Supporting Information for: Assessing model prediction performance for the expected cumulative number of recurrent events

1 Situations with right-censoring and no terminal event

This section considers the situation when no terminal event exists. In this case, the counting process of interest $N^*(t)$ counts the number of recurrent events that have occurred before time t and the regression model is defined by:

$$\mathbb{E}[dN^*(t) \mid \bar{X}(t)] = \lambda^*(t \mid \bar{X}(t))dt.$$

Let $\int_0^t \lambda^*(u \mid \bar{X}(u))du$ be the true cumulative rate function. In the absence of a terminal event, this cumulative function has a direct interpretation as the expected cumulative number of recurrent events given the covariate process: $\int_0^t \lambda^*(u \mid \bar{X}(u))du = \mu^*(t \mid \bar{X}(t))$ where, as defined at the beginning of Section (2) of the main paper, $\mu^*(t \mid \bar{X}(t)) = \mathbb{E}[N^*(t) \mid X(u) : 0 \leq u \leq t]$. In the presence of censoring, a variable C is observed such that the observed recurrent event process is now $N(t) = N^*(t \wedge C)$. We assume independent censoring (see [1]) in the following way:

$$\mathbb{E}[dN^*(t) \mid \bar{X}(t)] = \mathbb{E}[dN^*(t) \mid I(C \geq t), \bar{X}(t)].$$

This assumption implies that C does not convey any additional information on the probability of a jump of the recurrent event process. Under this assumption, we have

$$\mathbb{E}[dN(t) \mid I(C \geq t), \bar{X}(t)] = I(C \geq t)\lambda^*(t \mid \bar{X}(t))dt, \tag{1}$$

where $I(\cdot)$ denotes the indicator function. This last equation justifies the use of our criterion (1) in the main paper since $\mathbb{E}[dN(t) \mid \bar{X}(t)] = (1 - G(t- \mid \bar{X}(t)))\lambda^*(t \mid \bar{X}(t))dt$ and therefore Equation (3) of the main paper holds. Next, we assume Assumption 1 of the main paper and we make the following additional assumption.

Assumption (RC)**.** *We assume that there exists a constant $\tau > 0$ and a constant $c > 0$ such that*

1. $\forall t \in [0, \tau]$, $\mathbb{P}[C \geq t \mid \bar{X}(t)] \geq c$ *almost surely,*

2. $N(\tau)$ *is almost surely bounded by a constant.*

Note that condition *1.* was also assumed in [2]. It is stronger than simply assuming $\mathbb{P}[C \geq \tau] \geq c$. Condition *2.* is standard for recurrent event data, see for instance [1]. Finally, note that through Equation (1), conditions *1.* and *2.* imply that $\mathbb{E}[\mu^*(\tau \mid \bar{X}(\tau))] < \infty$.

On the basis of i.i.d. replications $(N_i(t), X_i(t) : 0 \leq t \leq \tau)$, let $\widehat{\mu} \in \mathcal{M}$ be an estimator of μ^* where $\mathcal{M}$ is a class of models that are assumed to be bounded. We propose to evaluate the prediction ability of this estimator through criterion $\widehat{\text{MSE}}(t, \widehat{\mu})$ defined in Equation (1) of the main paper. This criterion involves an estimator of G, the conditional cumulative distribution function of the censoring variable. If C and $X(\cdot)$ are independent, one can estimate G using the empirical cumulative distribution function of the censored variables since all these variables are observed. If C depends on X the conditional distribution of C must also be modelled. This can be done using kernel based estimators such as the Nadaraya-Watson estimator for the binary response variable $I(C \leq t)$ or extensions of this model. For instance, in [3], a local logistic method and an adjusted Nadaraya-Watson estimator are proposed. If the dimension of the covariates that are assumed to depend on the censoring distribution is too large, then a dimension reduction technique can first be employed, for example through a Single-Index-Modelling approach (see [4]). The theoretical results in Section 3 of the main paper are valid under Assumption 1 of the main paper and Assumption (RC). The proof can be found in Section 8 of the main paper.

2 Situations with dependence on prior counts

In this section, we model the probability of a jump of the recurrent events process as a function of the number of previous recurrent events. We still consider a terminal event T^* and we assume that the counting process of interest N^* verifies:

$$\mathbb{E}[dN^*(t) \mid I(T^* \geq t), N^*(t-), \bar{X}(t)] = \sum_{l=1}^{L} I(T^* \geq t, N^*(t-) = l-1)\lambda_l^*(t \mid \bar{X}(t))dt, \tag{2}$$

where $\lambda_l^*(t \mid \bar{X}(t))$ is the true rate function that depends on $l-1$ which represents the number of previous recurrent events that have already occurred and L represents the maximum number of events that can occur. This model was first introduced by [5] and further studied by [6]. It is also important to stress that Equation (2) defines a multi-state model with a different state for each value of $N^*(t-)$ and an absorbing state for the terminal event. In total, there are $L+2$ different states and L different hazard rates for the recurrent event process. In general, in order to estimate μ^* it is usually necessary to also model the hazard rate for the terminal event T^*. It is then possible to specify $L+1$ hazard rates models for the terminal event that depend on the number of recurrent events previously experienced. An example of the multi-state representation with $L=5$ is illustrated in Figure 1. Note that the hazard rates for the terminal event are also represented on the figure and are denoted $\lambda_l^{T^*}$.

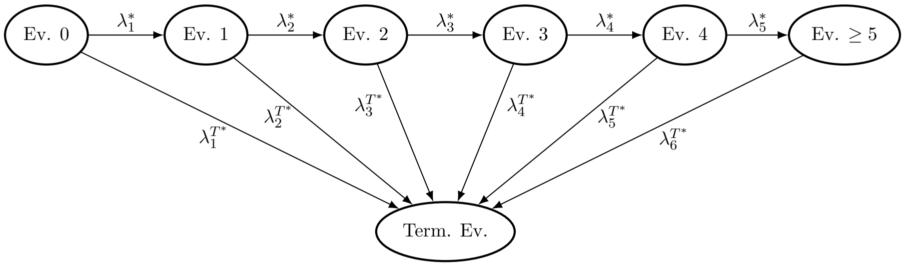

Figure 1: Illustration of a recurrent event model with dependence on prior events as a multi-state situation. The model is defined by Equation (2) with $L=5$. Individuals start in the state Ev. 0 and then can move to the other states as time increases. The state Term. Ev. represents the terminal event and is an absorbing state. The state Ev. ≥ 5 encompasses all recurrent events greater than 5.

Under this model, we observe that

$$\mu^*(t \mid \bar{X}(t)) = \sum_{l=1}^{L} \int_0^t \mathbb{P}[T^* \geq u, N^*(u-) = l-1 \mid \bar{X}(u)]\lambda_l^*(u \mid \bar{X}(u))du, \tag{3}$$

where $\mathbb{P}[T^* \geq u, N^*(u-) = l-1 \mid \bar{X}(u)]$ is usually named the state probability in the usual multi-state framework. An alternative formula for the expected cumulative number of recurrent events can be found in [7], formula (6.51) page 226. It can easily be shown that the two formulas coincide.

In this context, the independent censoring assumption is expressed in the following way:

$$\mathbb{E}[dN^*(t) \mid I(T^* \geq t), N^*(t-), \bar{X}(t)] = \mathbb{E}[dN^*(t) \mid I(T \geq t), N^*(t-), \bar{X}(t)].$$

As before, we denote $T = T^* \wedge C$ the minimum between terminal event and censoring, $N(t) = N^*(T \wedge t)$ the observed counting process and $Y_l(t) = I(T \geq t, N(t-) = l)$ the observed at-risk process. Under the independent censoring assumption, it can be shown that

$$\mathbb{E}[dN(t) \mid I(T \geq t), N(t-), \bar{X}(t)] = \sum_{l=1}^{L} Y_l(t)\lambda_l^*(t \mid \bar{X}(t))dt. \tag{4}$$

We assume Assumption (1) of the main paper and we make the following additional assumption.

Assumption (MSM)**.** *We assume that there exists a constant $\tau > 0$ and a constant $c > 0$ such that*

1. *$\forall l = 0, \dots, L-1$, $\forall t \in [0, \tau]$, $\mathbb{P}[T \geq t, N(t-) = l \mid \bar{X}(t)] \geq c$ almost surely,*
2. *$N(\tau)$ is almost surely bounded by a constant.*

We also assume that C is conditionally independent of $(N^*(\cdot), T^*)$ given $\bar{X}(\cdot)$.

Note that condition 2. of the assumption is automatically satisfied from our model (2). Note also that the independent assumption can be equivalently stated as: C is conditionally independent of $N^*(\cdot)$ given $\bar{X}(\cdot)$ and T^* is conditionally independent of C given $(N^*(\cdot), \bar{X}(\cdot))$. Using Equality (4) one can easily observe that

$$\begin{aligned}\mathbb{E}[dN(t) \mid \bar{X}(t)] &= \sum_{l=1}^{L} \mathbb{P}[T \geq u, N(u-) = l-1 \mid \bar{X}(u)]\lambda_l^*(u \mid \bar{X}(u))du,\\ &= \sum_{l=1}^{L} G_c(u \mid \bar{X}(u))\mathbb{P}[T^* \geq u, N^*(u-) = l-1 \mid \bar{X}(u)]\lambda_l^*(u \mid \bar{X}(u))du,\end{aligned}$$

where we used the fact that $\mathbb{P}[T \geq u, N(u-) = l-1 \mid \bar{X}(u)] = \mathbb{P}[T \geq u, N^*(u-) = l-1 \mid \bar{X}(u)]$ and the independent censoring hypothesis. We then directly see that Equation (3) of the main paper holds.

On the basis of i.i.d. replications $(N_i(t), X_i(t) : 0 \leq t \leq \tau)$, let $\widehat{\mu} \in \mathcal{M}$ be an estimator of μ^* where $\mathcal{M}$ is a class of models that are assumed to be bounded. We propose to evaluate the prediction ability of this estimator through criterion $\widehat{\mathrm{MSE}}(t, \widehat{\mu})$ defined in Equation (1) of the main paper. An estimator of G, the conditional cumulative distribution function of the censoring variable, can be proposed in the same manner as in Section 2.2 of the main paper. The theoretical results in Section 3 of the main paper are valid under Assumption 1 of the main paper and Assumption (MSM). The proof can be found in Section 8 of the main paper.

3 Computation of $A(t)$ in the simulation section 5.1

While the inseparability term cannot be computed on real data, it is possible to obtain its expression when the distribution of all variables are known. In this section, we provide the explicit expression of the inseparability term $A(t)$, in the simulation setting of Section 5.1 of the main paper. To this end, we use Equation (13) of the main paper. We first notice that, for $u < v$, $dN(u)dN(v) = I(C \geq v)dN^*(u)dN^*(v)$ and

$$\begin{aligned}
\mathbb{E}[dN(u)dN(v)] &= \mathbb{E}[I(C \geq v)\mathbb{E}[dN^*(u)dN^*(v) \mid X, C]] \\
&= \mathbb{E}[I(C \geq v)\lambda_0(u)\lambda_0(v)\exp(2\theta_0^\top X)]dudv \\
&= (1 - G(v-))\lambda_0(u)\lambda_0(v)\mathbb{E}[\exp(2\theta_0^\top X)]dudv,
\end{aligned}$$

where we used the fact that $\mathbb{E}[dN^*(u)dN^*(v) \mid X, C] = \mathbb{E}[dN^*(u) \mid X]\mathbb{E}[dN^*(v) \mid X] = \lambda_0(u)\lambda_0(v)\exp(2\theta_0^\top X)dudv$, since under our simulation scheme, N^* is independent of C and $dN^*(u)$ is independent of $dN^*(v)$ conditionally on X, for $u \neq v$. Let $\gamma = 3$, such that C follows a uniform distribution on $[0, \gamma]$. For $t < \gamma$, we have:

$$\mathbb{E}\left[\left(\int_0^t \frac{dN(u)}{1 - G(u- \mid X(u))}\right)^2\right] = A_1(t) + A_2(t),$$

where

$$A_1(t) = 2\iint_{0<u<v<t} \frac{\lambda_0(u)\lambda_0(v)}{1 - G(u-)}dudv\mathbb{E}[\exp(2\theta_0^\top X)], \tag{5}$$

$$A_2(t) = \int_0^t \frac{\mathbb{E}[dN(u)]}{(1 - G(u-))^2}. \tag{6}$$

We now compute

$$\int_0^v \frac{\lambda_0(u)}{1 - G(u-)}du = 2 \times \frac{\gamma}{\beta^2}\int_0^v \frac{u}{\gamma - u}du = 2 \times \frac{\gamma}{\beta^2} \times \left(\gamma \log\left(\frac{\gamma}{\gamma - v}\right) - v\right),$$

where we replaced λ_0 by the hazard of a Weibull distribution with shape parameter $\alpha = 2$, scale parameter β and the last equality was obtained from the change of variables $w = \gamma - u$. We then need to compute the following integral in $A_1(t)$:

$$\begin{aligned}
\int_0^t \left(\gamma \log\left(\frac{\gamma}{\gamma - v}\right) - v\right)\lambda_0(v)dv &= \frac{2}{\beta^2}\int_0^t \left(\gamma \log\left(\frac{\gamma}{\gamma - v}\right) - v\right)vdv \\
&= \frac{2}{\beta^2}\left(\frac{\gamma t^2}{2}\log(\gamma) - \gamma\int_0^t v\log(\gamma - v)dv - \frac{t^3}{3}\right).
\end{aligned}$$

The last integral is computed using integration by parts and then by using the change of variables $w = \gamma - v$:

$$\begin{aligned}
\int_0^t v\log(\gamma - v)dv &= \int_0^t \frac{v^2}{2}\frac{dv}{\gamma - v} + \frac{t^2}{2}\log(\gamma - t) \\
&= \int_{\gamma - t}^{\gamma} \frac{(\gamma - w)^2}{2w}dw + \frac{t^2}{2}\log(\gamma - t) \\
&= \frac{\gamma^2}{2}\log\left(\frac{\gamma}{\gamma - t}\right) - \gamma t + \frac{\gamma^2}{4} - \frac{(\gamma - t)^2}{4} + \frac{t^2}{2}\log(\gamma - t) \\
&= \frac{t^2 - \gamma^2}{2}\log(\gamma - t) - \frac{\gamma t}{2} - \frac{t^2}{4} + \frac{\gamma^2}{2}\log(\gamma).
\end{aligned}$$

Gathering all the parts in $A_1(t)$, we have:

$$\begin{aligned}
&A_1(t) \\
&= \frac{8\gamma}{\beta^4}\left(\frac{\gamma t^2}{2}\log(\gamma) - \frac{\gamma}{2}(t^2 - \gamma^2)\log(\gamma - t) + \frac{\gamma^2 t}{2} + \frac{\gamma t^2}{4} - \frac{\gamma^3}{2}\log(\gamma) - \frac{t^3}{3}\right)\mathbb{E}[\exp(2\theta_0^\top X)].
\end{aligned}$$

On the other hand, computation of $A_2(t)$ is straightforward, using the relation (see Section 1) $\mathbb{E}[dN(t) \mid X(t)] = (1 - G(t-))\lambda^*(t \mid X(t))dt$. For $t < \gamma$, we have

$$\begin{aligned} A_2(t) &= \int_0^t \frac{\lambda_0(t)}{1 - G(t-)} dt \mathbb{E}[\exp(\theta_0^\top X)] \\ &= 2 \times \frac{\gamma}{\beta^2} \times \left(\gamma \log \left(\frac{\gamma}{\gamma - t} \right) - t \right) \mathbb{E}[\exp(\theta_0^\top X)]. \end{aligned}$$

Finally, according to Equation (13) of the main paper, we need to compute $A_3(t) = \mathbb{E}\left[(\mu^*(t \mid X(t)))^2\right]$. From Equation (11) of the main paper, we directly have

$$A_3(t) = \left(\frac{t}{\beta}\right)^{2\alpha} \mathbb{E}[\exp(2\theta_0^\top X_i)].$$

To conclude, $A(t) = A_1(t) + A_2(t) - A_3(t)$ and the terms involved in this equation including $\mathbb{E}[\exp(\theta_0^\top X_i)]$ or $\mathbb{E}[\exp(2\theta_0^\top X_i)]$ can easily be computed using Monte-Carlo simulations.

4 Supplementary simulations in the context of right-censored data with no terminal event

Using the same simulation setting as in Section 5.1 of the main paper, we compare the performance of four different models based on the Cox and Aalen models, implemented using either only the first covariate $X_{i,1}$ or the two covariates $X_{i,1}$ and $X_{i,2}$. Figure 2 displays the prediction scores for 100 training samples of size 50 and a unique test sample of size $1,000$. The reference model is the one that uses no covariates and is estimated from the Nelson-Aalen estimator. Roughly, we see that all models have a better prediction performance than the Nelson-Aalen estimator as time increases especially from time equal to 1.5 and time equal to 2, for the models with one covariate and the models with two covariates, respectively. The models with two covariates clearly outperform the ones with one covariate with a slightly better performance for the Cox model as compared to the Aalen model. This is further illustrated in Table 1 where we compare the mean score of those four different models based on 500 training samples of size 20 and 50 and one single test sample of size $1,000$. We clearly observe that the correctly specified Cox model with two covariates outperforms all other models on average, for all time points and sample sizes. However, it tends to have a slightly bigger standard deviation, especially for $n_{\text{train}} = 20$ and $t = 2$ or $t = 2.9$. We have also computed the best prediction score that could be attained, using the correctly specified Cox model with two covariates. In that case, our prediction score reduces to the difference of MSE between the reference and the correctly specified model. Since the latter is equal to 0, the prediction score is equal to the mean squared error between $\mu^*(t \mid X(t))$ and $\mu_0(t)$, where μ_0 is the expected cumulative number of recurrent events for the reference model. Since we are using a non-parametric estimator for $\widehat{\mu_0}$ it converges towards $\mu_0(t) = (t/\gamma)^\alpha \mathbb{E}[\exp(\theta_0^\top X)]$. To conclude, for the correctly specified model, the prediction score converges towards

$$\mathbb{E}\left[\Big(\mu^*(t \mid X(t)) - \mu_0(t)\Big)^2\right] = \left(\frac{t}{\gamma}\right)^{2\alpha} \mathbb{V}[\exp(\theta_0^\top X)].$$

This quantity can be computed using Monte-Carlo simulations and is equal to 1.7, 27.7, 122.4 at times $t = 1$, 2, 2.9, respectively.

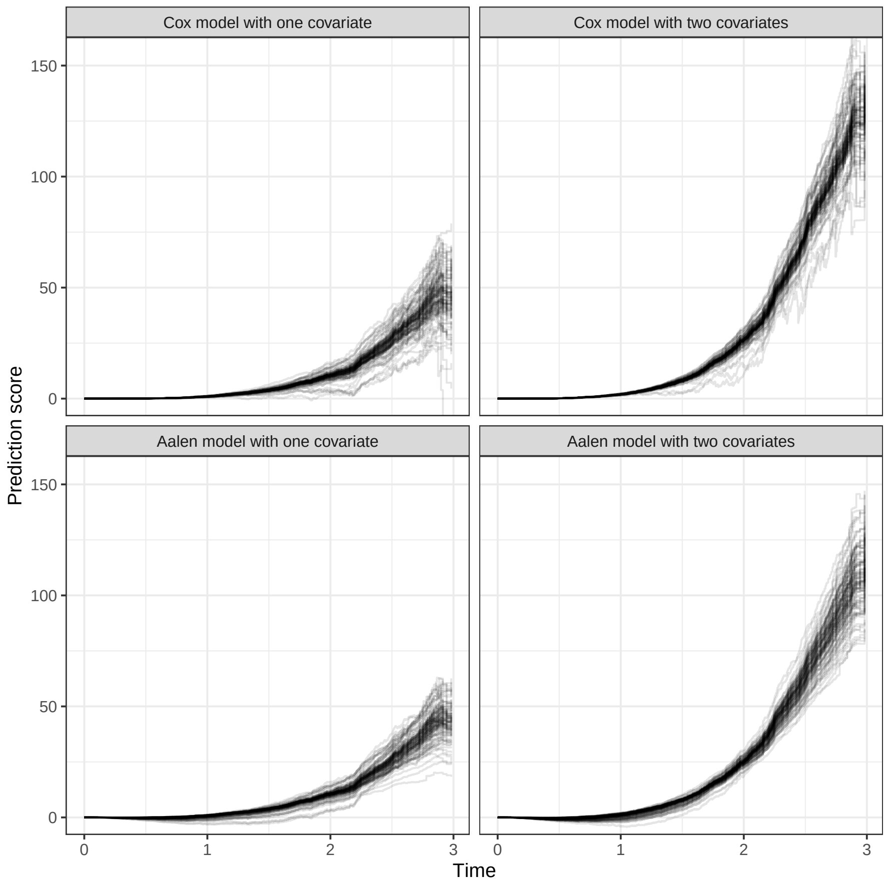

Figure 2: Prediction scores (see Equation (4) of the main paper) using four different models. The data were generated from a Cox model with two covariates and the expected cumulative number recurrent of events was predicted using the Cox model with one covariate, the Cox model with two covariates, the Aalen model with one covariate, the Aalen model with two covariates, respectively. The reference model uses no covariates and was estimated from the Nelson-Aalen estimator. The prediction scores are computed for 100 training samples of size $n_{\text{train}} = 50$ and a unique test sample of size $n_{\text{test}} = 1,000$.

5 Sensibility analysis for the estimation of the censoring distribution on simulated data

In this section, we study the effect of the misspecification of the censoring model. For simplicity, we consider the no terminal event situation of Section 5.1 in the main paper, but this time the censoring times are generated following a Cox model. We use the same two covariates as the one used to generate the recurrent events, but with a different effect β_0^C, and the baseline is supposed to be exponential:

$$\lambda^C(t \mid X_i) = \lambda_0^C \exp(X_i^\top \beta_0^C),$$

with $\lambda_0^C = 1/3$ and $\beta_0^C = (\log(0.8), \log(1.5))^\top$. Using those parameters, we observe 0 or 1 recurrent event for 45% of the individuals, less than or equal to 5 events for 64% of the individuals, and less than or equal to 12 events for 76% of the individuals. On average, we observe approximately 16.7 recurrent events per individual.

In Figure 3 the prediction criterion (see Equation (1) of the main paper) is computing using four different censoring models based on 100 training samples of size $n_{\text{train}} = 20$ and a unique test sample of size $n_{\text{test}} = 2,000$. The expected cumulative number of recurrent of events was

	$n_{\text{train}} = 20$			$n_{\text{train}} = 50$		
	$t = 1$	$t = 2$	$t = 2.9$	$t = 1$	$t = 2$	$t = 2.9$
Cox one cov.	0.89 (0.33)	9.17 (6.04)	40.6 (54.93)	0.96 (0.15)	10.06 (2.19)	46.93 (16.06)
Cox two cov.	1.8 (0.56)	24.69 (8.58)	119.11 (45.56)	1.95 (0.17)	26.33 (2.54)	127.52 (17.73)
Aalen one cov.	0.14 (1.33)	9.53 (4.4)	42.72 (16.27)	0.34 (0.82)	10.09 (1.99)	44.7 (9.21)
Aalen two cov.	0.44 (1.76)	24.18 (4.8)	103.78 (25.28)	0.51 (1.22)	25.13 (1.9)	109.36 (14.21)

Table 1: Means and standard deviations (in bracket) over 500 simulations for the prediction score of the expected number of recurrent events. Large values indicate better predictive performances. Using Monte-Carlo experiments, the best predictive values for times $t = 1, 2, 2.9$ are equal to 1.7, 27.7 and 122.4, respectively

predicted using the correct Cox model but the prediction criterion in Equation (1) of the main paper is implemented using either the true censoring distribution, the Kaplan-Meier estimator, the correct Cox model with estimated parameters or the random survival forest from [8]. The mean over 500 training samples where the censoring cumulative distribution function was calculated using the true censoring distribution was computed and is represented by a dashed line on all four plots. We clearly see that using the correct specification of the Cox model gives reasonable estimates with a very slight overestimation of the prediction criterion while the random survival forest also provides very accurate estimates with a very slight underestimation of the prediction criterion. On the other hand, the Kaplan-Meier estimator that omits the covariate effects on the censoring distribution performs poorly: there is a clear overestimation of the estimated prediction criterion.

6 Prediction of the terminal event for the Atrial Fibrillation dataset

In this section, we evaluate the prediction performance for the terminal event in the Atrial Fibrillation dataset, with the Cox models with age only and with age, AF type and diabetes, the Aalen model with age and the random survival forests with age. The reference model is taken as the Kaplan-Meier estimator and the score is computed using formula (4) of the main paper, where the prediction criterion $\widehat{\text{MSE}}$ is computed from the Kaplan-Meier estimator of the censoring variable. The random survival forests were implemented from the `rfsrc` package (see [8]). The results are presented in Table 2 and Figure 4. We clearly see that the survival random forests perform poorly, especially before time 1 000 where the Kaplan-Meier shows a better performance. The Cox model with age, AF type and diabetes shows a better performance for all time points and the Aalen and Cox models with age show very similar performances and outperform all four models. Other models were also investigated with the different combinations of all three variables with each algorithm and the results were similar and are therefore omitted. In the main manuscript, we then decided to use the Cox model with age for the modelling of the terminal event.

	$t = 1000$	$t = 1500$	$t = 2000$
Aalen with age	0.011 $[0.002, 0.017]$	0.015 $[0.001, 0.039]$	0.038 $[-0.001, 0.035]$
Cox with age	0.010 $[-0.006, 0.023]$	0.016 $[-0.001, 0.042]$	0.037 $[-0.005, 0.053]$
Cox with age, AF type, diabetes	0.006 $[-0.009, 0.019]$	0.011 $[-0.007, 0.04]$	0.027 $[-0.002, 0.04]$
RSF with age	0.005 $[-0.021, 0.032]$	0.010 $[-0.014, 0.039]$	0.010 $[-0.031, 0.045]$

Table 2: Means and 80% intervals (in curly bracket) over 10-folds cross validation for the prediction score of the survival function of the terminal event in the atrial fibrillation dataset. With the Kaplan-Meier estimator as the reference, four different models are compared at three different time points: the Aalen and Cox models with covariate age, the Cox model with covariates age, AF type and diabetes and the Random Survival Forest model with covariate age.

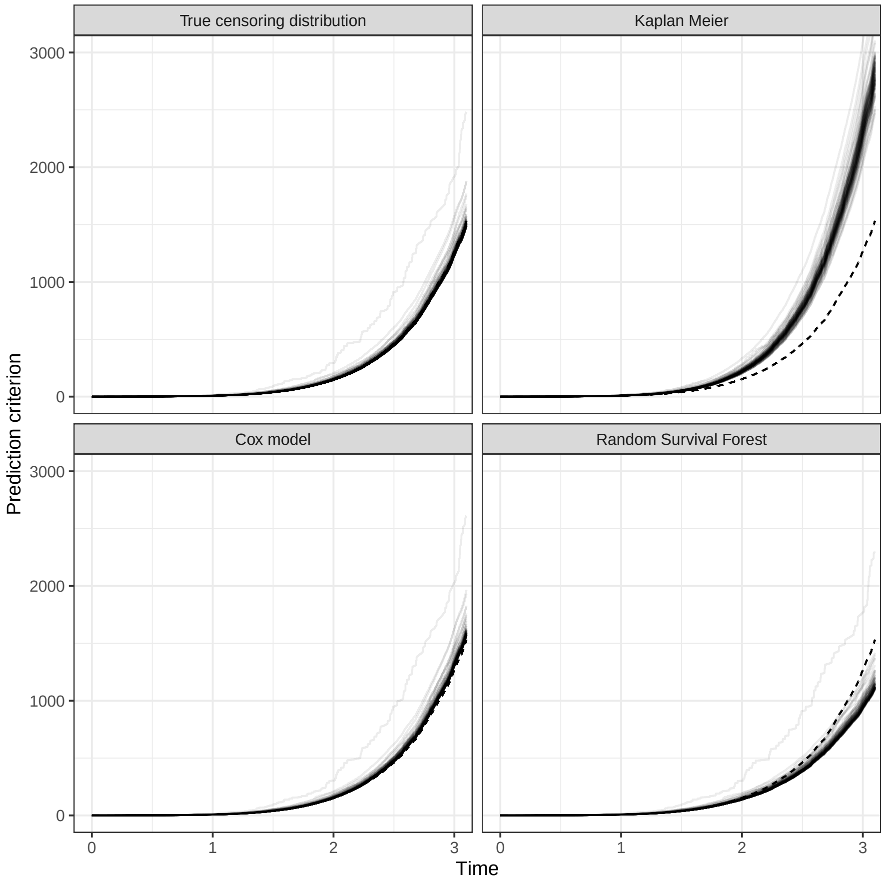

Figure 3: Prediction criterions (see Equation (1) of the main paper) using four different censoring models. The recurrent event data and the censoring distribution were generated from a Cox model with two covariates. The expected cumulative number recurrent of events was predicted using the correct Cox model but the prediction criterion in Equation (1) of the main paper is implemented with four different censoring models: with the true censoring distribution, with the Kaplan-Meier estimator, with the correct Cox model (but the parameters are estimated) or with the random survival forest. The prediction criterions are computed for 100 training samples of size $n_{\text{train}} = 20$ and a unique test sample of size $n_{\text{test}} = 2,000$. The mean over 500 training samples was computed when the censoring cumulative distribution function is calculated using the true censoring distribution and is represented by a dashed line on all four plots.

References

[1] Danyu Y Lin, Lee-Jen Wei, I Yang, and Zhiliang Ying. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730, 2000.

[2] Thomas A Gerds and Martin Schumacher. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040, 2006.

[3] Peter Hall, Rodney CL Wolff, and Qiwei Yao. Methods for estimating a conditional distribution function. *Journal of the American Statistical association*, 94(445):154–163, 1999.

[4] Michel Delecroix, Wolfgang Härdle, and Marian Hristache. Efficient estimation in conditional single-index regression. *Journal of Multivariate Analysis*, 86(2):213–226, 2003.

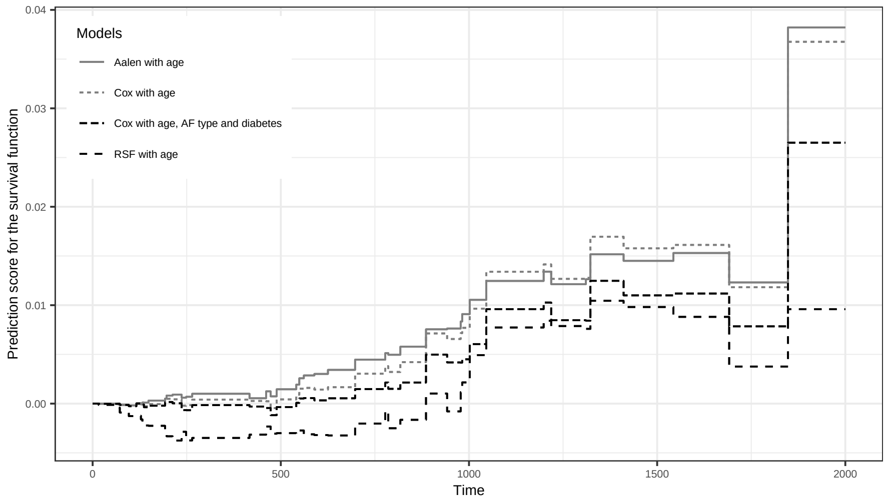

Figure 4: Prediction scores for the survival function of the terminal event in the atrial fibrillation dataset, computed using 10-fold cross validation. With the Kaplan-Meier estimator as the reference, four different models are compared. For ease of visualisation, we describe those models in increasing order of their scores at time $t = 2000$: the Cox model with covariates age, AF type and diabetes (score = 0.067), the Random Survival Forest (RSF) model with covariate age (score = 0.069), the Aalen model with covariate age (score = 0.094) and the Cox model with covariate age (score = 0.105).

[5] Ross L Prentice, Benjamin J Williams, and Arthur V Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.

[6] Olivier Bouaziz and Agathe Guilloux. A penalized algorithm for event-specific rate models for recurrent events. *Biostatistics*, 16(2):281–294, 2015.

[7] Richard J Cook and Jerald Lawless. *The statistical analysis of recurrent events*. Springer Science & Business Media, 2007.

[8] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.