

## Article

# Dynamic Risk Prediction via a Joint Frailty-Copula Model and IPD Meta-Analysis: Building Web Applications

Takeshi Emura <sup>1,2,\*</sup>, Hirofumi Michimae <sup>3</sup>  and Shigeyuki Matsui <sup>2,4</sup><sup>1</sup> Biostatistics Center, Kurume University, Kurume 830-0011, Japan<sup>2</sup> Research Center for Medical and Health Data Science, The Institute of Statistical Mathematics, Tokyo 190-8562, Japan; smatsui@med.nagoya-u.ac.jp<sup>3</sup> Department of Clinical Medicine (Biostatistics), School of Pharmacy, Kitasato University, Tokyo 108-8641, Japan; michimae@pharm.kitasato-u.ac.jp<sup>4</sup> Department of Biostatistics, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan

\* Correspondence: takeshiemura@gmail.com

**Abstract:** Clinical risk prediction formulas for cancer patients can be improved by dynamically updating the formulas by intermediate events, such as tumor progression. The increased accessibility of individual patient data (IPD) from multiple studies has motivated the development of dynamic prediction formulas accounting for between-study heterogeneity. A joint frailty-copula model for overall survival and time to tumor progression has the potential to develop a dynamic prediction formula of death from heterogeneous studies. However, the process of developing, validating, and publishing the prediction formula is complex, which has not been sufficiently described in the literature. In this article, we provide a tutorial in order to build a web-based application for dynamic risk prediction for cancer patients on the basis of the R packages *joint.Cox* and *Shiny*. We demonstrate the proposed methods using a dataset of breast cancer patients from multiple clinical studies. Following this tutorial, we demonstrate how one can publish web applications available online, which can be manipulated by any user through a smartphone or personal computer. After learning this tutorial, developers acquire the ability to build an online web application using their own datasets.

**Keywords:** clustered data; copula; frailty model; meta-analysis; risk prediction; Shiny; survival analysis



**Citation:** Emura, T.; Michimae, H.; Matsui, S. Dynamic Risk Prediction via a Joint Frailty-Copula Model and IPD Meta-Analysis: Building Web Applications. *Entropy* **2022**, *24*, 589. <https://doi.org/10.3390/e24050589>

Academic Editor: Raúl Alcaraz

Received: 4 March 2022

Accepted: 21 April 2022

Published: 22 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Clinicians often wish to predict the risk of death for their patients on the basis of patient-level information such as clinicopathological status and tumor gene expressions. A variety of statistical and machine learning methods have been developed for clinical risk prediction in patients with breast cancer [1–6], gastric cancer [7], ovarian cancer [8–10], lymphoma [11–13], head and neck cancer [14], and mixed cancer types [15–17]. The prediction methods for clinicians help in contributing to the development of personalized medicine by allowing patients to consider their future and clinicians to choose an optimal therapy.

Traditional prediction methods are designed to predict survival from the date of the initial treatment (e.g., the first-line treatment). The classical Cox regression model [18,19] and many other methods have been applied to predict the risk of death from the initial treatment (e.g., 5-year survival). However, a patient could ask about the risk of death in case they develop (or do not develop) metastasis after the initial treatment [20]. Clinicians may need *dynamic* prediction results, where the prediction time can be set arbitrarily, e.g., one year after treatment [20–22]. Dynamic prediction is informative for clinicians regarding the subsequent stages of treatments (e.g., the timing of second-line treatment or the termination of treatment) [22,23]. The use of intermediate event information has been suggested to construct a risk prediction scheme using dynamic prediction models [10,20–24].

For a prediction model to be clinically valid, the model needs to follow some guidelines such as the TRIPOD statement [25]. Due to the increased accessibility of datasets of individual patient data (IPD) from multiple studies, different countries, and diverse ethnic groups, the TRIPOD statement is on the way to an extension (<https://www.tripod-statement.org/clustered/> accessed on 1 April 2022) [26].

A *joint frailty-copula model* for IPD meta-analyses [9] has demonstrated its ability to develop both clinical prediction and surrogate validation models from heterogeneous studies [10]. However, the implementation remains technically complex. For clinicians to employ the abilities of the joint frailty-copula model, it is better to implement a web calculator for predicting survival. In addition, there has been a strong demand for IPD meta-analyses in order to validate surrogate endpoints [27–35], which is related to the prediction of overall survival (the true endpoint) by its surrogate. In any case, statistical methods for IPD meta-analyses have to account for between-study heterogeneity in the prediction [10,26], estimation [30–33,36–38], and surrogate validation [30,32–35].

This article presents a template for developing a web application for dynamic risk prediction based on a joint frailty-copula model and IPD meta-analysis. The main computational tools are the R packages *joint.Cox* [39] and *Shiny* [40]. To bridge developers and users, we effectively connected the development of a complex dynamic prediction scheme and its implementation to a patient's prognosis. We provide a Shiny-based template for transforming a prediction model to a web application. A dataset for breast cancer was used for illustration.

After learning the proposed template, developers have the ability to make an online web application using their own dataset. The resultant applications are accessible to clinicians without knowledge of R and can be manipulated by any user through a smartphone or personal computer.

This article is organized as follows. Section 2 reviews the background of this research. Section 3 proposes validation methods for dynamic prediction. Section 4 proposes a template for developing a web application with a step-by-step tutorial. Section 5 reports the results of developing a web-based risk prediction tool following the proposed template with a dataset of breast cancer patients. Section 6 concludes with the discussion.

## 2. Background

### 2.1. Review on Dynamic Prediction

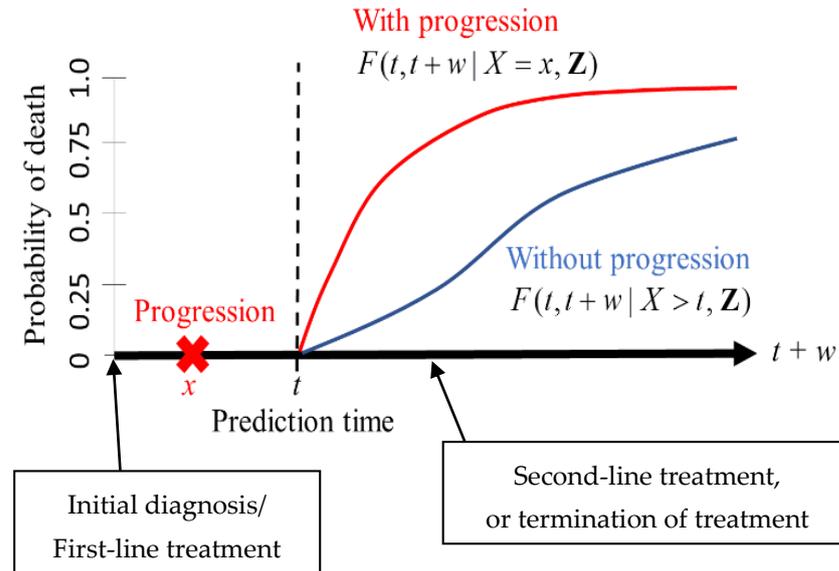
We adopted the framework of *dynamic prediction* so that a clinician can set a prediction time point for a patient after the initial diagnosis (or the first-line treatment) [10,20–24]. For instance, a clinician can assess the risk of death for a patient who remains disease-free in the next  $t$  years after the initial diagnosis or a patient who develops tumor progression within the next  $t$  years after the diagnosis (Figure 1). The risk assessment at a certain point of time may be informative for clinicians to make decisions on second-line treatments or on the termination of treatment.

The measure of prediction is the conditional probability of death between  $t$  and  $t + w$  given the observed status of a patient at time  $t$  (Figure 1). For instance, one can set  $t = 2$  (years) and  $w = 5$  (years) when predicting the risk of death within five years for a patient who has survived two years. An important feature of dynamic prediction is the flexibility of choosing a prediction time  $t$  at which clinicians can utilize the observed tumor progression status before time  $t$  (Figure 1). A statistical model has to be imposed to derive the mathematical expressions of the conditionally predicted probabilities; the details are presented below.

### 2.2. Review on a Joint Frailty-Copula Model

To derive a risk prediction scheme from IPD from multiple studies, we utilized a joint frailty-copula model describing the dependence between time to tumor progression (TTP) and overall survival (OS) [9,10,28–30]. The joint frailty-copula model differs from the standard joint models for longitudinal measurements [20,22]. The former is tailored

to IPD meta-analyses of multiple studies, where frailty accounts for the between-study heterogeneity, and a copula accounts for the dependence between two event times (TTP and OS). The joint frailty-copula model is especially useful for fitting IPD datasets for validating surrogate endpoints of OS in the meta-analytic framework [28–30].



**Figure 1.** A risk prediction scheme in the framework of dynamic prediction. The measure of prediction is the conditional probability of death between  $t$  and  $t + w$  given the observed status of a patient at time  $t$ . The expressions for  $F$  are defined in Section 2.3.

An IPD meta-analysis provides a dataset from  $G$  studies with the  $i$ -th study containing  $N_i$  patients for  $i = 1, 2, \dots, G$ . Let  $X_{ij}$  be TTP and  $D_{ij}$  be OS for  $i = 1, 2, \dots, G$  and  $j = 1, 2, \dots, N_i$ . Additionally, let  $\mathbf{Z}_{1ij}$  and  $\mathbf{Z}_{2ij}$  be covariates for TTP and OS, respectively. Both  $X_{ij}$  and  $D_{ij}$  can be censored administratively at censoring time  $C_{ij}$ ;  $X_{ij}$  can also be censored by death at time  $D_{ij}$ . What we observe in the dataset are the first-occurring event time  $T_{ij} = \min(X_{ij}, D_{ij}, C_{ij})$ , the status of tumor progression  $\delta_{ij} = I(T_{ij} = X_{ij})$ , the terminal event time  $T_{ij}^* = \min(D_{ij}, C_{ij})$ , and the status for death  $\delta_{ij}^* = I(T_{ij}^* = D_{ij})$ , where  $I(\cdot)$  is the indicator function. Thus, the observed data are expressed as  $(T_{ij}, T_{ij}^*, \delta_{ij}, \delta_{ij}^*, \mathbf{Z}_{1ij}, \mathbf{Z}_{2ij})$  for  $i = 1, 2, \dots, G$  and  $j = 1, 2, \dots, N_i$ .

To model between-study heterogeneity, we introduced “frailty terms”, defined by unobserved positive numbers:  $u_i, i = 1, 2, \dots, G$  [9,10,27–34]. All the patients within the  $i$ -th study are assumed to have the common risk value  $u_i$ . Conditional on the value  $u_i$ , the risk of each patient is defined by the hazard function for TTP and OS, denoted, respectively, as  $r_{ij}(t|u_i, \mathbf{Z}_{1ij}) = P(X_{ij} \leq t + dt|X_{ij} \geq t, u_i, \mathbf{Z}_{1ij})/dt$  and  $\lambda_{ij}(t|u_i, \mathbf{Z}_{2ij}) = P(D_{ij} \leq t + dt|D_{ij} \geq t, u_i, \mathbf{Z}_{2ij})/dt$ . Additionally, conditional dependence between OS and TTP is modeled by the joint survival function,  $P(X_{ij} > x, D_{ij} > y|u_i, \mathbf{Z}_{2ij})$ .

The joint frailty-copula model [9] specifies hazard functions and the joint survival function in the following form:

$$\begin{cases} r_{ij}(t|u_i, \mathbf{Z}_{1ij}) = u_i r_0(t) \exp(\beta'_1 \mathbf{Z}_{1ij}); & \text{the hazard function for TTP,} \\ \lambda_{ij}(t|u_i, \mathbf{Z}_{2ij}) = u_i^\alpha \lambda_0(t) \exp(\beta'_2 \mathbf{Z}_{2ij}); & \text{the hazard function for OS,} \\ C_\theta[\exp\{-u_i R_0(x) \exp(\beta'_1 \mathbf{Z}_{1ij})\}, \exp\{-u_i^\alpha \Lambda_0(y) \exp(\beta'_2 \mathbf{Z}_{2ij})\}] & \text{the joint survival function,} \end{cases}$$

where  $\beta.$  is a vector of regression coefficients,  $(r_0(\cdot), \lambda_0(\cdot))$  are the baseline hazard functions,  $(R_0(\cdot), \Lambda_0(\cdot))$  are the cumulative baseline hazard functions, and  $C_\theta$  is a copula with parameter  $\theta$ . The frailty term  $u_i > 0$  represents the  $i$ -th study’s baseline risk not explained by patient-level covariates, assumed to be gamma-distributed with  $E(u) = 1$  and  $Var(u) = \eta > 0$ , where  $\eta$  represents the degree of heterogeneity [9,27–30]. The parameter

$\alpha \in (-\infty, \infty)$  differentiates the effects of  $u$  between TTP and OS. Special cases are the null effect on OS ( $\alpha = 0$ ) and the shared effects between TTP and OS ( $\alpha = 1$ ).

For computational simplicity, we particularly adopted the Clayton copula

$$C_\theta(v, w) = \left( v^{-\theta} + w^{-\theta} - 1 \right)^{-1/\theta}, \quad \theta > 0,$$

where  $\theta$  specifies the dependence between TTP and OS [9,28,29] and can be rescaled into Kendall's tau:  $\tau = \theta / (\theta + 2)$ . Under the Clayton copula, the presence of tumor progression leads to a  $(\theta + 1)$ -fold higher risk of death [10,28] than the absence of it. Thus,  $(\theta + 1)$  is a relative risk parameter.

All the parameters in the joint frailty-copula model must be estimated by fitting a training dataset. The main advantage of IPD meta-analyses with the joint frailty-copula model is the ability to accommodate a large number of patients collected across different counties and ethnic groups. A model developed from mixed populations is often more precise than a model from a single population [26].

Based on the observed data, one can obtain the parameter estimates  $(\hat{\alpha}, \hat{\theta}, \hat{\eta}, \hat{\beta}_1, \hat{\beta}_2, \hat{r}_0, \hat{\lambda}_0)$  under the Clayton copula through the *joint.Cox* R package [28]. This package implements a penalized likelihood estimation method using the five-parameter spline for the baseline hazard functions:

$$\hat{r}_0(t) = \sum_{\ell=1}^5 \hat{g}_\ell M_\ell(t), \quad \hat{\lambda}_0(t) = \sum_{\ell=1}^5 \hat{h}_\ell M_\ell(t),$$

where  $M_1(t), \dots, M_5(t)$  are the M-spline basis functions defined by [9,28], and  $\hat{g}_\ell > 0$  and  $\hat{h}_\ell > 0$  are estimated parameters.

### 2.3. Dynamic Prediction under the Joint Frailty-Copula Model

Dynamic prediction is formulated by the *conditional failure function* [21]. This function is the conditional probability of death between  $t$  and  $t + w$  given that a patient remains alive at  $t$ . Under the joint frailty-copula model, the conditional failure function separates into two cases [10]: First, if a patient does not experience tumor progression before  $t$ , the prediction formula is defined as

$$F(t, t + w | X > t, \mathbf{Z}) \equiv P(D \leq t + w | D > t, X > t, \mathbf{Z}),$$

where  $\mathbf{Z} \equiv (\mathbf{Z}_1, \mathbf{Z}_2)$  are covariates for a patient. Second, if a patient experiences tumor progression before  $t$ , the prediction formula is defined as

$$F(t, t + w | X = x, \mathbf{Z}) \equiv P(D \leq t + w | D > t, X = x, \mathbf{Z}), \quad x \leq t.$$

The two formulas may be combined into one formula by writing  $F(t, t + w | H(t, X), \mathbf{Z})$ , where  $H(t, X)$  represents the tumor status of the patient.

The computation of  $F(t, t + w | \cdot)$  proceeds as follows: First, we fit a training dataset to the joint frailty-copula model to obtain the parameter estimates  $(\hat{\alpha}, \hat{\theta}, \hat{\eta}, \hat{\beta}_1, \hat{\beta}_2, \hat{r}_0, \hat{\lambda}_0)$ . Next, we compute the survival functions  $\hat{S}_X(t|u) = \exp\{-u\hat{R}_0(t) \exp(\hat{\beta}'_1 \mathbf{Z}_1)\}$  and  $\hat{S}_D(t|u) = \exp\{-u^\alpha \hat{\Lambda}_0(t) \exp(\hat{\beta}'_2 \mathbf{Z}_2)\}$ . Finally, we compute

$$\begin{aligned} \hat{F}(t, t + w | X > t, \mathbf{Z}) &= \frac{\int_0^\infty (C_\theta[\hat{S}_X(t|u), \hat{S}_D(t|u)] - C_\theta[\hat{S}_X(t|u), \hat{S}_D(t + w|u)]) f_{\hat{\eta}}(u) du}{\int_0^\infty C_\theta[\hat{S}_X(t|u), \hat{S}_D(t|u)] f_{\hat{\eta}}(u) du}, \\ \hat{F}(t, t + w | X = x, \mathbf{Z}) &= \frac{\int_0^\infty \left( C_\theta^{[0,1]}[\hat{S}_X(x|u), \hat{S}_D(t|u)] - C_\theta^{[1,0]}[\hat{S}_X(x|u), \hat{S}_D(t + w|u)] \right) \hat{r}_X(x|u) \hat{S}_X(x|u) f_{\hat{\eta}}(u) du}{\int_0^\infty C_\theta^{[1,0]}[\hat{S}_X(x|u), \hat{S}_D(t|u)] \hat{r}_X(x|u) \hat{S}_X(x|u) f_{\hat{\eta}}(u) du}, \end{aligned} \tag{1}$$

where  $\hat{r}_X(x|u) = -\partial\{\log \hat{S}_X(x|u)\} / \partial x$  and  $C_\theta^{[1,0]}(v, w) = \partial C_\theta^{[1,0]}(v, w) / \partial v$ .

We implemented the computation of  $\hat{F}(t, t + w | \cdot, \mathbf{Z})$  in the R package *joint.Cox*. The 95% confidence interval (CI) for  $\hat{F}(t, t + w | \cdot, \mathbf{Z})$  can be obtained via the Monte Carlo method

described in Appendix B of [10]. We suggest a graphical display for the 95% CI to visualize the interval length.

#### 2.4. Online Web Applications

In a series of papers, Fournier et al. [41], Asar et al. [42], and Lenain et al. [43] developed and validated a web application for dynamic risk predictions of kidney graft failure, making it available online (<https://shiny.idbc.fr/DynPG/> accessed on 1 April 2022). Their prediction formula employed a simplified joint model of longitudinal measurements and time to graft failure. To our knowledge, their study was the only one to provide a validated web application for dynamic risk prediction. Following their study, but adopting a different model (the joint frailty-copula model), we provide statistical validation methods and a tutorial to make a web application for developers (Sections 3 and 4). Based on this tutorial, we made two web applications for predicting overall survival for breast cancer (<https://takeshi.shinyapps.io/Breast-2022-0218/> accessed on 1 April 2022) and ovarian cancer (<https://takeshi.shinyapps.io/Ovarian-2022-0218/> accessed on 1 April 2022).

### 3. Validation Methods

We propose three statistical methods to validate the dynamic prediction formulas (of Section 2.3), which have not been sufficiently discussed in the context of the joint frailty-copula model.

#### 3.1. Calibration Plot

To assess the performance of the dynamic prediction formulas, we propose a calibration plot that measures the agreement between the observed outcomes and predictors [26]. However, the calibration plot as recommended by the TRIPOD statement for non-dynamic formulas under the Cox model [25,26] does not fit our goal. Below, we define a calibration plot for dynamic prediction under the joint frailty-copula model.

For a patient alive at  $t$  with his/her tumor status  $H(t, X)$  and covariates  $\mathbf{Z}$ , the survival outcome  $I(D > t + w)$  is calibrated by  $\hat{S}(t, t + w|H(t, X), \mathbf{Z}) = 1 - \hat{F}(t, t + w|H(t, X), \mathbf{Z})$ . The aim of the calibration plot is to graphically show the agreement between the outcome and predictor based on the observed patients, namely,  $I(T_{ij}^* > t + w)$  and  $\hat{S}(t, t + w|H(t, T_{ij}), \mathbf{Z}_{ij})$ . However,  $I(T_{ij}^* > t + w)$  is biased for  $I(D_{ij} > t + w)$  since  $T_{ij}^*$  is the censored version of  $D_{ij}$ . Thus, the following weight function [10,18,19] must be defined for the bias correction:

$$\hat{w}_{ij}(t, t + w) = \frac{\delta_{ij}^* \hat{G}(t)}{\hat{G}(T_{ij}^*)} I(T_{ij}^* \leq t + w) + \frac{\hat{G}(t)}{\hat{G}(t + w)} I(T_{ij}^* > t + w),$$

where  $\hat{G}(t)$  is the Kaplan–Meier (KM) estimate of the censoring survival function  $G(t) = P(C_{ij} > t)$  by treating  $T_{ij}^*$  as the event time and  $1 - \delta_{ij}^*$  as the event indicator.

We define the observed survival probability:

$$\text{Obs}(w) = \frac{1}{Y(t)} \sum_{ij} I(T_{ij}^* > t) \hat{w}_{ij}(t, t + w) I(T_{ij}^* > t + w).$$

We also define the predicted survival probability:

$$\text{Pred}(w) = \frac{1}{Y(t)} \sum_{ij} I(T_{ij}^* > t) \hat{w}_{ij}(t, t + w) \hat{S}(t, t + w|H(t, T_{ij}), \mathbf{Z}_{ij}), \tag{2}$$

where  $Y(t) = \sum_{ij} I(T_{ij}^* > t)$ . We finally define a calibration plot by points  $\{\text{Pred}(w_k), \text{Obs}(w_k)\}, k = 1, 2, \dots, K$ , where  $0 < w_1 < \dots < w_K$  is a prespecified sequence. If the plot is placed on the diagonal line defined by  $\text{Pred}(w_k) = \text{Obs}(w_k)$ , the ideal performance of the prediction formula is achieved.

### 3.2. Brier Score

A good calibration is not enough to show the effectiveness of the dynamic prediction formula. For instance, the dynamic KM estimator  $\hat{F}(t, t+w) = 1 - \hat{S}^{\text{KM}}(t+w)/\hat{S}^{\text{KM}}(t)$  may exhibit a good calibration plot, where  $\hat{S}^{\text{KM}}(\cdot)$  is calculated by partial data  $(T_{ij}^*, \delta_{ij}^*)$  for  $i = 1, 2, \dots, G$  and  $j = 1, 2, \dots, N_i$ . However, this does not mean that the resultant prediction is precise and efficient. Therefore, to assess the performance of the prediction formula, Emura et al. [10] utilized the Brier scores [44] for the joint model and KM estimator defined, respectively, by

$$\hat{\text{Err}}^{\text{JM}}(t, t+w) = \frac{1}{Y(t)} \sum_{ij} \mathbf{I}(T_{ij}^* > t) \hat{w}_{ij}(t, t+w) \{ \mathbf{I}(T_{ij}^* > t+w) - \hat{S}(t, t+w | H(t, T_{ij}), \mathbf{Z}_{ij}) \}^2,$$

$$\hat{\text{Err}}^{\text{KM}}(t, t+w) = \frac{1}{Y(t)} \sum_{ij} \mathbf{I}(T_{ij}^* > t) \hat{w}_{ij}(t, t+w) \{ \mathbf{I}(T_{ij}^* > t+w) - \hat{S}^{\text{KM}}(t+w)/\hat{S}^{\text{KM}}(t) \}^2.$$

The confidence interval of the Brier score can be obtained by a bootstrap method [10]. While the Brier score is one of many scoring methods for assessing the prediction performance [45], it is the most commonly used measure in joint models.

We require  $\hat{\text{Err}}^{\text{JM}}(t, t+w) < \hat{\text{Err}}^{\text{KM}}(t, t+w)$ . This requirement may not be trivially satisfied because the dynamic KM estimator itself is a good predictor that is superior to the non-dynamic KM estimator [23]. Therefore, we suggest using a joint model for prediction only if the upper limit of the 95% CI of  $\hat{\text{Err}}^{\text{JM}}(t, t+w)$  is less than  $\hat{\text{Err}}^{\text{KM}}(t, t+w)$ .

### 3.3. The C-Index for Discrimination Performance

We propose a method to calculate a dynamic version of the *c*-index. First, as our prediction shall be made for all individuals who are at risk at prediction time  $t$ , we define a risk set  $R(t) = \{(i, j); T_{ij}^* > t\}$ . Next, as our prediction horizon is up to  $t+w$ , we define a censored outcome,  $\{\min(T_{ij}^*, t+w), \delta_{ij}^* \mathbf{I}(T_{ij}^* \leq t+w)\}$ , for  $(i, j) \in R(t)$ . The *c*-index is defined as the concordance measure between this outcome and its predictor  $\hat{S}(t, t+w | H(t, T_{ij}), \mathbf{Z}_{ij})$  given  $t$  and  $t+w$ . To compute the *c*-index, one can use the R function “concordance(.)” in the survival package.

## 4. Tutorial: Building Web Applications

The implementation of the dynamic prediction tools with the joint frailty-copula model is a highly technical and complex process. Therefore, we present the roadmap for developers who wish to build a web application to implement the prediction method using the R packages *joint.Cox* and *Shiny*. The resultant web application is accessible to users, including clinicians and patients, without knowledge of R.

The proposed methods consist of the following major steps:

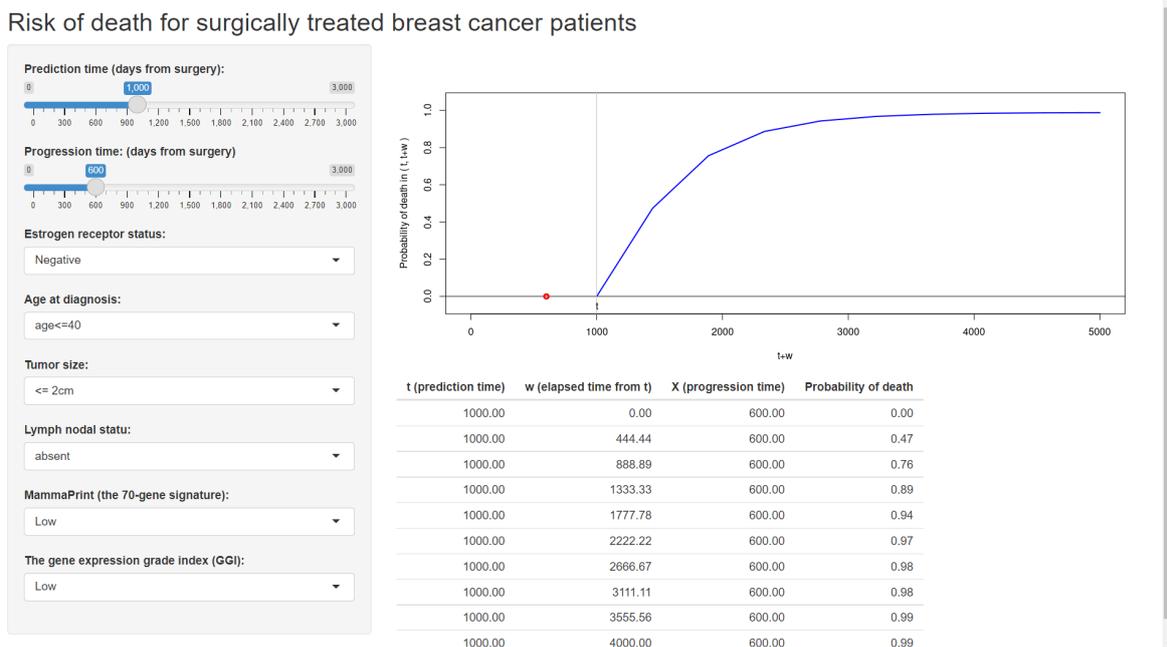
**Step 1:** Fit a training dataset to a joint frailty-copula model using the R package *joint.Cox*;

**Step 2:** Validate the fitted model;

**Step 3:** Use the “app.R” file to build a web application using the R package *Shiny*.

In Step 1, a developer obtains a training dataset and fits it to the joint frailty-copula model that is defined in Section 2.2. Step 2 validates the model by checking the prediction capability using the methods presented in Section 3. If the validation results are not satisfactory, the developer goes back to Step 1. If the model is satisfactorily validated, it is transformed to a web application in Step 3. An example of the “app.R” file in Step 3 is the file “app\_breast.R” available in the Supplementary Materials.

Figure 2 displays the web application made by Shiny (Step 3), which is already available online.



**Figure 2.** The web application for a clinical prediction tool made by applying the proposed methods to breast cancer data. The interactive version is available at <https://takeshi.shinyapps.io/Breast-2022-0218/> (accessed on 1 April 2022).

Steps 1–3 should be performed by developers who have some knowledge of the two R packages. Below are the step-by-step instructions for developers.

**Step 1: Fit a training dataset to a model.**

In this first step, all the unknown parameters in the model have to be estimated by fitting an appropriate training dataset.

We explain this step using the dataset of breast cancer patients from Haibe-Kains et al. [4]. The endpoint of interest is the time to death measured from the date of surgery. From their data, we extracted 720 breast cancer patients having the complete information for the metastasis and death events and their associated covariates (Table 1). All the patients were treated with surgery, and some of them were treated with adjuvant systemic therapy in their first-line treatments.

**Table 1.** The breast cancer dataset of Haibe-Kains et al. [4].

Maximum Follow-Up Days	Dataset <sup>a</sup>	N	The Number of Observed Events (Event Rates)		
			Metastasis	Death	Censoring
5165	CAL	109	24 (22%)	75 (69%)	34 (31%)
6694	NKI	295	101 (34%)	79 (27%)	216 (73%)
9108	TRANSBIG	196	62 (32%)	56 (29%)	140 (71%)
8267	UCSF	120	19 (16%)	39 (32%)	81 (68%)
9108	Total	720	206 (29%)	249(35%)	471 (65%)

**Notes:** The R code for obtaining the data is available in the Supplementary Materials. The data are a subset from the file “jnci-JNCI-11-0924-s02.csv” available in the Supplementary Data of Haibe-Kains et al. [4]; the file is available on the journal webpage. <sup>a</sup> Datasets are signified as acronyms: CAL = dataset of breast cancer patients from the University of California, San Francisco, and the California Pacific Medical Center (United States); NKI = National Kanker Institute (the Netherlands); TRANSBIG = dataset collected by the TransBIG consortium (Europe); UCSF = University of California, San Francisco (United States). The extracted data are the subset having complete values of “t.dmfs: time for distant metastasis-free survival (days)”, “e.dmfs: event for distant metastasis-free survival”, “t.os: time for overall survival (days)”, and “e.os: event for overall survival”, as well as covariates (ER, Size, Node, Age, MammaPrint, and GGI). The median follow-up time was calculated from the Kaplan–Meier estimator for the time to censoring for each study. The event rates were calculated separately for each study.

A developer can fit the joint frailty-copula model to the breast cancer data using the R function *jointCox.reg(.)* and then derive a prediction formula using the R function *F.prediction(.)*. These R functions are the main tools in the R package *joint.Cox*. The R code for applying these functions to the breast cancer data is available in the Supplementary Materials. The usage of the package is detailed in the book of Emura et al. [28] or its preprint available at <https://sites.google.com/view/takeshi-emura> (accessed on 1 April 2022).

In the joint frailty-copula model, the following covariates should be included:

- Estrogen receptor status (**ER** = 1 for positive; = 0 for negative);
- Tumor size (**Size** = 1 for > 2 cm; = 0 for ≤ 2 cm);
- Lymph nodal status (**Node** = 1 for present; =0 for absent);
- Age at diagnosis (**Age** = 1 for age ≤40; =2 for 40 < age ≤ 50; = 3 for age>50);
- The 70-gene signature developed by [1] (**MammaPrint** = 1 for high; = −1 for low);
- The gene expression grade index (GGI) defined by [3] (**GGI** = 1 for high; = −1 for low).

Briefly, the above covariates were selected by optimizing the likelihood cross-validation (LCV) criterion [28]. The LCV accounts for the number of parameters in the model in a similar manner to the AIC. The details are available upon request to the corresponding author.

### Step 2: Validate the prediction formula.

Before prediction results are reported to a patient, it is important for a developer to:

- (i) Check the confidence interval (CI) for the prediction formula;
- (ii) Check the calibration plot, Brier score, and *c*-index.

From (i), one can see the variability of the predicted risk formulas by the length of the CI. A wide CI results from a number of reasons, such as an insufficient number of samples/events and a large value of *t* or *w*. Due to its complex nature, it seems unrealistic to determine a sample size requirement formula for the joint frailty-copula model to achieve the required length of the CI. For this reason, the CI is provided only to show the variability of the prediction results, not as a tool to determine the sample size [46].

From (ii), one can assure that the predicted risk is close to the true risk of a patient. Often, a large value of *t* or *w* produces a large prediction error [10]. This is because the training dataset may not contain a sufficient amount of observed event times beyond a large value of *t*. Even if *t* is small, a larger value of the prediction horizon *t* + *w* introduces a larger uncertainty of predicting the occurrence of an event.

Therefore, from (i) and (ii), a developer should choose appropriate values of *t* and *w*.

### Step 3: Make a web application using Shiny.

We suggest the R package *Shiny* to transfer the prediction model (results from Step 1) to a web application. With *Shiny*, one can easily convert R programs into a web application. Usually, this process is carried out by making an “app.R” file, a program file written in the R language. We developed a template file (available in the Supplementary Materials) so that developers can modify the parameters for their own prediction settings. Once the parameters are appropriately tuned by developers according to Steps 1–2, the web application is immediately built by running the template file. The web application can then be made publicly available through a free and self-service platform, *shinyapps.io* (<https://www.shinyapps.io/> accessed on 1 April 2022).

## 5. Results

We report the results of developing a web-based risk prediction tool following the tutorial.

The risk prediction tool was developed by fitting the joint frailty-copula model to the dataset of 720 breast cancer patients (Table 1). Each patient yields possibly censored outcomes: time to death and time to metastasis both measured from the date of surgery. These event times were treated as OS and TTP for the joint frailty-copula model. The covariates included in the model are the estrogen receptor status (ER=1 for positive; =0 for negative), tumor size (Size = 1 for >2 cm; = 0 for ≤2 cm), lymph nodal status (Node = 1 for

present; =0 for absent), age at diagnosis (Age = 1 for age  $\leq 40$ ; = 2 for  $40 < \text{age} \leq 50$ ; = 3 for age  $> 50$ ), MammaPrint (MammaPrint = 1 for high; = -1 for low), and the GGI defined by [3] (GGI = 1 for high; = -1 for low). The R code to reproduce the results of this section is available in the Supplementary Materials.

### 5.1. Model Fitting

We fitted the joint model to the breast cancer data using the R function *jointCox.reg(.)*. Consequently, we obtained the covariate effects on the hazard for time to metastasis and the covariate effects on the hazard for OS.

$$\beta_1' \mathbf{Z}_1 = (-0.14 \times \text{Age}) + (-0.23 \times \text{ER}) + (0.27 \times \text{Size}) + (0.20 \times \text{MammaPrint}) + (0.19 \times \text{GGI})$$

$$\beta_2' \mathbf{Z}_2 = (-0.36 \times \text{ER}) + (0.14 \times \text{Node}) + (0.27 \times \text{Size}) + (0.17 \times \text{MammaPrint}) + (0.25 \times \text{GGI}).$$

All the regression coefficients were significant ( $p < 0.05$ ). The estimate of the frailty variance was  $\eta = 0.067$ , and its coefficient was  $\alpha = 1$ , implying the presence of unobserved factors affecting both metastasis and death. The estimate of the Clayton copula parameter was  $\theta = 10.7$  (95% CI: 8.6–13.4), showing strong dependence between time to metastasis and time to death (Kendall's tau = 0.84). The baseline hazard functions were estimated as

$$\hat{\rho}_0(t) = 0.20 \times M_1(t) + 0.39 \times M_2(t) + 0.20 \times M_3(t) + 0.42 \times M_4(t) + 0.24 \times M_5(t),$$

$$\hat{\lambda}_0(t) = 0.05 \times M_1(t) + 0.38 \times M_2(t) + 0.37 \times M_3(t) + 0.09 \times M_4(t) + 0.00 \times M_5(t),$$

for  $t < 9108$  (days), which is the maximum follow-up date, where the  $M$ 's are the spline basis functions [28]. More details on the model fitting processes are available upon request.

### 5.2. Developing and Validating a Predictor

To develop and validate the prediction formula, we considered a hypothetical patient:

#### Patient 1:

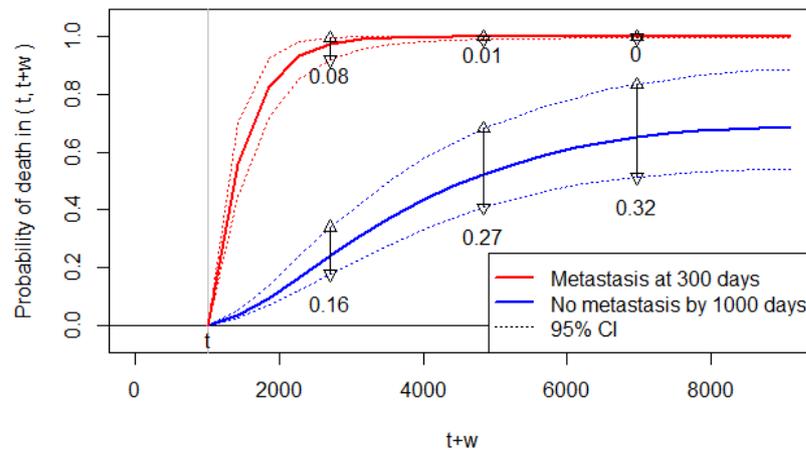
- Age at diagnosis: 45 years
- Estrogen receptor: positive
- Tumor size: >2cm
- Lymph nodal status: present
- MammaPrint: high
- GGI: high

There are 23 breast cancer patients (out of 720 patients) with the same covariate status as Patient 1. Among them, 5 patients experienced metastasis within their follow-up period (the other 18 patients were censored, and thus their metastatic status is unknown). Among the five patients, two developed metastasis within 1000 days, and the other three developed metastasis after 1000 days. Accordingly, we set the prediction time at  $t = 1000$  (days) and then considered two scenarios: (a) Patient 1 does not develop metastasis by 1000 days; (b) Patient 1 develops metastasis at 300 days.

Accordingly, Figure 3 displays the conditional probability of death predicted at  $t = 1000$  (days) for  $0 \leq w \leq 8108$ , which is the allowable range of  $w$  for this dataset; the figure displays  $F(1000, 1000 + w | X > t, \mathbf{Z})$  for scenario (a), and  $F(1000, 1000 + w | X = 300, \mathbf{Z})$  for scenario (b). We can observe that the predicted risk of death increases significantly by the metastasis at 300 days.

Below, we validate the prediction formulas by checking the CI, calibration plot, Brier score, and  $c$ -index.

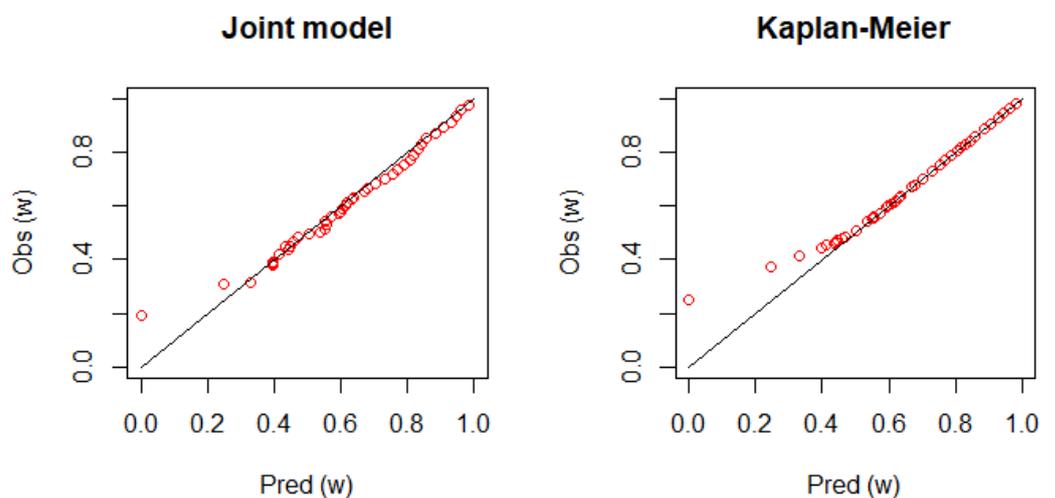
The 95% CIs for the predicted probabilities are sufficiently narrow to discriminate the two scenarios. However, the widths of the 95% CI are somewhat wide for the predicted probabilities without metastasis (Figure 3). Thus, while we are certain about the increased risk of death by metastasis, the predicted probabilities of death have some uncertainty for Patient 1 with no metastasis. For Patient 1 with metastasis at 300 days, the predicted probabilities almost reach the upper bound of one. Thus, we are almost certain about the future death of Patient 1 if she/he develops metastasis.



**Figure 3.** The predicted probability of death when the prediction time is set at  $t = 1000$  days. The 95% CIs are indicated by the dotted lines ( . . . . ), and their widths are shown by the vertical lines (and the number below the lines) at three time points.

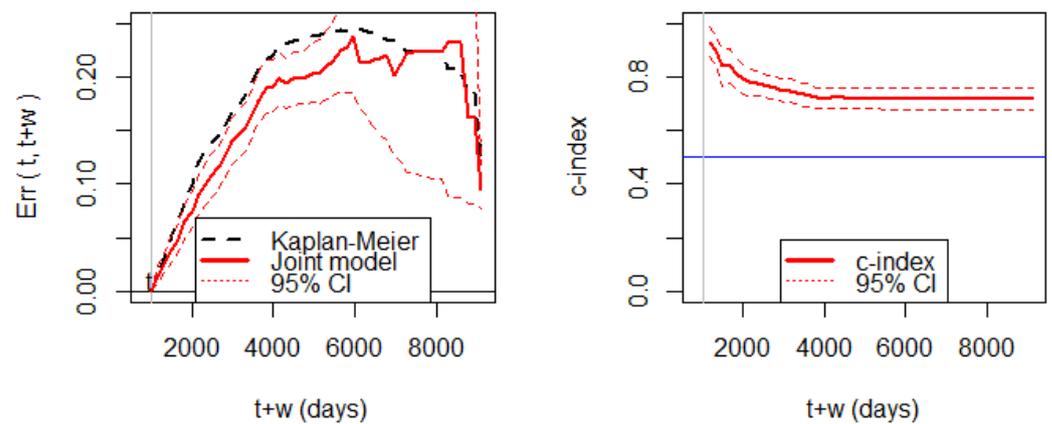
Note that Figure 3 shows the results under a specific prediction setting. A clinician may need a global risk measure for the increase in death after metastasis without specifying the prediction setting. One way is to report the relative risk parameter estimated as  $(\theta + 1) = 11.7$  (95% CI: 9.6–14.4). Thus, a patient with metastasis has an 11.7 times higher risk of death compared to a patient without metastasis.

Figure 4 shows the calibration plot for Patient 1. The figure displays a good agreement between the observed survival rate and predicted survival rate as the plot is placed on the diagonal line. However, it is not clear how the joint model performs better than the dynamic KM estimator.



**Figure 4.** The calibration plots comparing the observed and predicted survival rates, consisting of  $\{\text{Pred}(w_k), \text{Obs}(w_k)\}$  with equally spaced prediction horizons,  $0 < w_1 < \dots < w_{50} = 8108$  (days). If the plots are placed on the diagonal line, the ideal performance of the prediction formula is achieved. **Left panel:** the joint frailty-copula model; **right panel:** the dynamic KM estimator.

Figure 5 shows the Brier score for assessing prediction errors. We can observe that the prediction error of the joint model is significantly smaller than the prediction error of the KM estimator before 5000 days. The wide width of the 95% CI after 5000 days is due to the small number of patients alive. The  $c$ -index shows that the discrimination ability is consistently higher than 0.5.



**Figure 5.** (Left panel): prediction errors (Brier score) based on the breast cancer data at the prediction time at 1000 days. (Right panel): the  $c$ -index for discrimination ability with the 95% CI based on the same setting.

Hence, the joint model has enough predictive power up to 5000 days, though the maximum follow-up time is 9108 days. In other words, a reliable prediction is not possible beyond 5000 days. Therefore, when we set  $t = 1000$  (days), we have to choose the prediction horizon by  $0 \leq t + w \leq 5000$ . Thus,  $0 < w \leq 4000$  is the allowable range of  $w$  for this dataset and the joint model.

### 5.3. Upload a Web Application

On the basis of the developed prediction formula for breast cancer patients, we wrote a program file that can run under the *Shiny* package [40]. This file builds a web application and also publishes the application on the internet. For reference for users, we provide a template file, “app\_breast”, in the Supplementary Materials, which can be directly used to upload our web application on the internet. With this file, we converted the developed prediction formula into a web application and published it online (Figure 2).

The concrete steps to build an online application are as follows.

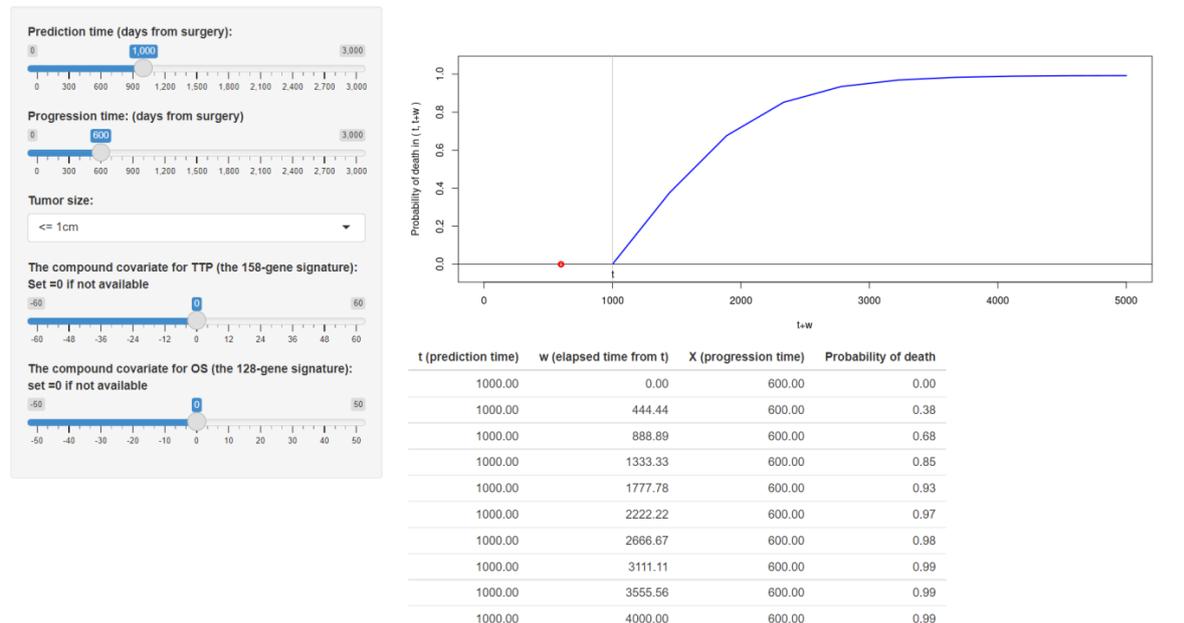
- Step 1:** Open the “app\_breast” file in R studio;
- Step 2:** Run the code in the file (as with the usual R code), and then the application is generated in a window;
- Step 3:** Check if the application works properly;
- Step 4:** Publish the application (click the “Publish” icon).

Developers may edit our template file “app\_breast” according to their training data. Developers working on their own data need to add/delete input variables and adjust all the model parameters and prediction time points. The editing process requires the basic programming skills for R and *Shiny*.

As an additional example, we also built our web application based on our previously validated formula for ovarian cancer patients [10,28]. The published web application is shown in Figure 6.

Following the tutorial, developers can publish their original web applications using their own validated prediction formulas for their targeted patients.

## Risk of death for surgically treated ovarian cancer patients



**Figure 6.** The web application for a clinical prediction tool made by applying the proposed methods to ovarian cancer data. The interactive version is available at <https://takeshi.shinyapps.io/Ovarian-2022-0218/> (accessed on 1 April 2022).

## 6. Conclusions and Discussion

In this article, we present a tutorial for developers who wish to build a web application for the personalized risk prediction method of [10]. While we reviewed the methods of [10], we also propose new validation tools: the calibration plot and the *c*-index, that were not considered in the context of the joint frailty-copula model. We provide a Shiny-based template to transfer the prediction model to a user-friendly web application. Following the proposed tutorial and template, we built two web applications for predicting the risk for breast cancer patients (Figure 2) and ovarian cancer patients (Figure 6). Developers learning our tutorial will have the ability to create a variety of online web applications using their own datasets.

According to the predicted risk of death, a clinician may suggest a second-line treatment, or he/she may suggest stopping the treatment [23]. The key factor is whether the clinician regards tumor progression as the failure or the inadequacy of the initial treatment. In this respect, clinicians can make their decision on the second-line treatment with the help of the risk of death with or without tumor progression (Figure 1).

Throughout this article, we used a breast cancer dataset as a demonstrative example for developing a dynamic risk prediction scheme. This is our second example of demonstrating the joint frailty-copula model as a reliable dynamic prediction scheme. Our previous work [10] employed the ovarian cancer dataset of Ganzfried et al. [47] to develop a risk prediction formula of death according to tumor progression status, residual tumor size, and genetic covariates. Other cancer datasets from other clinical studies can be utilized for building web applications, such as a dataset for bladder cancer with an appropriate treatment of competing risks [48,49].

Note that the developed prediction formula for breast cancer patients (Figure 2) has many more clinicopathological covariates than the model for ovarian cancer patients (Figure 6). This is because many of the clinicopathological covariates in ovarian cancer are not significant, nor do they effectively improve the simple model of the residual tumor size alone. The poor predictive power of clinical covariates has motivated a number of survival prediction models with gene expressions in ovarian cancers [8–10,50,51]. While the residual

tumor size is considered the strongest clinical predictor of survival, the tumor progression information has even stronger predictive ability for overall survival [10], providing a rationale for applying dynamic prediction tools. According to these experiences, the proposed dynamic prediction scheme with the joint model would be promising for other cancers as long as the clinician can properly use it.

In this study, we employed the R package *Shiny* to transfer complex prediction formulas to a user-friendly web application. Recently, Fournier et al. [41], Asar et al. [42], and Lenain et al. [43] made a Shiny-based web application for the dynamic prediction of long-term kidney graft failure. Clearly, this type of dynamic prediction tool can promote the development of personalized medicine, allowing clinicians to utilize powerful prediction algorithms without detailed knowledge of R.

The issue that we did not discuss in this article is the construction of a multigene predictor, an important element in personalized medicine. For the breast cancer dataset, we employed the commercially available prognostic signatures MammaPrint [1] and GGI [3]. If there is no established signature, one often has to apply a gene selection (feature selection) method for a number of available genes. One of the most frequently used methods is to select genes according to the significance levels of the associations between each gene and survival [12,13,52–54]. One may apply the *compound.Cox* R package [52] to select an optimal set of genes while assessing their false discovery rate.

The issue of dynamic prediction is not restricted to medical research, which can be seen in the reliability analysis of mechanical items or systems [55–59]. In the reliability prediction of mechanical equipment [56–59], the conditional failure/survival probability was suggested as a measure of prediction in a similar fashion to our prediction scheme in Figure 1. The present tutorial can be a useful guide for building a web application for these engineering issues. Since statistical models in engineering applications involve time-dependent covariates [58] or dependent component failures [60–62], the derivation of a prediction formula is challenging.

This article considers two endpoints (TTP and OS) via the joint frailty-copula model. While more than two endpoints could be observable in clinical trials, the computational cost for implementing multivariate joint models is high, and some approximation techniques are always necessary (see [63–67] and references therein). Currently, the joint frailty-copula model has not been extended to multivariate settings due to its computational difficulty, mainly caused by the need for numerical integrations for frailty terms. Parametric joint frailty-copula models, such as the Weibull and Pareto joint frailty-copula models, reduce the computational cost to some degree [68,69], though they are less flexible compared to spline-based models.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/e24050589/s1>, Supplementary Materials of this article include the following items: File S1. R code to analyze the breast cancer data; File S2. A template file “app\_breast. R” for Shiny.

**Author Contributions:** Conceptualization, T.E.; Formal analysis, T.E.; Funding acquisition, H.M.; Methodology, T.E.; Supervision, S.M.; Writing—original draft, T.E., H.M. and S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by JSPS KAKENHI (JP21K12127) and (JP22K11948).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the numerical results of the article are reproduced by the R code in Supplementary Materials.

**Acknowledgments:** We thank the two reviewers for their helpful comments that improved this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Van't Veer, L.J.; Dai, H.; Van De Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; Van Der Kooy, K.; Marton, M.J.; Witteveen, A.T.; et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **2002**, *415*, 530–536. [[CrossRef](#)] [[PubMed](#)]
2. Shukla, N.; Hagenbuchner, M.; Win, K.T.; Yang, J. Breast cancer data analysis for survivability studies and prediction. *Comput. Methods Programs Biomed.* **2018**, *155*, 199–208. [[CrossRef](#)] [[PubMed](#)]
3. Sotiriou, C.; Wirapati, P.; Loi, S.; Harris, A.; Fox, S.; Smeds, J.; Nordgren, H.; Farmer, P.; Praz, V.; Haibe-Kains, B.; et al. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. *J. Natl. Cancer Inst.* **2006**, *98*, 262–272. [[CrossRef](#)] [[PubMed](#)]
4. Haibe-Kains, B.; Desmedt, C.; Loi, S.; Culhane, A.; Bontempi, G.; Quackenbush, J.; Sotiriou, C. A Three-Gene Model to Robustly Identify Breast Cancer Molecular Subtypes. *J. Natl. Cancer Inst.* **2012**, *104*, 311–325. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, S.; Ivy, J.S.; Payton, F.C.; Diehl, K.M. Modeling the impact of comorbidity on breast cancer patient outcomes. *Health Care Manag. Sci.* **2009**, *13*, 137–154. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, S.; Ivy, J.S.; Wilson, J.R.; Diehl, K.M.; Yankaskas, B.C. Competing risks analysis in mortality estimation for breast cancer patients from independent risk groups. *Health Care Manag. Sci.* **2014**, *17*, 259–269. [[CrossRef](#)] [[PubMed](#)]
7. Neto, C.; Brito, M.; Lopes, V.; Peixoto, H.; Abelha, A.; Machado, J. Application of Data Mining for the Prediction of Mortality and Occurrence of Complications for Gastric Cancer Patients. *Entropy* **2019**, *21*, 1163. [[CrossRef](#)]
8. Waldron, L.; Haibe-Kains, B.; Culhane, A.C.; Riester, M.; Ding, J.; Wang, X.V.; Ahmadifar, M.; Tyekucheva, S.; Bernau, C.; Risch, T.; et al. Comparative Meta-analysis of Prognostic Gene Signatures for Late-Stage Ovarian Cancer. *J. Natl. Cancer Inst.* **2014**, *106*, 49. [[CrossRef](#)] [[PubMed](#)]
9. Emura, T.; Nakatochi, M.; Murotani, K.; Rondeau, V. A joint frailty-copula model between tumour progression and death for meta-analysis. *Stat. Methods Med Res.* **2017**, *26*, 2649–2666. [[CrossRef](#)] [[PubMed](#)]
10. Emura, T.; Nakatochi, M.; Matsui, S.; Michimae, H.; Rondeau, V. Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model. *Stat. Methods Med. Res.* **2018**, *27*, 2842–2858. [[CrossRef](#)] [[PubMed](#)]
11. Rosenwald, A.; Wright, G.; Chan, W.C.; Connors, J.M.; Campo, E.; Fisher, R.I.; Gascoyne, R.D.; Muller-Hermelink, H.K.; Smeland, E.B.; Giltnane, J.M.; et al. The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-Cell Lymphoma. *N. Engl. J. Med.* **2002**, *346*, 1937–1947. [[CrossRef](#)] [[PubMed](#)]
12. Matsui, S. Predicting survival outcomes using subsets of significant genes in prognostic marker studies with microarrays. *BMC Bioinform.* **2006**, *7*, 156. [[CrossRef](#)]
13. Matsui, S. Statistical issues in clinical development and validation of genomic signatures. In *Design and Analysis of Clinical Trials for Predictive Medicine*; Matsui, S., Buyse, M., Simon, R., Eds.; CRC Press: Boca Raton, FL, USA, 2015; pp. 207–226.
14. Keek, S.; Wesseling, F.; Woodruff, H.; van Timmeren, J.; Nauta, I.; Hoffmann, T.; Cavalieri, S.; Calareso, G.; Primakov, S.; Leijenaar, R.; et al. A Prospectively Validated Prognostic Model for Patients with Locally Advanced Squamous Cell Carcinoma of the Head and Neck Based on Radiomics of Computed Tomography Images. *Cancers* **2021**, *13*, 3271. [[CrossRef](#)]
15. Michiels, S.; Koscielny, S.; Hill, C. Prediction of cancer outcome with microarrays: A multiple random validation strategy. *Lancet* **2005**, *365*, 488–492. [[CrossRef](#)]
16. Choi, J.; Oh, I.; Seo, S.; Ahn, J. G2Vec: Distributed gene representations for identification of cancer prognostic genes. *Sci. Rep.* **2018**, *8*, 13729. [[CrossRef](#)] [[PubMed](#)]
17. Kim, M.; Oh, I.; Ahn, J. An Improved Method for Prediction of Cancer Prognosis by Network Learning. *Genes* **2018**, *9*, 478. [[CrossRef](#)] [[PubMed](#)]
18. Graf, E.; Schmoor, C.; Sauerbrei, W.; Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Stat. Med.* **1999**, *18*, 2529–2545. [[CrossRef](#)]
19. Gerds, T.A.; Schumacher, M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biom. J.* **2006**, *48*, 1029–1040. [[CrossRef](#)]
20. Proust Lima, C.; Blanche, P. Dynamic Predictions. In *Wiley StatsRef: Statistics Reference Online*; Wiley: Hoboken, NJ, USA, 2014; pp. 1–6.
21. van Houwelingen, H.C.; Putter, H. *Dynamic Prediction in Clinical Survival Analysis*; CRC Press: New York, NY, USA, 2011.
22. Sène, M.; Taylor, J.M.G.; Dignam, J.J.; Jacqmin-Gadda, H.; Proust-Lima, C. Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation. *Stat. Methods Med. Res.* **2016**, *25*, 2972–2991. [[CrossRef](#)] [[PubMed](#)]
23. Kheirandish, M.; Catanzaro, D.; Crudu, V.; Zhang, S. Integrating landmark modeling framework and machine learning algorithms for dynamic prediction of tuberculosis treatment outcomes. *J. Am. Med Inform. Assoc.* **2022**, *29*, 900–908. [[CrossRef](#)]
24. Thompson, R.H.; Leibovich, B.C.; Lohse, C.M.; Cheville, J.C.; Zincke, H.; Blute, M.L.; Frank, I. Dynamic Outcome Prediction in Patients With Clear Cell Renal Cell Carcinoma Treated With Radical Nephrectomy: The D-SSIGN Score. *J. Urol.* **2007**, *177*, 477–480. [[CrossRef](#)] [[PubMed](#)]
25. Moons, K.G.M.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.A.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [[CrossRef](#)] [[PubMed](#)]

26. Riley, R.D.; Ensor, J.; Snell, K.; Debray, T.; Altman, D.G.; Moons, K.G.M.; Collins, G. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *BMJ* **2016**, *353*, i3140. [[CrossRef](#)] [[PubMed](#)]
27. Rondeau, V.; Pignon, J.-P.; Michiels, S.; on behalf of the MACH-NC Collaborative Group. A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. *Stat. Methods Med. Res.* **2015**, *24*, 711–729. [[CrossRef](#)] [[PubMed](#)]
28. Emura, T.; Matsui, S.; Rondeau, V. *Survival Analysis with Correlated Endpoints, Joint Frailty-Copula Models*; JSS Research Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2019.
29. Emura, T.; Sofeu, C.L.; Rondeau, V. Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials. *Stat. Methods Med. Res.* **2021**, *30*, 2634–2650. [[CrossRef](#)] [[PubMed](#)]
30. Sofeu, C.L.; Emura, T.; Rondeau, V. A joint frailty-copula model for meta-analytic validation of failure time surrogate endpoints in clinical trials. *Biom. J.* **2021**, *63*, 423–446. [[CrossRef](#)] [[PubMed](#)]
31. Michiels, S.; Baujat, B.; Mahé, C.; Sargent, D.; Pignon, J. Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses. *J. Clin. Epidemiol.* **2005**, *58*, 238–245. [[CrossRef](#)]
32. Rotolo, F.; Paoletti, X.; Burzykowski, T.; Buyse, M.; Michiels, S. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Stat. Methods Med. Res.* **2019**, *28*, 170–183. [[CrossRef](#)] [[PubMed](#)]
33. Rotolo, F.; Paoletti, X.; Michiels, S. *surrosv*: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. *Comput. Methods Programs Biomed.* **2018**, *155*, 189–198. [[CrossRef](#)] [[PubMed](#)]
34. Peng, M.; Xiang, L.; Wang, S. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Comput. Stat. Data Anal.* **2018**, *124*, 53–70. [[CrossRef](#)]
35. Burzykowski, T.; Molenberghs, G.; Buyse, M.; Geys, H.; Renard, D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2001**, *50*, 405–422. [[CrossRef](#)]
36. Schneider, S.; Demarqui, F.; Colosimo, E.A.; Mayrink, V.D. An approach to model clustered survival data with dependent censoring. *Biom. J.* **2019**, *62*, 157–174. [[CrossRef](#)] [[PubMed](#)]
37. Ha, I.D.; Vaida, F.; Lee, Y. Interval estimation of random effects in proportional hazards models with frailties. *Stat. Methods Med. Res.* **2016**, *25*, 936–953. [[CrossRef](#)]
38. Ha, I.D.; Lee, Y. A review of h-likelihood for survival analysis. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1157–1178. [[CrossRef](#)]
39. Emura, T. *jointCox: Joint Frailty-Copula Models for Tumour Progression and Death in Meta-Analysis*, CRAN. 2022. Available online: <https://CRAN.R-project.org/package=jointCox> (accessed on 1 April 2022).
40. Winston Chang, J.C.; Allaire, J.J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for R*. CRAN. 2021. Available online: <https://CRAN.R-project.org/package=shiny> (accessed on 1 April 2022).
41. Fournier, M.-C.; Foucher, Y.; Blanche, P.; Legendre, C.; Girerd, S.; Ladrrière, M.; Morelon, E.; Buron, F.; Rostaing, L.; Kamar, N.; et al. Dynamic predictions of long-term kidney graft failure: An information tool promoting patient-centred care. *Nephrol. Dial. Transplant.* **2019**, *34*, 1961–1969. [[CrossRef](#)] [[PubMed](#)]
42. Asar, Ö.; Fournier, M.-C.; Dantan, E. Dynamic predictions of kidney graft survival in the presence of longitudinal outliers. *Stat. Methods Med. Res.* **2021**, *30*, 185–203. [[CrossRef](#)] [[PubMed](#)]
43. Lenain, R.; Dantan, E.; Giral, M.; Foucher, Y.; Asar, Ö.; Naesens, M.; Hazzan, M.; Fournier, M.-C. External Validation of the DynPG for Kidney Transplant Recipients. *Transplantation* **2021**, *105*, 396–403. [[CrossRef](#)]
44. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [[CrossRef](#)]
45. Hughes, G.; Topp, C.F. Probabilistic Forecasts: Scoring Rules and Their Decomposition and Diagrammatic Representation via Bregman Divergences. *Entropy* **2015**, *17*, 5450–5471. [[CrossRef](#)]
46. Riley, R.D.; Ensor, J.; Snell, K.I.; Harrell, F.E.; Martin, G.P.; Reitsma, J.B.; Moons, K.G.; Collins, G.; Van Smeden, M. Calculating the sample size required for developing a clinical prediction model. *BMJ* **2020**, *368*, m441. [[CrossRef](#)] [[PubMed](#)]
47. Ganzfried, B.F.; Riester, M.; Haibe-Kains, B.; Risch, T.; Tyekucheva, S.; Jazic, I.; Wang, X.V.; Ahmadifar, M.; Birrer, M.J.; Parmigiani, G.; et al. *curatedOvarianData*: Clinically annotated data for the ovarian cancer transcriptome. *Database* **2013**, *2013*, bat013. [[CrossRef](#)] [[PubMed](#)]
48. Lee, M.; Ha, I.D.; Lee, Y. Frailty modeling for clustered competing risks data with missing cause of failure. *Stat. Methods Med. Res.* **2017**, *26*, 356–373. [[CrossRef](#)] [[PubMed](#)]
49. Emura, T.; Shih, J.-H.; Ha, I.D.; Wilke, R. Comparison of the marginal hazard model and the sub-distribution hazard model for competing risks under an assumed copula. *Stat. Methods Med. Res.* **2019**, *29*, 2307–2327. [[CrossRef](#)]
50. Yoshihara, K.; Tajima, A.; Yahata, T.; Kodama, S.; Fujiwara, H.; Suzuki, M.; Onishi, Y.; Hatae, M.; Sueyoshi, K.; Fujiwara, H.; et al. Gene Expression Profile for Predicting Survival in Advanced-Stage Serous Ovarian Cancer Across Two Independent Datasets. *PLoS ONE* **2010**, *5*, e9615. [[CrossRef](#)]
51. Yoshihara, K.; Tsunoda, T.; Shigemizu, D.; Fujiwara, H.; Hatae, M.; Fujiwara, H.; Masuzaki, H.; Katabuchi, H.; Kawakami, Y.; Okamoto, A.; et al. High-Risk Ovarian Cancer Based on 126-Gene Expression Signature Is Uniquely Characterized by Downregulation of Antigen Presentation Pathway. *Clin. Cancer Res.* **2012**, *18*, 1374–1385. [[CrossRef](#)]
52. Emura, T.; Matsui, S.; Chen, H.-Y. *compoundCox*: Univariate feature selection and compound covariate for predicting survival. *Comput. Methods Programs Biomed.* **2019**, *168*, 21–37. [[CrossRef](#)] [[PubMed](#)]

53. Emura, T.; Chen, Y.-H. Gene selection for survival data under dependent censoring: A copula-based approach. *Stat. Methods Med. Res.* **2016**, *25*, 2840–2857. [[CrossRef](#)]
54. Redekar, S.S.; Varma, S.L.; Bhattacharjee, A. Identification of key genes associated with survival of glioblastoma multiforme using integrated analysis of TCGA datasets. *Comput. Methods Programs Biomed. Update* **2022**, *2*, 100051. [[CrossRef](#)]
55. Kordestani, M.; Saif, M.; Orchard, M.E.; Razavi-Far, R.; Khorasani, K. Failure Prognosis and Applications—A Survey of Recent Literature. *IEEE Trans. Reliab.* **2021**, *70*, 728–748. [[CrossRef](#)]
56. Hong, Y.; Meeker, W.Q.; McCalley, J.D. Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Ann. Appl. Stat.* **2009**, *3*, 857–879. [[CrossRef](#)]
57. Mitra, D.; Kundu, D.; Balakrishnan, N. Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1019–1048. [[CrossRef](#)]
58. Zheng, R.; Najafi, S.; Zhang, Y. A recursive method for the health assessment of systems using the proportional hazards model. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108379. [[CrossRef](#)]
59. Emura, T.; Michimae, H. A Review of Field Failure Data Analysis Involving Left-Truncation and Right-Censoring. *Qual. Reliab. Eng. Int* **2022**, *In Revision*.
60. Ota, S.; Kimura, M. Effective estimation algorithm for parameters of multivariate Farlie–Gumbel–Morgenstern copula. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1049–1078. [[CrossRef](#)]
61. Jia, X.; Wang, L.; Wei, C. Reliability Research of Dependent Failure Systems Using Copula. *Commun. Stat.-Simul. Comput.* **2014**, *43*, 1838–1851. [[CrossRef](#)]
62. Fan, T.-H.; Wang, Y.-F.; Ju, S.-K. A Competing Risks Model With Multiply Censored Reliability Data Under Multivariate Weibull Distributions. *IEEE Trans. Reliab.* **2019**, *68*, 462–475. [[CrossRef](#)]
63. Mazroui, Y.; Mathoulin-Pélissier, S.; MacGrogan, G.; Brouste, V.; Rondeau, V. Multivariate frailty models for two types of recurrent events with a dependent terminal event: Application to breast cancer data. *Biom. J.* **2013**, *55*, 866–884. [[CrossRef](#)]
64. Król, A.; Mauguen, A.; Mazroui, Y.; Laurent, A.; Michiels, S.; Rondeau, V. Tutorial in Joint Modeling and Prediction: A Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event. *J. Stat. Softw.* **2017**, *81*, 1–52. [[CrossRef](#)]
65. Murray, J.; Philipson, P. A fast approximate EM algorithm for joint models of survival and multivariate longitudinal data. *Comput. Stat. Data Anal.* **2022**, *170*, 107438. [[CrossRef](#)]
66. Flórez, A.J.; Molenberghs, G.; Van der Elst, W.; Abad, A.A. An efficient algorithm to assess multivariate surrogate endpoints in a causal inference framework. *Comput. Stat. Data Anal.* **2022**, *172*, 107494. [[CrossRef](#)]
67. Philipson, P.; Hickey, G.L.; Crowther, M.J.; Kolamunnage-Dona, R. Faster Monte Carlo estimation of joint models for time-to-event and multivariate longitudinal data. *Comput. Stat. Data Anal.* **2020**, *151*, 107010. [[CrossRef](#)]
68. Wu, B.-H.; Michimae, H.; Emura, T. Meta-analysis of individual patient data with semi-competing risks under the Weibull joint frailty–copula model. *Comput. Stat.* **2020**, *35*, 1525–1552. [[CrossRef](#)]
69. Lin, Y.H.; Sun, L.H.; Tseng, Y.J.; Emura, T. The Pareto type I joint frailty-copula model for clustered bivariate survival data. *Commun. Stat. -Simul. Comput.* **2022**. [[CrossRef](#)]