), time-sensitive risk considerations (as depicted in Eq. 2), and regularization within the learning objective, we undertake exhaustive ablation studies to dissect the contributions of distinct components. Four distinct variants are considered for analysis: 1) SSP-censor disregards the censoring likelihood and employs only Eq. 1 to determine model parameters; 2) SSP-decay omits the time-dependent decay in cumulative risk computation, with $\eta(u, \tau)$ set to 1 in Eq. 2; 3) SSP-static neglects repeated measures, utilizing Cox proportional hazard instead of Eq. 2; 4) SSP-reg: In this variant, regularization is excluded, rendering $P(\mathbf{W}) = 0$ in the learning objective function as presented in Eq. 5. The results (see Fig. 3) consistently underscore the supremacy of SSP relative to all the variants. Notably, the improvement of SSP over SSP-censor underscores the enhancement in predictive capacity gained by estimating the censoring likelihood. The superiority of SSP in contrast to SSP-decay underscores the effectiveness of the time-decay setting. Furthermore, the comparison between SSP and SSP-static accentuates the significance of incorporating all repeated measures. In comparison with SSP-reg, SSP exhibits higher AUC and C-index values, illustrating the potential efficacy of regularization.

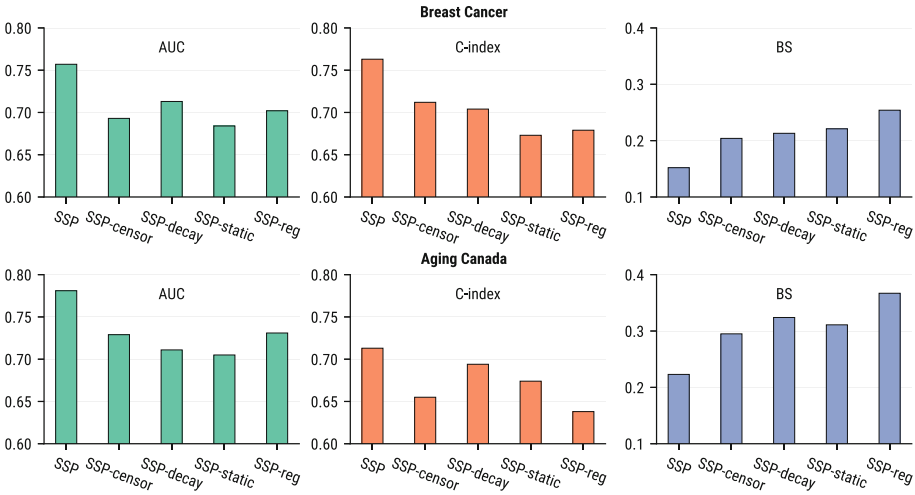

Fig. 3. Comparison of variants' performance

6 Conclusion

This paper proposes a new repeated measure modeling method that quantifies the impact of measures on health status and time-to-event. The new method optimizes the parameters on full data and the regularization approach. The model learning is performed to transform survival prediction into multiple logistic regression learning tasks at different time points. This approach eliminates the need for any assumptions regarding the distribution of unknown data. The outcomes of comparative experiments conducted on real-world data substantiate that the proposed method surpasses state-of-the-art models. It demonstrates superior performance and efficacy in predicting both survival probability and time-to-event outcomes.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Discovery Grant RGPIN-2020-07110 and Discovery Accelerator Supplements Grant RGPAS-2020-00089, and the National Natural Science Foundation of China (NSFC) under Grant No. U1805263.

References

1. Aalen, O.: Nonparametric estimation of partial transition probabilities in multiple decrement models. Ann. Stat., 534–545 (1978)
2. Cox, D.R.: Regression models and life tables. J. R. Stat. Soc. Ser. B Stat. Methodol. **34**, 187–220 (1972)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol., 1–38 (1977)
4. Fan, J., Zhang, W.: Statistical methods with varying coefficient models. Stat. Interface **1**(1), 179 (2008)

5. Faraggi, D., Simon, R.: A neural network model for survival data. Stat. Med. **14**(1), 73–82 (1995)
6. Fernández, T., Rivera, N., Teh, Y.W.: Gaussian processes for survival analysis. In: NeurIPS, pp. 5021–5029 (2016)
7. Fisher, L.D., Lin, D.Y.: Time-dependent covariates in the Cox proportional-hazards regression model. Annu. Rev. Public Health **20**(1), 145–157 (1999)
8. Giunchiglia, E., Nemchenko, A., van der Schaar, M.: RNN-SURV: a deep recurrent model for survival analysis. In: ICANN, pp. 23–32 (2018)
9. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests. Ann. Appl. Stat., 841–860 (2008)
10. Jenkins, S.P.: Survival analysis unpublished Manuscript, Institute for Social and Economic Research, Chapter 3, University of Essex, Colchester, UK (2005)
11. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. **53**(282), 457–481 (1958)
12. Katzman, J., Shaham, U., Bates, J., Cloninger, A., Jiang, T., Kluger, Y.: DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med. Res. Methodol., 18–24 (2018)
13. Kim, M., Pavlovic, V.: Variational inference for gaussian process models for survival analysis. In: UAI, pp. 435–445 (2018)
14. Kirch, W. (ed.): Pearson's Correlation Coefficient, pp. 1090–1091. Springer, Netherland(2008). https://doi.org/10.1007/978-1-4020-5614-7_2569
15. Lee, C., Zame, W.R., Yoon, J., van der Schaar, M.: Deephit: a deep learning approach to survival analysis with competing risks. In: AAAI, pp. 2314–2321 (2018)
16. Lee, E.T., Wang, J.W.: Statistical Methods for Survival Data Analysis. John Wiley & Sons, 4th edn. (2013)
17. Li, Y., Wang, L., Wang, J., Ye, J., Reddy, C.K.: Transfer learning for survival analysis via efficient l2,1-norm regularized Cox regression. In: ICDM, pp. 231–240 (2017)
18. Lin, H.c., Baracos, V., Greiner, R., Chun-nam, J.Y.: Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: NeurIPS, pp. 1845–1853 (2011)
19. Liu, M., Lu, W., Shore, R.E., Zeleniuch-Jacquotte, A.: Cox regression model with time-varying coefficients in nested case - control studies. Biostatistics **11**(4), 693–706 (2010)
20. Liu, S., Trenkler, G.: Hadamard, khatri-rao, kronecker and other matrix products. Int. J. Inf. Syst. Sci. **4**(1), 160–177 (2008)
21. Moghaddass, R., Rudin, C.: The latent state hazard model, with application to wind turbine reliability. Ann. Appl. Stat. **9**(4), 1823–1863 (2014)
22. Sun, Y., Sundaram, R., Zhao, Y.: Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood. Scand. J. Stat. **36**(3), 444–462 (2009)
23. Tian, L., Zucker, D., Wei, L.: On the Cox model with time-varying regression coefficients. J. Amer. Statist. Assoc. **100**(469), 172–183 (2005)
24. Vinzamuri, B., Li, Y., Reddy, C.K.: Active learning based survival regression for censored data. In: CIKM, pp. 241–250 (2014)
25. Wang, P., Li, Y., Reddy, C.K.: Machine learning for survival analysis: A survey. ACM Comput. Surv. **51**(6), 1–36 (2019)
26. Wei, L.J.: The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Stat. Med. **11**(14–15), 1871–1879 (1992)
27. Zhang, J., Wang, S., Chen, L., Guo, G., Vanasse, A.: Time-dependent survival neural network for remaining useful life prediction. In: PAKDD, pp. 441–452 (2019)

28. Zhang, J., Wang, S., Courteau, J., Chen, L., Bach, A., Vanasse, A.: Predicting COPD failure by modeling hazard in longitudinal clinical data. In: ICDM, pp. 639–648 (2016)
29. Zhang, Z., Reinikainen, J., Adeleke, K.A., Pieterse, M.E., Groothuis-Oudshoorn, C.G.: Time-varying covariates and coefficients in Cox regression models. Annals Trans, Med. **6**(7), 121 (2018)
30. Zhou, M.: Understanding the Cox regression models with time-change covariates. Am. Stat. **55**(2), 153–155 (2001)
31. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. **67**(2), 301–320 (2005)