BMC Medical Research Methodology

RESEARCH

Open Access

Development of time to event prediction models using federated learning

Rasmus Rask Kragh Jørgensen[1,2,3*], Jonas Faartoft Jensen[1], Tarec El-Galaly[1,3,4,5], Martin Bøgsted[2,6], Rasmus Froberg Brøndum[2,6], Mikkel Runason Simonsen[1,7] and Lasse Hjort Jakobsen[1]

Abstract

Background In a wide range of diseases, it is necessary to utilize multiple data sources to obtain enough data for model training. However, performing centralized pooling of multiple data sources, while protecting each patients' sensitive data, can require a cumbersome process involving many institutional bodies. Alternatively, federated learning (FL) can be utilized to train models based on data located at multiple sites.

Method We propose two methods for training time-to-event prediction models based on distributed data, relying on FL algorithms, for time-to-event prediction models. Both approach incorporates steps to allow prediction of individual-level survival curves, without exposing individual-level event times. For Cox proportional hazards models, the latter is accomplished by using a kernel smoother for the baseline hazard function. The other proposed methodology is based on general parametric likelihood theory for right-censored data. We compared these two methods in four simulation and with one real-world dataset predicting the survival probability in patients with Hodgkin lymphoma (HL).

Results The simulations demonstrated that the FL models performed similarly to the non-distributed case in all four experiments, with only slight deviations in predicted survival probabilities compared to the true model. Our findings were similar in the real-world advanced-stage HL example where the FL models were compared to their non-distributed versions, revealing only small deviations in performance.

Conclusion The proposed procedures enable training of time-to-event models using data distributed across sites, without direct sharing of individual-level data and event times, while retaining a predictive performance on par with undistributed approaches.

*Correspondence:
Rasmus Rask Kragh Jørgensen
Rasmus.rask@rn.dk
[1] Department of Hematology, Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark
[2] Center for Clinical Data Science, Aalborg University and Aalborg University Hospital, Aalborg, Denmark
[3] Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[4] Department of Medicine Solna, Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden
[5] Department of Hematology, Odense University Hospital, Odense, Denmark
[6] Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark
[7] Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark

Introduction

Clinical prediction models, such as the Cox proportional hazards model, are widely used in medical practice, with many models designed to predict, e.g., a patient's expected survival prospects. The performance of a clinical prediction model often depends on the amount of data available for training the model, with more data yielding improved model performance. Therefore, it is often desirable to pool data from multiple sources, or *sites*, for the development of new clinical predictive models, which can be a cumbersome process for several reasons. First, all input variables for the model may not be available across sites. Second, acquiring sufficient data for rare diseases may require the involvement of many sites, necessitating the navigation of many legal processes, to protect individual-level patient data. As an alternative to pooling site-specific data, federated learning (FL) can be applied to use distributed data to fit a clinical prediction model, without having to pool data at a central repository [1–3].

In recent years, FL methodologies have been developed to fit time-to-event models, particular Cox proportional hazards models. One of the first time-to-event FL methodologies was WebDISCO, a web-based service for the distributed fitting of Cox proportional hazards models [2]. WebDISCO applied an iterative Newton-Raphson approach to update parameter estimates by cycling through sites, and yielded estimates identical to those from a traditional centralized approach with pooled data [2, 4]. However, concerns about privacy issues were raised, due to the possibility of data leakage when risk-sets comprise of a single subject, as this enabled re-identification of the individual [5]. To safeguard against privacy violations, an encryption domain with secret sharing was added to the WebDISCO procedure [6]. Additionally, these methods involve an initial sharing of event times and event indicators to form risk-sets, as the Cox partial likelihood is non-separable [5]. This means that the Cox partial log-likelihood cannot be written as a sum of terms, each depending on a single subject. To mitigate data leakage issues, it was proposed that a risk-set includes a minimum of five individuals, with aggregated event times and added noise [7]. This led to the use of aggregated event times for aligning the risk-sets across sites.

It has also been suggested to combine surrogate partial likelihoods with weight-based methods, rather than directly maximizing the Cox partial likelihood. For example, the One-shot Distributed Algorithm Cox model (ODAC) iteratively aggregates locally fitted models using a weighted average [8]. Another study proposed a weight-based integrated Cox model (WICOX) in which local loss functions are estimated by repeated local data-splitting combined with variance-adjusted weights [9]. These methods provide a framework for survival predictions by weighting each site's model prediction to form a global prediction. However, both the ODAC and WICOX framework require an initial sharing of person sensitive event times. One-shot FL methods have gained increased attentions, as these algorithms aggregates models after a single round of communication, thereby improving communication efficiency, although assuming reliable local convergence across all sites. In a time-to-event setting, this assumption is often violated due to data heterogeneity, such as skewed non-independent and identically distributed data (e.g., varying censoring rates) [10].

Previous studies of FL with time-to-event models have generally focused on using coefficient estimation or weighted aggregation of site-specific models to predict survival curves. However, these methods often require sharing of sensitive outcome information, either by directly aligning the risk-sets at each unique event time, or by sharing the event indicator and event ordering. Methods to circumvent this problem have been investigated, converting the time-to-event data into a binary classification problem with time as a covariate. This methodology enables approximation of the non-separable Cox partial likelihood function using a separable discrete-time model [5]. Notably, by finely splitting the data and specifying a single rate parameter for each time interval, the regression parameter from the Cox model can be estimated using a Poisson regression [11]. Another method for using aggregated gradients across sites has been shown to optimize a site-stratified Cox partial likelihood [5]. Since many of the previously mentioned studies involve an initial sharing of event times, estimation of the survival function would be straightforward. However, even if participating sites were only allowed to share outcomes among each other but not publicly, it would still be possible to make inferences about individuals' event times from a predicted survival curve [12]. Notably, the alignment of risk-sets and predicted survival curves, can be interpreted as sharing sensitive information, when sensitive information is defined as data that can be directly traced to an individual, which would include event times and indicators [13].

The purpose of the present study was to introduce a new method in which FL is used to build time-to-event prediction models without pooling individual-level data across sites. In this method, each site includes full information on its own share of subjects (horizontal data) and only allows the sharing of non-informative data, such as parameter estimates or other aggregated information. Specifically, our approach focuses on enabling time-to-event predictions without explicitly exposing event times.

Method

In this section, we consider right-censored time-to-event outcomes (T, Δ) dependent on covariates x, where the observed time $T = \min(E, C)$ is either the event time E or censoring time C, whichever comes first, and with status indicator $\Delta = 1[E \leq C]$. Suppose that for $n \in \mathbb{N}$ individuals we have observed covariates and survival outcomes $(t, \delta, x) = (t_i, \delta_i, x_i)_{i=1,\ldots,n}$. For simplicity we assume that the t_i's are distinct event times. In a Cox proportional hazards model the hazard function takes the semi-parametric form as follows:

$$h(t|x) = h_0(t)\exp(x^\intercal \beta), \tag{1}$$

where β is a vector of regression coefficients [14]. Importantly, the baseline hazard $h_0(t)$ is unspecified, and is not dependent on any covariates. The model parameter β can be estimated from the observed data using Cox partial likelihood, which in the case of no ties is given as follows:

$$L(\beta|(t, \delta, x)) = \prod_{i \in \mathcal{D}} \frac{\exp(x_i^\intercal \beta)}{\sum_{j \in \mathcal{R}_i} \exp(x_j^\intercal \beta)}, \tag{2}$$

where $\mathcal{D} = \{i \in \{1, 2, \ldots, n\} | \delta_i = 1\}$ represents the indices for individuals experiencing the event, and $\mathcal{R}_i = \{j \in \{1, 2, \ldots, n\} | t_j > t_i\}$ represents the risk-set comprising individuals still at risk at time t_i. Usually, the partial log-likelihood

$$l(\beta) = \sum_{i \in \mathcal{D}} \left(x_i^\intercal \beta - \log\left(\sum_{j \in \mathcal{R}_i} \exp(x_j^\intercal \beta) \right) \right) \tag{3}$$

is maximized to obtain the parameter estimate $\hat{\beta}$ [15]. In the context of distributed data, optimization of the partial log-likelihood requires sharing of event times between sites to build risk-sets, potentially enabling privacy violations. The predicted survival probability at time t for an individual i with covariates x_i can then be estimated as follows:

$$\begin{aligned} \hat{S}(t|x_i, \hat{\beta}) &= \exp\left(-\int_0^t \hat{h}_0(u)\exp(x_i^\intercal \hat{\beta}) du \right) \\ &= \hat{S}_0(t)^{\exp(x_i^\intercal \hat{\beta})}, \end{aligned} \tag{4}$$

where $\hat{S}_0(t) = \exp(-\hat{H}_0(t))$ is the baseline survival function. The cumulative baseline hazard $H_0(t)$ at time t is commonly estimated using the Breslow estimator [16], which produces a step function that increases at all observed event times before t, given as

$$\hat{H}_{0,BR}(t|\hat{\beta}) = \sum_{\substack{i \in \mathcal{D}, \\ t_i \leq t}} \left(\frac{1}{\sum_{j \in \mathcal{R}_i} \exp(x_j^\intercal \hat{\beta})} \right). \tag{5}$$

Thus, even in cases where the estimates of the regression coefficients in the Cox model can be obtained without pooling data across sites, sharing of event times and risk-sets is still necessary to allow prediction of survival probabilities. To overcome this problem, one could instead apply a kernel-based estimator of the baseline hazard function, such as the Ramlau-Hansen estimator [17], to obtain a smoothed version of the baseline hazard function. The formulation of a kernel estimator of the baseline hazard function [18] is given as follows:

$$\begin{aligned} \hat{h}_{0,RH}(t|\hat{\beta}) &= \frac{1}{b} \int \mathcal{K}\left(\frac{t-u}{b} \right) d\hat{H}_{0,BR}(u|\hat{\beta}) \\ &= \frac{1}{b} \sum_{i \in \mathcal{D}} \mathcal{K}\left(\frac{t-t_i}{b} \right) \frac{1}{\sum_{j \in \mathcal{R}_i} \exp(x_j^\intercal \hat{\beta})}, \end{aligned} \tag{6}$$

where $\mathcal{K}$ is a kernel function, such as the Gaussian kernel $\mathcal{K}(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$, which will be used throughout this paper. Furthermore, b is a non-negative bandwidth determining the degree to which event times further away from t should contribute to the estimate. Additional considerations and sensitivity analysis regarding the selection of the bandwidth parameter are discussed in the supplementary material (Figure S10). The smoothed version, $\hat{h}_{0,RH}(t|\hat{\beta})$, can thus be shared without directly exposing event times. Subsequently, estimates of the cumulative baseline hazard $\hat{H}_{0,RH}(t|\hat{\beta})$ can be obtained by integration of $\hat{h}_{0,RH}(t|\hat{\beta})$.

Federated learning for the Cox model

In this section, we aim to explore how FL can be applied with a Cox model to make accurate time-to-event predictions, without sharing event times across sites. Our procedure is two-fold. First parameter estimates $\hat{\beta}$ are obtained. Second these estimates are used to produce survival predictions by applying the above-mentioned kernel smoother. Suppose that the data are horizontally distributed across $S \in \mathbb{N}$ sites (with all sites having access to the same covariates), and n_s denotes the number of subjects at site $s = 1, \ldots S$, then $(t_{s,i}, \delta_{s,i}, x_{s,i})$ will be the observed data for individual i at site s.

Estimation of regression coefficients

If the likelihood function is separable with respect to site—i.e., if the total likelihood can be written as the product of site-specific likelihoods—then β can be estimated by sharing the site-specific likelihood term in a iterative optimization procedure, e.g., Newton-Raphson. For the Cox partial likelihood,

$$\begin{aligned} L(\beta) &= \prod_{s=1}^{S} \prod_{i=1}^{n_s} L_{s,i}(\beta|(t, \delta, x)) \\ &= \prod_{s=1}^{S} \prod_{i=1}^{n_s} \left(\frac{\exp(x_{s,i}^\intercal \beta)}{\sum_{j \in \mathcal{R}_i} \exp(x_{s,j}^\intercal \beta)} \right)^{\delta_i}, \end{aligned} \tag{7}$$

the risk set, $\mathcal{R}_i$, includes all participants at risk regardless of site, such that the partial likelihood is not immediately separable. If $\mathcal{R}_i$ is replaced with $\mathcal{R}_{s,i} = \{j \in \{1, 2, \ldots, n_s\} | t_{s,j} > t_{s,i}\}$, i.e., the site specific risk sets, this makes the likelihood separable across sites and corresponds to using the likelihood function of a site-stratified Cox model, as previously proposed [5, 19]. To estimate the parameters of the stratified Cox model, we utilize Newton-Raphson optimization based on weighted gradients and Hessians of the site-specific log partial likelihoods to mimic a federated stochastic descent methodology [20]. Specifically, the weighted gradients and Hessians are given as follows:

$$\overline{\nabla l(\hat{\beta}_{FL})} = \sum_{s=1}^{S} w_s \cdot \nabla l_s(\hat{\beta}_{FL}) \tag{8}$$

$$\overline{\nabla^2 l(\hat{\beta}_{FL})} = \sum_{s=1}^{S} w_s \cdot \nabla^2 l_s(\hat{\beta}_{FL}). \tag{9}$$

Here, w_s are weights satisfying the criteria $w_s > 0$ and $\sum_{s=1}^{S} w_s = 1$, which determine the contribution of each local site. Throughout, w_s is calculated as $w_s = n_s/n$. These can then be applied in a Newton-Raphson update as follows:

$$\hat{\beta}_{FL}^{p+1} = \hat{\beta}_{FL}^{p} - \left(\overline{\nabla^2 l(\hat{\beta}_{FL}^{p})}\right)^{-1} \overline{\nabla l(\hat{\beta}_{FL}^{p})}, \tag{10}$$

where p is the current iteration. After each update, $\hat{\beta}_{FL}^{p+1}$ is re-distributed to all sites for a new round of iterations. This process is repeated until $||\hat{\beta}_{FL}^{p+1} - \hat{\beta}_{FL}^{p}|| < \varepsilon$ for some small value $\varepsilon > 0$, and formalized in Algorithm 1.

Algorithm 1 FL-Cox β estimation

Initialize and distribute $\hat{\beta}_{FL}^{0}$, $w_s = n_s/n$
while $||\hat{\beta}_{p+1} - \hat{\beta}_{p}|| \geq \varepsilon$ **do**
  for s in 1:S **do**
    Compute $\nabla l_s(\hat{\beta}_{FL}^{p})$ and $\nabla^2 l_s(\hat{\beta}_{FL}^{p})$
    Share $\nabla l_s(\hat{\beta}_{FL}^{p})$ and $\nabla^2 l_s(\hat{\beta}_{FL}^{p})$ with the central site
  end for
  Compute $\overline{\nabla l(\hat{\beta}_{FL}^{p})} = \sum_{s=1}^{S} w_s \cdot \nabla l_s(\hat{\beta}_{FL}^{p})$
  Compute $\overline{\nabla^2 l(\hat{\beta}_{FL}^{p})} = \sum_{s=1}^{S} w_s \cdot \nabla^2 l_s(\hat{\beta}_{FL}^{p})$
  Update $\hat{\beta}_{FL}^{p+1} = \hat{\beta}_{FL}^{p} - \left(\overline{\nabla^2 l(\hat{\beta}_{FL}^{p})}\right)^{-1} \overline{\nabla l(\hat{\beta}_{FL}^{p})}$
  Redistribute $\hat{\beta}_{FL}^{p+1}$ to each site
end while

After convergence, the next step is to use the kernel-smoother to enable survival predictions without sharing events times.

Baseline hazard smoothing

Applying the kernel estimator in Eq. (6) smooths the local baseline hazard function, allowing each site to expose its respective baseline hazards without explicitly sharing event times. More precisely, each site s estimates $\hat{h}_{0,RH,s}(t|\hat{\beta})$ for a predefined grid of time-points t according to (6) where $\hat{\beta}$ are the coefficients estimated from Algorithm 1. The central site aggregates estimates $\hat{h}_{0,RH,s}(t|\hat{\beta})$ for each site and for each pre-defined time point by a weighted average. A regression smoother, e.g., a loess curve is used to model the aggregated estimates to form a smooth global baseline hazard function $\hat{h}_{0,L}(t)$ [21]. The procedure is outlined in Algorithm 2.

Algorithm 2 Baseline hazard function estimation [22]

Distribute $\hat{\beta}$, and a pre-defined grid of time-values $t \in \{t_1, t_2, \ldots, t_{max}\}$
for s in 1:S **do**
  Compute $\hat{h}_{0,RH,s}(t|\hat{\beta})$ by Equation (6) at each t
  Distribute $\hat{h}_{0,RH,s}(t|\hat{\beta})$ to the central aggregator
end for
The central site aggregates estimates of $\overline{\hat{h}_{0,RH}(t|\hat{\beta})} = \sum_{i=1}^{S} w_i \cdot \hat{h}_{0,RH,i}(t|\hat{\beta})$ at each time-point t, and modeled using loess regression after log-transforming to ensure strictly positive hazard [22]
Compute $\hat{H}_{0,L}(t) = \int_0^t \hat{h}_{0,L}(u)du$, for each t
Use $\hat{H}_{0,L}(t)$ and $\hat{\beta}_{FL}$ to form survival predictions using Equation (4)

Time-points can be selected arbitrarily (e.g., $10, 20, 30, \ldots$ days), depending on the desired granularity of the baseline hazard function. A straight-forward approach used throughout, is an evenly spaced grid of time-points, e.g., $0, 20, 40, \ldots, 1825$, which serves as our prediction horizon, i.e predicted survival curve. Ideally, the grid of time-points should be chosen to allow the capture of sufficient changes in the baseline hazard over time. This can be accommodated using expert knowledge or by estimating the average distance between event times across sites, to build an equidistant grid of time points.

Federated learning with a parametric model

The general likelihood for a parametric survival model [23] based on n right-censored individuals is as follows:

$$L(\beta) = \prod_{i=1}^{n} S(t_i) h(t_i)^{\delta_i}. \tag{11}$$

In a distributed setting, it is possible to factorize across sites, and the likelihood in (11) can then be rewritten as follows:

$$L(\beta) = \prod_{s=1}^{S} \prod_{i=1}^{n_s} S(t_{s,i}) h(t_{s,i})^{\delta_{s,i}}. \tag{12}$$

Thus, parametric survival models can be fitted if each site shares the corresponding site-specific term of the likelihood, which does not require sharing of individual-level data (if the sites consists of more than one subject). Both simple parametric survival models, like the Weibull or exponential model and more flexible spline-based models can be used with this approach. The spline-based survival models described by Carstensen enables fitting of flexible survival models using ordinary Poisson regression [11]. For a Poisson regression model one can split the entire follow-up into smaller pre-determined time intervals, and assume constant hazards within each sub-interval by modeling the timescale as a covariate [11]. Specifically, the time bins should be selected finely enough to provide an adequate description of the data [11]. In particular, we model the effect of the timescales using splines, which requires selection of the number of knots and knot positions. The proposed methodology for a FL-Poisson model is formalized in Algorithm 3.

Algorithm 3 FL-Poisson model

Initialize $\hat{\beta}^0_{FL}$, pre-defined time grid $t \in \{t_1, t_2, \ldots, t_{max}\}$, number of knots and knot positions
while $||\hat{\beta}^{p+1}_{FL} - \hat{\beta}^p_{FL}|| \geq \varepsilon$ **do**
  for s in 1:S **do**
    Estimate $\nabla l_s(\hat{\beta}^p_{FL})$ and $\nabla^2 l_s(\hat{\beta}^p_{FL})$ based on first- and second order derivatives of (12)
    Distribute $\nabla l_s(\hat{\beta}^p_{FL})$ and $\nabla^2 l_s(\hat{\beta}^p_{FL})$ to the central site
  end for
  Compute $\nabla l(\hat{\beta}^p_{FL}) = \sum_{s=1}^{S} \nabla l_s(\hat{\beta}^p_{FL})$
  Compute $\nabla^2 l(\hat{\beta}^p_{FL}) = \sum_{s=1}^{S} \nabla^2 l_s(\hat{\beta}^p_{FL})$
  Update $\hat{\beta}^{p+1}_{FL} = \hat{\beta}^p_{FL} - \left(\nabla^2 l(\hat{\beta}^p_{FL})\right)^{-1} \nabla l(\hat{\beta}^p_{FL})$
  Redistribute $\hat{\beta}^{p+1}_{FL}$ to each site
end while

Results

To investigate the performance of the proposed FL methods, we used both simulated experiments (each simulated 1000 times) and a real-world data scenario of patients diagnosed with Hodgkin lymphoma. The set up was intended to mimic a predictive study design, in which models are trained based on a training dataset, and then validated in an independent external cohort [24].

Simulation specifications

The four simulation experiments were simulated with outcomes conditioned on four independent features, drawn from a standard normal distribution. Specifically, the observed time-to-event, T_E, was simulated from a Weibull distribution as follows:

$$T_E = -\left(\frac{\log(U)}{\lambda \cdot \exp(x^\intercal \beta)}\right)^{1/\upsilon},$$

where $U \sim U(0,1)$, and where λ and υ are the scale and shape parameters, respectively [25]. Throughout, the scale and shape parameter are fixed at $\lambda = 0.5$ and $\upsilon = 2$, respectively. The time to censoring, C, was simulated from an independent uniform distribution $U(0, T_{C_{max}})$, where $T_{C_{max}}$ is the maximum follow-up time. The simulated follow-up time was then $T = \min(T_E, C)$, with status indicator $1[T_E \leq C]$.

In each experiment, the set up differed from the others in either the number of subjects n_s per site, the number of sites, or the maximum length of follow-up $T_{C_{max}}$ across sites. In simulation 1, which mimicked a rare disease setting, each site had access to a small number of subjects. In simulation 2, each site had a varying $T_{C_{max}}$ to investigate scenarios with varying follow-up and censoring mechanisms. Simulation 3 investigated the impact of varying the number of subjects n_s per site. Finally, in simulation 4, both the number of subjects n_s and $T_{C_{max}}$ varied across sites. Table 1 lists the exact simulation specifications.

In Table 1, $n_s \in \{...\}$ and $T_{C_{max}} \in \{...\}$ are notations used to clarify differing numbers of patients and differing lengths of follow-up, respectively, across sites within each experiment. For example, in Simulation 4, site 1 included 20 subjects with a maximum follow-up of 1 year, site 2 had 30 subjects and maximally 2 years of follow-up, etc. For each of the four simulations, an independent external validation cohort of 1000 subjects was simulated with the same fixed underlying covariate effect, and with $T_{C_{max}} = 5$ years. To assess the methods in a heterogeneous population, additional simulations were conducted with varying site-specific changes to the underlying distributions, as detailed in the supplementary material (Tables S1 and S2).

Applied models

In each experiment, four models were fit using different learning schemes and training procedures, and compared to the true underlying model from which the data was simulated. Additionally, it was assumed that the underlying dependent covariates were known, such that no feature selection was considered in this set-up. In total, two versions of the Cox model and two versions of the Poisson model were fitted. All initial values, β_0, were set to 0, for each FL model.

Benchmark models

The baseline models comprised a Cox model and a Poisson model trained in a non-distributed setting, where all data were pooled at a single site. The Cox model will be termed *Cox-Pool*, which is a Cox model estimated by optimizing the regular Cox partial likelihood in (2), and with survival estimates based on (4) using the Breslow estimator from

Table 1 Experiment specifications for each of the four simulation experiments performed

	Simulation 1	Simulation 2	Simulation 3	Simulation 4
Number of sites S	30	10	20	20
Patients per site n_s	20	100	$n_s \in \{20, 30, \ldots, 210\}$	$n_s \in \{20, 30, \ldots, 210\}$
$T_{C_{max}}$ (Years)	5	$T_{C_{max}} \in \{1, 2, \ldots, 10\}$	5	$T_{C_{max}} \in \{1, 2, \ldots, 20\}$
Patient distribution	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(\mu, \sigma^2)$
Mean μ_s	0	0	0	0
SD σ_s^2	1	1	1	1
Covariate effect β				
β_1	log(1.2)	log(1.2)	log(1.2)	log(1.2)
β_2	log(0.8)	log(0.8)	log(0.8)	log(0.8)
β_3	log(2.0)	log(2.0)	log(2.0)	log(2.0)
β_4	log(0.7)	log(0.7)	log(0.7)	log(0.7)
Bandwidth b_s	4	4	4	4

(5). The Poisson model will be termed *Poisson-Pool*, which seeks to optimize the likelihood from (11) in which exact event times are used to form discrete time bins modeled as covariates using splines [11].

FL-Cox model

The FL-Cox model was developed by training a Cox model in a distributed setting, where data were allocated at different sites. The β coefficient for the FL-Cox model was estimated, using the procedure described in Algorithm 1. Additionally, survival predictions were performed by approximating the baseline hazards using a kernel smoother as described in Algorithm 2.

FL-Poisson model

The FL-Poisson model involved optimizing the separable likelihood in (11), following the procedure described in Algorithm 3. As event times cannot be shared across sites, a predefined grid of time-points, $t \in \{0, 20, 40, \ldots, 1825\}$ (up to about 5 years measured in days), was selected to align time bins across sites. The baseline-hazard were modeled using a natural spline with 6 knots (including boundary knots), and knot positions were chosen by averaging evenly spaced quantiles of local event times (plus random noise) ensuring an even distribution across event times, as previously suggested [26].

Performance measures

To accurately compare predictions, each model's predicted survival probability over a specified time period $[0, t']$ was compared to the predictions of the true model as follows:

$$ISD(t'|x_i) = \int_0^{t'} |S_m(u|x_i) - S_{True}(u|x_i)|du, \; i = 1, \ldots, n, \tag{13}$$

where $ISD(t'|x_i)$ is the integrated survival difference over the time period $[0, t']$ for each subject i in the external test. Here, $S_m(u|x_i)$ is the survival predicted by the m'th model (Cox-Pool, Poisson-Pool, FL-Cox, and FL-Poisson), and $S_{True}(u|x_i)$ is the true survival derived from the true model for the same subject with covariate structure x_i. To summarize the difference across subjects, the mean integrated absolute difference (MIAD) was estimated by computing $ISD(t'|x_i)$ for all individuals i, up to time t', and constructing a sample mean as follows:

$$MIAD = \frac{1}{n} \sum_{i=1}^{n} ISD(t'|x_i).$$

Additionally, to compare the predicted survival with the observed survival over the time period $[0, t']$, we used the integrated Brier score (IBS) [27]. To assess each model's ability to rank subjects according to risk, the concordance index (C-index) [28] was estimated using the predicted survival probability at time-point t'.

Simulation results

In all four experiments, we observed no noteworthy difference between the estimated C-index of the fitted models versus the true model across the 1000 simulations (Fig. 1 and Table 2). Similar results were observed for the IBS measure, with almost identical distributions of performance estimates across all simulations (Fig. 2). We noted minor differences in IBS distribution between the true model and the other models; however, the median IBS differences across all simulations were < 0.001 in each experiment for all models (Table 2). Therefore, the predictive performance seemed identical across all

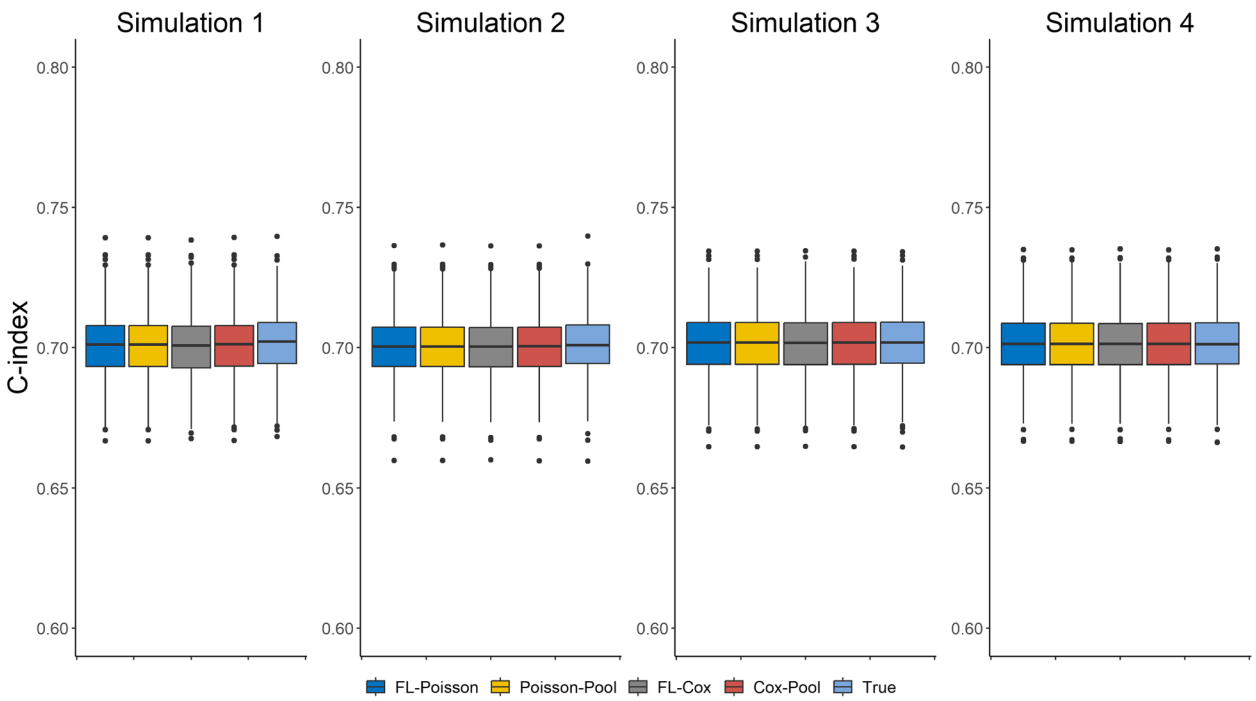

Fig. 1 C-index for each model in each simulation on a simulated external data set containing 1000 subjects. C-index, Concordance-index; FL, Federated learning

Table 2 Performance estimates across the 1000 simulations for each of the four experiments

	Models (median [range])	Simulation 1	Simulation 2	Simulation 3	Simulation 4
C-index	True model	0.702 [0.668, 0.740]	0.701 [0.660, 0.740]	0.702 [0.665, 0.734]	0.701 [0.666, 0.735]
	FL-pois	0.701 [0.667, 0.739]	0.700 [0.660, 0.736]	0.702 [0.665, 0.734]	0.701 [0.667, 0.735]
	Poisson-Pool	0.701 [0.667, 0.739]	0.700 [0.660, 0.737]	0.702 [0.665, 0.734]	0.701 [0.667, 0.735]
	FL-Cox	0.701 [0.668, 0.738]	0.700 [0.660, 0.736]	0.702 [0.665, 0.735]	0.701 [0.667, 0.735]
	Cox-Pool	0.701 [0.667, 0.739]	0.700 [0.660, 0.736]	0.702 [0.665, 0.734]	0.701 [0.667, 0.735]
C-index Diff	FL-Poisson	−0.0007 [−0.0095, 0.0041]	−0.0003 [−0.0115, 0.0027]	0.0000 [−0.0052, 0.0024]	0.0000 [−0.0028, 0.0029]
	Poisson-Pool	−0.0007 [−0.0094, 0.0042]	−0.0003 [−0.0114, 0.0027]	0.0000 [−0.0052, 0.0024]	0.0000 [−0.0029, 0.0029]
	FL-Cox	−0.0009 [−0.0137, 0.0050]	−0.0003 [−0.0151, 0.0036]	−0.0001 [−0.0059, 0.0025]	0.0000 [−0.0033, 0.0025]
	Cox-Pool	−0.0007 [−0.0097, 0.0041]	−0.0003 [−0.0114, 0.0029]	0.0000 [−0.0052, 0.0024]	0.0000 [−0.0029, 0.0029]
IBS	True model	0.1231 [0.1135, 0.1327]	0.1231 [0.1141, 0.1327]	0.1230 [0.1138, 0.1345]	0.1229 [0.1124, 0.1339]
	FL-Poisson	0.1240 [0.1134, 0.1342]	0.1237 [0.1143, 0.1334]	0.1233 [0.1138, 0.1351]	0.1233 [0.1131, 0.1346]
	Poisson-Pool	0.1239 [0.1133, 0.1341]	0.1235 [0.1141, 0.1332]	0.1233 [0.1137, 0.1348]	0.1232 [0.1128, 0.1344]
	FL-Cox	0.1241 [0.1138, 0.1343]	0.1238 [0.1141, 0.1333]	0.1233 [0.1137, 0.1347]	0.1233 [0.1126, 0.1347]
	Cox-Pool	0.1240 [0.1133, 0.1343]	0.1236 [0.1143, 0.1333]	0.1233 [0.1138, 0.1348]	0.1232 [0.1128, 0.1345]
IBS-Diff	FL-Poisson	0.0008 [−0.0008, 0.0040]	0.0005 [−0.0010, 0.0041]	0.0003 [−0.0004, 0.0021]	0.0003 [−0.0006, 0.0032]
	Poisson-Pool	0.0007 [−0.0007, 0.0040]	0.0005 [−0.0007, 0.0036]	0.0002 [−0.0006, 0.0017]	0.0002 [−0.0006, 0.0026]
	FL-Cox	0.0009 [−0.0005, 0.0055]	0.0006 [−0.0006, 0.0039]	0.0003 [−0.0008, 0.0022]	0.0003 [−0.0005, 0.0021]
	Cox-Pool	0.0008 [−0.0005, 0.0043]	0.0005 [−0.0007, 0.0034]	0.0002 [−0.0006, 0.0017]	0.0002 [−0.0006, 0.0025]
MIAD	FL-Poisson	19.97 [6.16, 46.71]	16.97 [5.63, 39.03]	12.61 [3.55, 34.26]	12.46 [3.87, 32.61]
	Poisson-Pool	18.83 [6.06, 42.62]	15.60 [3.69, 36.69]	11.01 [3.46, 29.07]	10.87 [4.03, 28.12]
	FL-Cox	21.18 [7.70, 48.43]	17.88 [6.27, 48.42]	12.79 [4.78, 37.80]	12.87 [3.58, 35.47]
	Cox-Pool	19.68 [6.95, 43.39]	16.32 [7.24, 36.07]	11.24 [3.93, 28.22]	10.98 [4.22, 28.34]

C-index Concordance indexM, *FL* Federated learning, *IBS* Integrated Brier score, *Diff* Difference, *MIAD* Mean integrated absolute difference

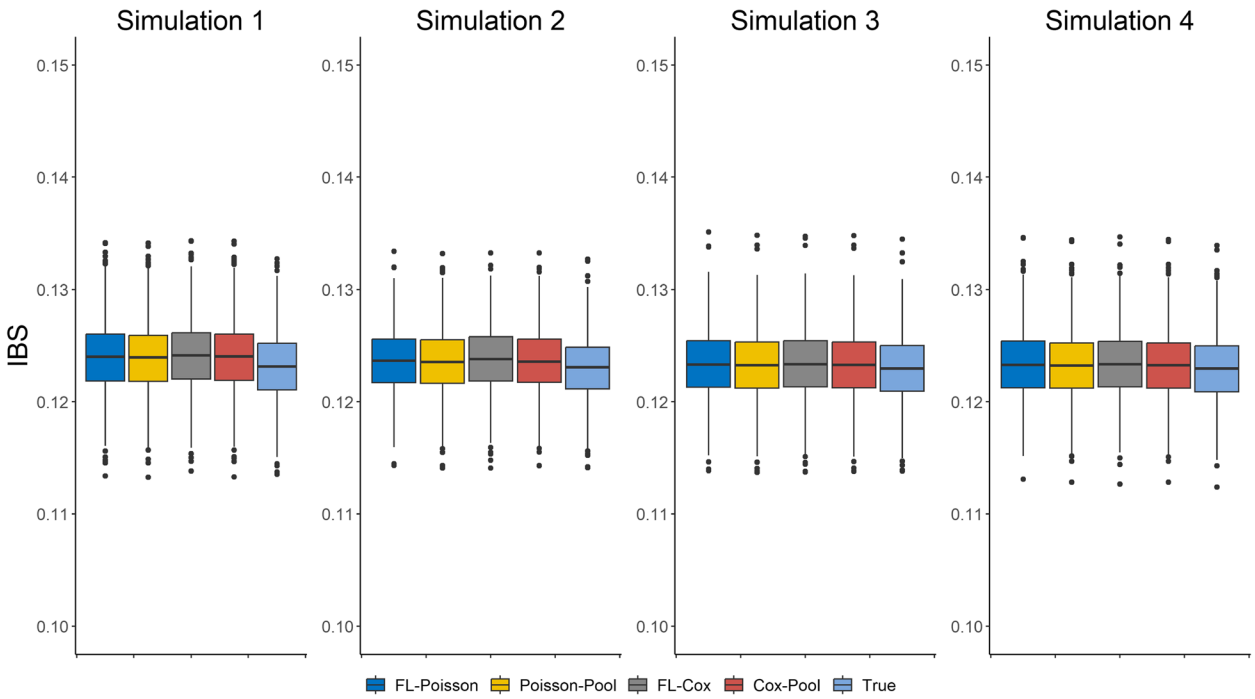

Fig. 2 IBS for each model in each simulation on a simulated external data set containing 1000 subjects. IBS, Integrated Brier score; FL, Federated learning

models, indicating that a preserved ratio between the estimated covariate effects and reasonable risk ranking of subjects using the FL methodology.

The MIAD measure was observed to differ across experiments, and as the total number of subjects increased between experiments, each model's predicted survival probability coincided with the true model (Fig. 3). The Poisson-Pool model in experiment 4 exhibited the smallest median MIAD of 10.87 (range: 4.03–28.12) (Table 2), where as the distributed FL-Poisson model attained a median MIAD of 12.46 (range: 3.87–32.61). In experiment 4, the FL-Cox model exhibited the largest median MIAD of 12.87 (range: 3.58–35.47). The FL-Cox model in experiment 1 showed the overall largest median MIAD of 21.18 (range: 7.70–48.43), and its non-distributed version attained a median MIAD of 19.68 (range: 6.95–43.39). Overall, the MIAD estimates for the pooled models seemed to be systematically lower than for the FL versions, with similar results found across all experiments. These findings, highlighted a negligible difference between the FL versions and the pooled versions. The same negligible difference was found in a more heterogeneous simulated population, as listed in the supplementary material (Figure S1-S6). Analysis of the coefficient estimates for β indicated great alignment between the FL and pooled versions, when comparing to the true coefficient values (Supplementary Figure S7 and Table S3). Also, a negligible difference was found when comparing the coefficient estimates between the FL-Poisson model and the pooled version (Supplementary Figure S8). Across simulation 1–6, the FL-Poisson converged after 13–14 communication rounds, whereas the FL-Cox converged between 5–8 communication rounds (Supplementary material Table S4).

Real-world advanced-stage Hodgkin Lymphoma

To strengthen the applicability of the proposed methods, we performed an analysis using real-world data, including patients derived from the Danish National Lymphoma register (LYFO) [29]. The experiment included 707 patients with advanced-stage Hodgkin lymphoma. Additional details about these patients are described elsewhere [30]. Out of the 707 patients, 569 ($\approx$ 80.5%) were censored, and 138 experienced an event. Patients were randomly split 70/30 into a model development cohort and a separated validation cohort. Additionally, the development cohort was divided into 8 separate sites, to mimic different hospital departments in Denmark. The validation cohort was kept separate from the development cohort and used to estimate the performance of the applied methods by predicting the 5-year survival probability after diagnosis. A total of seven features were considered in this real-world scenario, which were chosen to mimic the features included in the international

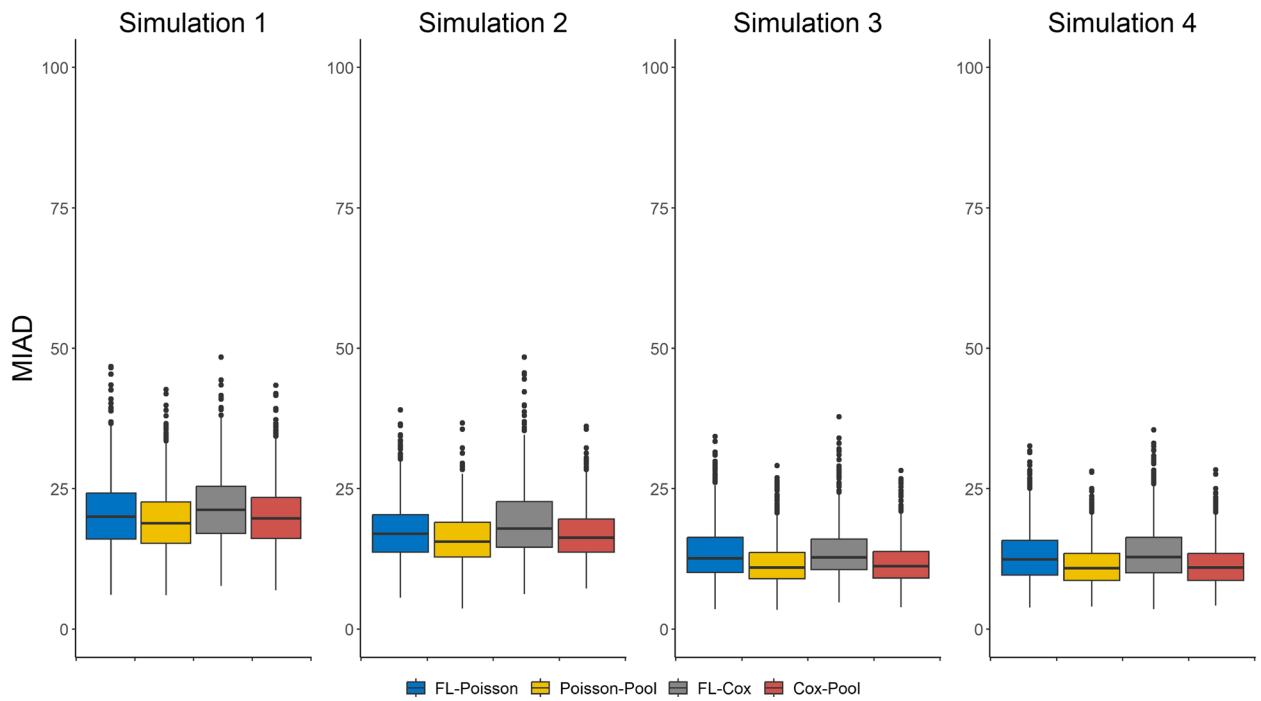

Fig. 3 MIAD for each model in each simulation on a simulated external data set containing 1000 subjects. MIAD, Mean integrated absolute difference; FL, Federated learning

prognostic score (IPS-7) [31]. Since the true underlying model for the real-world data was unknown, the non-distributed versions of each model (Cox-Pool and Poisson-Pool) were compared against the distributed version (FL-Cox and FL-Poisson, respectively).

Real-world results

We assessed the performance of the proposed models applied in a real-world patient cohort, which revealed great alignment of predictive performance between the FL models and the pooled versions. The Poisson-Pool and FL-Poisson attained similar C-index of 0.755 and 0.755, respectively, compared to 0.754 and 0.768 for the Cox-Pool and FL-Cox models, respectively(Table 3). Additionally, no notable differences were observed for the IBS estimates as all models obtained almost identical estimates (Table 3). We compared the individual predicted survival probabilities from the FL-Poisson model against the Poisson-Pool model which yielded an MIAD estimate of 4.12. Comparison between the FL-Cox and Cox-Pool models yielded an MIAD estimated of was 34.28.

Table 3 Performance estimates for the real-world experiment

Measure	Model	Estimates
C-index	FL-Poisson	0.755
	Poisson-Pool	0.755
	FL-Cox	0.768
	Cox-Pool	0.754
IBS	FL-Poisson	0.082
	Poisson-Pool	0.082
	FL-Cox	0.081
	Cox-Pool	0.082
MIAD	FL-Poisson vs. Poisson-Pool	4.12
	FL-Cox vs. Cox-Pool	34.28

C-index Concordance index, *FL* Federated learning, *IBS* Integrated Brier score, *Diff* Difference, *MIAD* Mean integrated absolute difference

Discussion

In the current study, we addressed decomposing the (partial) likelihood of the Cox and Poisson models into site-specific likelihood contributions in order to estimate model parameters in a distributed setting while also including a comment on the use of kernel smoothers to form survival predictions without directly exposing event times. The proposed FL methodologies were compared to a non-distributed version in which data were pooled at a central site, to mimic an optimal case. The loss in predictive performance was minimal with negligible differences, indicating the applicability of the proposed method. Additionally, experiment 1 tested a set-up in which the data mimicked

the setting of a rare disease, which yielded promising results regarding the stability of the proposed method for combining datasets with limited sample sizes.

FL is receiving increasing attention in healthcare research, as a method for enabling interdisciplinary and cross-institutional collaborations while preserving patient privacy. Despite the large potential, the application of FL to survival analysis presents specific methodological challenges, particularly the necessity to share event times across sites [2, 7–9]. Since sensitive information is defined as data that can be directly linked to an individual, event times can be considered sensitive and can potentially invoke application of data protection regulations, such as the General Data Protection Regulation act (GDPR) [13, 32]. The FL landscape is evolving rapidly, with continuous advancements in algorithms, privacy preserving mechanisms, and domain specific adaptations. To address the challenge of shared event times, a kernel based smoother approach was developed to enable anonymous survival predictions in a federated setting without sharing event times. By locally estimating a smoothed baseline hazard and sharing only aggregated information, this method preserves privacy with minimal loss in predictive performance on par with a centralized approach and with similar coefficient estimates.

The current study had several limitations. First, in this study, we did not describe the methods for variable selection, which is an important aspect of predictive studies. Methods involving regularization or dimension reduction techniques have previously been proposed for the Cox model in a FL setup; however, that requires alignment of risk-sets between sites, which would expose event times [3, 33–35]. Second, we did not consider vertical distributed data in which sites does not share attributes, but each site contains the same set of subjects. Since the main focus of this paper was to apply FL using time-to-event data without explicitly exposing event times, we did not apply feature selection techniques. Moreover, the proposed techniques do not ensure any guarantee of privacy against the risk that subjects may be re-identified from the aggregation of gradients and Hessians [36]. However, adding a carefully crafted noise parameter to local estimates could help ensure, e.g., differential privacy. As is inherent to the FL methodology is that an iterative communication between sites and the central aggregator are necessary, of which can have several practical challenges, but these are described elsewhere.

Another limitation of the presented method is the requirement to specify the knot placements, number of knots, and data split required to fit a Poisson model. Also, a similar approach to our Poisson-model, using a site-weighted base generalized Bayesian federated inference framework, has been proposed previously [37]. However, the Bayesian federated inference framework, like other one-shot methods, assumes that the statistical model can be fitted to each site's local data, which may be infeasible when local datasets are small [38]. In general, models can be sensitive to the choice of knot placement, especially in cases where over- or under-fitting is an issue, e.g., when event times and choice of knot position are misaligned. Sensitivity analysis is often recommended when fitting spline based models. However, this further complicates the FL setup requiring additional information sharing [39], and thereby increasing the possibility of exposing sensitive information. The selection of time splits is usually dependent on the event times themselves, i.e., time bins are formed using the exact event times [11]. However, since these could not be shared, a pre-defined grid of time splits was used, leading to a difference between the FL and the pooled models in the real-world comparisons (Table 3). The results presented in Fig. 3 and Table 2 highlighted a difference between the cases in which data were pooled compared to the FL version. However, since the performance estimates for the IBS and C-index were almost identical among all models, the minimal loss of accuracy observed in the MIAD estimates seemed to be within an acceptable range. Overall, the models seemed robust against site heterogeneity with respect to sample size, follow-up time, and distribution of covariates.

A natural next step would be to extend the presented methodology to a competing risks setting, since many clinical prediction models based on time-to-event data are subject to competing events. This could be accomplished by using the presented method on cause-specific hazards, and applying these to obtain the absolute risk on an individual-patient level. Likewise, the methodology could be extended to a recurrent event setting, in which the intensity function is estimated using the Andersen-Gill model, and marginal means are computed using the estimator proposed by Cook and Lawless [40]. As with the Breslow estimator, which exposed event times, it would still be necessary to apply a smoothing process to avoid sharing sensitive information.

This paper introduced new methods for performing FL using time-to-event data, without explicitly exposing event times and status indicators. The results demonstrated the preserved prediction accuracy of the proposed FL models compared to fitting similar models using pooled data. Additionally, the proposed baseline hazard algorithm (Algorithm 2) enabled anonymous survival predictions for the FL-Cox model. Thus, the presented methods enable investigators to develop clinical survival prediction models without sharing sensitive patient data. With the increasing complexity of data sharing, e.g., due to the GDPR, such methods hold great potential for continued development of useful clinical tools.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-025-02598-y.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
RRKJ and LHJ developed the concept and design. RRKJ, JF and MRS developed the methodology. RRKJ, LHJ, TC, JF, MRS, RF, MB did data analysis and interpretation. RRKJ, LHJ and TC, provided the data. RRKJ wrote the main manuscript. All authors reviewed the manuscript and are accountable of all aspect of the work.

Funding
Mikkel Runason Simonsen is supported by the Danish Data Science Academy, which is funded by the Novo Nordisk Foundation (NNF21SA0069429) and VILLUM FONDEN (40516).

Data availability
Data are only accessible through the official data owners in Denmark (https://www.sundk.dk/). Therefore, it is not possible to share the data used in this study.

Declarations

Ethics approval and consent to participate
The study was approved by the North Denmark Region (ID: 2021-002).

Consent for publication
Not applicable.

Competing interests
Rasmus Rask Kragh Jørgensen: No conflicts of interest. Tarec El-Galaly: No conflicts of interest. Jonas Faartoft Jensen: No conflicts of interest. Lasse Hjort Jakobsen: Honoraria, Roche, Novo Nordisk. Mikkel Runason Simonsen: No conflicts of interest. Rasmus Froberg Brøndum: No conflicts of interest. Martin Bøgsted: No conflicts of interest.

Received: 9 January 2025 Accepted: 15 May 2025
Published online: 26 May 2025

References

1. Yang Q, Liu Y, Chen T, Tong Y. Federated Machine Learning: Concept and Applications. ACM Trans Intell Syst Technol. 2019;10(2). https://doi.org/10.1145/3298981.
2. Lu CL, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc. 2015;22(6):1212–9. https://doi.org/10.1093/jamia/ocv083.
3. Späth J, Matschinske J, Kamanu FK, Murphy SA, Zolotareva O, Bakhtiari M, et al. Privacy-aware multi-institutional time-to-event studies. PLoS Digit Health. 2022;1(9):e0000101. https://doi.org/10.1371/journal.pdig.0000101.
4. Nocedal J, Wright SJ. Numerical optimization. 2nd ed. Springer Series in Operations Research and Financial Engineering. Springer New York; 2006. https://doi.org/10.1201/b19115-11.
5. Andreux M, Manoel A, Menuet R, Saillard C, Simpson C. Federated Survival Analysis with Discrete-Time Cox Models. CoRR. 2020. abs/2006.0. http://arxiv.org/abs/2006.08997. Accessed 20 May 2025.
6. Kamphorst B, Rooijakkers T, Veugen T, Cellamare M, Knoors D. Accurate training of the Cox proportional hazards model on vertically-partitioned data while preserving privacy. BMC Med Inform Decis Mak. 2022;22(1):49. https://doi.org/10.1186/s12911-022-01771-3.
7. Juwara L, Yang YA, Velly AM, Saha-Chaudhuri P. Privacy-preserving analysis of time-to-event data under nested case-control sampling. Stat Methods Med Res. 2024;33(1):96–111. https://doi.org/10.1177/09622802231215804.
8. Duan R, Luo C, Schuemie MJ, Tong J, Liang CJ, Chang HH, et al. Learning from local to global: An efficient distributed algorithm for modeling time-to-event data. J Am Med Inform Assoc. 2020;27(7):1028–36. https://doi.org/10.1093/jamia/ocaa044.
9. Park JA, Kim TH, Kim J, Park YR. WICOX: Weight-Based Integrated Cox Model for Time-to-Event Data in Distributed Databases Without Data-Sharing. IEEE J Biomed Health Inform. 2023;27(1):526–37. https://doi.org/10.1109/JBHI.2022.3218585.
10. Liu X, Tang Z, Li X, Song Y, Ji S, Liu Z, et al. One-shot Federated Learning Methods: A Practical Guide; 2025. https://arxiv.org/abs/2502.09104.
11. Carstensen B. Demography and epidemiology: Practical use of the Lexis diagram in the computer age or "Who needs the Cox model anyway?" Technical report. 2006. https://bendixcarstensen.com/rr-06-2.pdf. Accessed 20 May 2025.
12. Wei Y, Royston P. Reconstructing Time-to-event Data from Published Kaplan-Meier Curves. Stata J Promot Commun Stat Stata. 2017;17(4):786–802. https://doi.org/10.1177/1536867X1801700402.
13. O'Keefe CM, Rubin DB. Individual privacy versus public good: Protecting confidentiality in health research. Stat Med. 2015;34(23):3081–103. https://doi.org/10.1002/sim.6543.
14. Cox DR. Regression Models and Life-Tables. J R Stat Soc Ser B Stat Methodol. 1972;34(2):187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
15. Andersen PK, Ravn H. Models for Multi-State Survival Data. Boca Raton: Chapman and Hall/CRC; 2023. https://doi.org/10.1201/9780429029684.
16. Lin DY. On the Breslow estimator. Lifetime Data Anal. 2007;13(4):471–80. https://doi.org/10.1007/s10985-007-9048-y.
17. Ramlau-Hansen H. Smoothing Counting Process Intensities by Means of Kernel Functions. Ann Stat. 1983;11(2):453–66. https://doi.org/10.1214/aos/1176346152.
18. Basha L, Hoxha F. Kernel Estimation of the Baseline Function in the Cox Model. Eur Sci J ESJ. 2019;15(6). https://doi.org/10.19044/esj.2019.v15n6p105.
19. Kleinbaum DG, Klein M. The Stratified Cox Procedure. 2012. pp. 201–40. https://doi.org/10.1007/978-1-4419-6646-9_5.
20. McMahan HB, Moore E, Ramage D, Hampson S, Arcas BAy. Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017. 2016. http://arxiv.org/abs/1602.05629. Accessed 20 May 2025.
21. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. J Am Stat Assoc. 1979;74(368):829–36. https://doi.org/10.1080/01621459.1979.10481038.
22. Chambers JM, Hastie TJ. Statistical Models in S. Chambers JM, Hastie TJ, editors. Heidelberg: Routledge; 2017. https://doi.org/10.1201/9780203738535.
23. Rao MB, Klein JP, Moeschberger ML. Survival Analysis Techniques for Censored and Truncated Data. Technometrics. 1998;40(2):159. https://doi.org/10.2307/1270658.
24. Steyerberg EW. Clinical Prediction Models. vol. 77 of Statistics for Biology and Health. Cham: Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-16399-0.
25. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Stat Med. 2005;24(11):1713–23. https://doi.org/10.1002/sim.2059.
26. Harrell FE. Regression Modeling Strategies. Springer Series in Statistics. Cham: Springer International Publishing; 2015. https://doi.org/10.1007/978-3-319-19425-7.
27. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med. 1999;18(17–18):2529–45. https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5.
28. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011;30(10):1105–17. https://doi.org/10.1002/sim.4154.

29. Arboe B, Josefsson P, Jørgensen J, Haaber J, Jensen P, Poulsen CB, et al. Danish National Lymphoma Registry. Clin Epidemiol. 2016;8:577–81. https://doi.org/10.2147/CLEP.S99470.
30. Rask Kragh Jørgensen R, Bergström F, Eloranta S, Tang Severinsen M, Bjøro Smeland K, Fosså A, et al. Machine Learning-Based Survival Prediction Models for Progression-Free and Overall Survival in Advanced-Stage Hodgkin Lymphoma. JCO Clinical Cancer Informatics. 2024;(8):e2300255. https://doi.org/10.1200/CCI.23.00255.
31. Hasenclever D, Diehl V, Armitage JO, Assouline D, Björkholm M, Brusamolino E, et al. A Prognostic Score for Advanced Hodgkin's Disease. N Engl J Med. 1998;339(21):1506–14. https://doi.org/10.1056/nejm199811193392104.
32. Coppel KC P. General Data Protection Regulation. In: Information Rights. Bloomsbury Publishing Plc; 2023. pp. 557–645. https://doi.org/10.5040/9781509967339.V2_0031.
33. Miao G, Yu L, Yang J, Bennett DA, Zhao J, Wu SS. Learning from vertically distributed data across multiple sites: An efficient privacy-preserving algorithm for Cox proportional hazards model with variable selection. J Biomed Inform. 2024;149:104581. https://doi.org/10.1016/j.jbi.2023.104581.
34. Saha-Chaudhuri P, Juwara L. Survival analysis under the Cox proportional hazards model with pooled covariates. Stat Med. 2021;40(4):998–1020. https://doi.org/10.1002/sim.8816.
35. Imakura A, Tsunoda R, Kagawa R, Yamagata K, Sakurai T. DC-COX: Data collaboration Cox proportional hazards model for privacy-preserving survival analysis on multiple parties. J Biomed Informa. 2023;137:104264. https://doi.org/10.1016/j.jbi.2022.104264.
36. Ebadi H, Sands D, Schneider G. Differential Privacy. In: Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. New York: ACM; 2015. pp. 69–81. https://doi.org/10.1145/2676726.2677005.
37. Pazira H, Massa E, Weijers JAM, Coolen ACC, Jonker MA. Bayesian Federated Inference for Survival Models. 2024. https://api.semanticscholar.org/CorpusID:269430589. Accessed 20 May 2025.
38. Jordan MI, Lee JD, Yang Y. Communication-Efficient Distributed Statistical Inference. J Am Stat Assoc. 2019;114(526):668–81. https://doi.org/10.1080/01621459.2018.1429274.
39. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. vol. 103 of Springer Texts in Statistics. 1st ed. New York: Springer New York; 2013. https://doi.org/10.1007/978-1-4614-7138-7.
40. Lawless JF, Nadeau C, Cook RJ. Analysis of Mean and Rate Functions for Recurrent Events. New York: Springer US; 1997. pp. 37–49. https://doi.org/10.1007/978-1-4684-6316-3_4.

Publisher's Note