

An architectural rendering of a modern building with a brick and dark panel facade, featuring large windows. In the foreground, a wide, paved plaza is populated with people walking, sitting, and pushing a stroller. The scene is set against a clear blue sky with lush green trees lining the plaza.

Time to Federation

DSTS Two-day meeting

Rasmus Rask Kragh Jørgensen, 4/5/2026

RESEARCH

Open Access

Development of time to event prediction models using federated learning

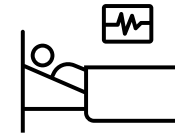
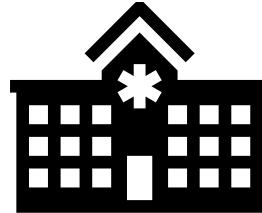


Rasmus Rask Kragh Jørgensen^{1,2,3*} , Jonas Faartoft Jensen¹, Tarec El-Galaly^{1,3,4,5}, Martin Bøgsted^{2,6}, Rasmus Froberg Brøndum^{2,6}, Mikkel Runason Simonsen^{1,7} and Lasse Hjort Jakobsen¹

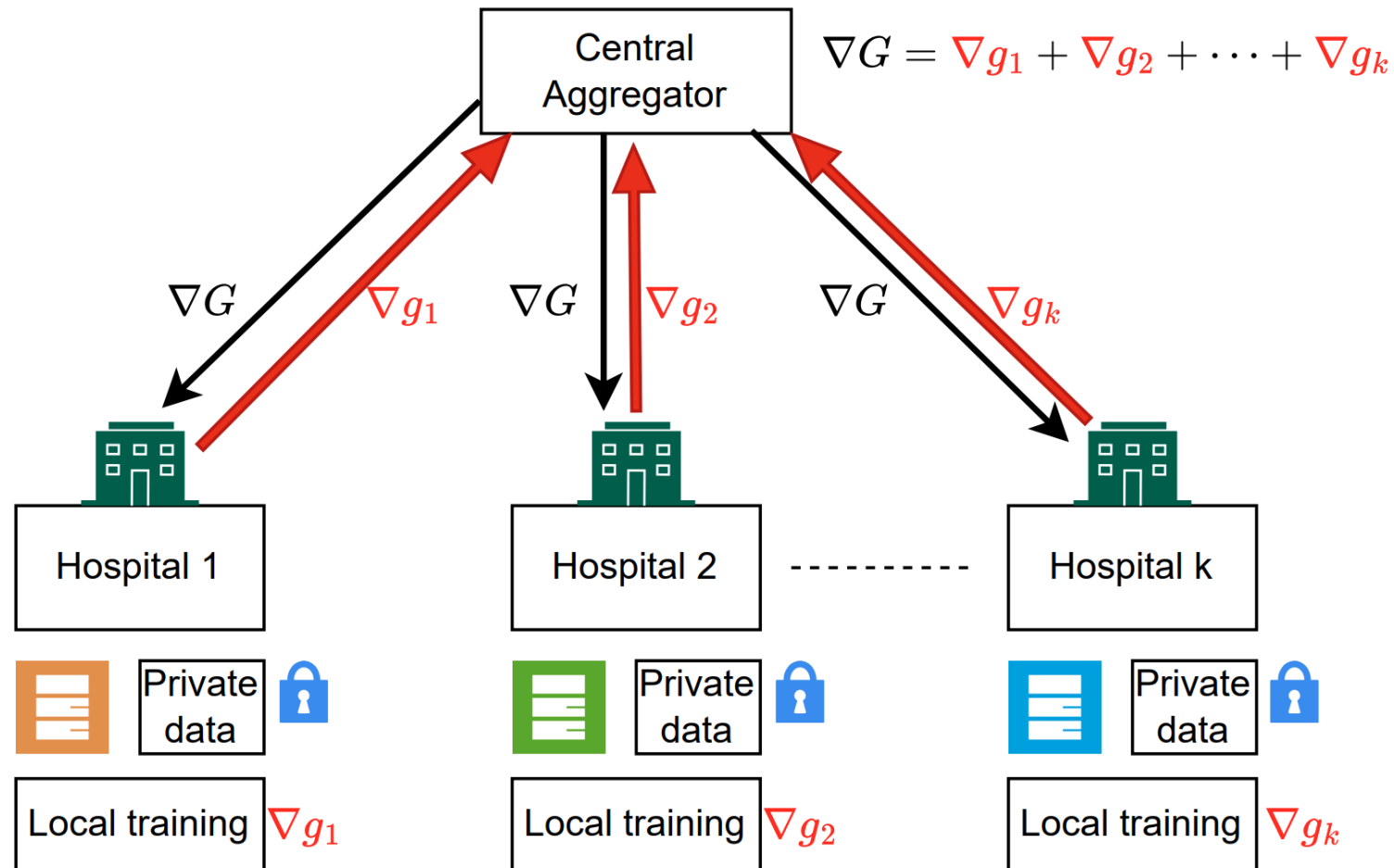


Introduction – Survival prediction with Cox

$$\hat{S}(t|x_i, \hat{\beta}) = \exp\left(-\int_0^t \hat{h}_0(u) \exp(x_i^T \hat{\beta}) du\right) = \hat{S}_0(t)^{\exp(x_i^T \hat{\beta})},$$



Introduction – Federated learning



Introduction – Previous studies

i	...	T_d	...
...	...	...	...
i	...	$\sum_{j \in R(T_d)} \exp\{(\hat{\beta}_{WICOX}^{(i)})^T x_j\}$	...
...	...	$= \sum_{k=1}^3 \sum_{j \in R_k(T_d)} \exp\{(\hat{\beta}_{WICOX}^{(i)})^T x_j\}$	...
...	...	...	...

At each
At master part

- Output
- $\hat{S}(t)^{(i)}$
- $\hat{\Lambda}_o(t)^{(i)}$

or distributed

inter-site update in the distributed Cox model

RECEIVED 24 November 2014

ODAC algorithm

(1) Initialization

In site $k = 1$ to K ,

do

Fit a Cox model and obtain the local estimate $\hat{\beta}_k$ and the variance estimate $\hat{V}_k$; **broadcast** $\hat{\beta}_k$, $\hat{V}_k$, and the **set of unique event time points in site k**.

end

(2) Local surrogate estimator

In Site $k = 1$ to K ,

do

obtain $\tilde{\beta}$ using (4), and **all the unique event time points across all sites** $t_1 \dots t_d$; calculate and broadcast the intermediate summary-level statistics $U_j(T)$, $W_j(T)$ and $Z_j(T)$; construct the surrogate likelihood $\tilde{L}_k(\beta)$ by (3) treating the k-th site as the local site; obtain and broadcast $\tilde{\beta}_k$ and the variance $\tilde{V}_k$;

end

(3) Evidence synthesis

Obtain $\tilde{\beta}$ using (5) by plugging in $\tilde{\beta}_k$ and $\tilde{V}_k$.

Return $\tilde{\beta}$.

Step 1 and Step 2 for prediction and log (HR) estimation

Fig. 1. Illustration of overall framework of the WICOX algorithm under the

denote the i th datasets in k party. The m models are constructed at each party through $\{Z_{k1}^{(i)}\}_{i=1,2,\dots,m}$, and the estimated parameters, $\{(\hat{\beta}_k^{(i)})'\}_{i=1,2,\dots,m}$ and $\{(\hat{V}(\hat{\beta}_k^{(i)}))'\}_{i=1,2,\dots,m}$, and total set of event times, $t_{k1}, t_{k2}, \dots, t_{kd_{k1}}$, are exchanged between parties.

Here, $(\hat{\beta}_k^{(i)})' = (\hat{\beta}_{k1}^{(i)}, \hat{\beta}_{k2}^{(i)}, \dots, \hat{\beta}_{kP}^{(i)})$ is a vector of parameters estimated from $Z_{k1}^{(i)}$, and $(\hat{V}(\hat{\beta}_k^{(i)}))' = (\hat{V}(\hat{\beta}_{k1}^{(i)}), \hat{V}(\hat{\beta}_{k2}^{(i)}), \dots, \hat{V}(\hat{\beta}_{kP}^{(i)}))$ is a vector of variance of $\hat{\beta}_k^{(i)}$.

Park et al. IEEE JBHI 2023

4: Repeat

Lu et al. JAMIA 2015



CENTER FOR CLINICAL
DATA SCIENCE

AALBORG
UNIVERSITET



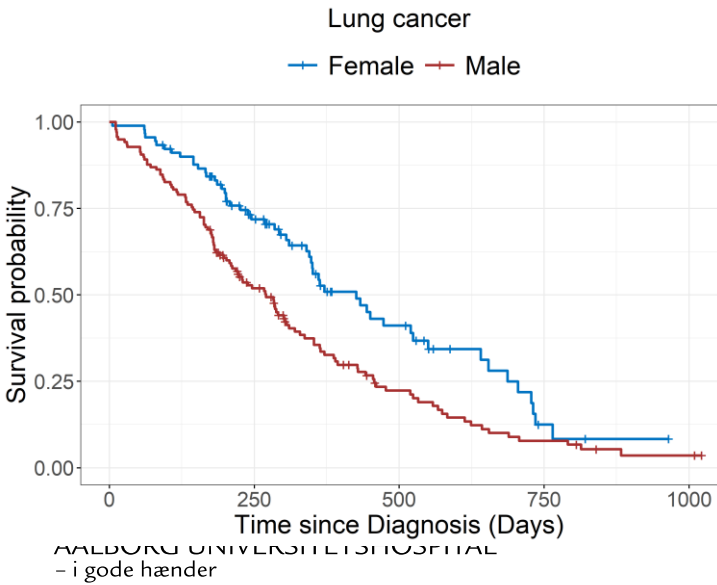
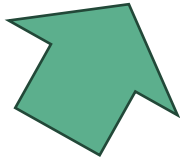
AALBORG UNIVERSITETSHOSPITAL
- i gode hænder

Duan et al. JAMIA 2020

Sensitive dataset

Time	Status	Sex	Age	ECOG	Disease
80	1	Male	27	1	Lung cancer
809	0	Female	33	3	Lung cancer
332	1	Male	66	2	Lung cancer
467	0	Female	56	3	Lung cancer
123	1	Female	87	1	Lung cancer
4	1	NA	18	3	Lung cancer

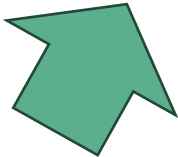
Time	Status	Sex	Age	ECOG	Disease
80	1	Male			Lung cancer
700-900	0	Female			Lung cancer
332	1	Male			Lung cancer
427-511	0	Female			Lung cancer
123	1	Female			Lung cancer



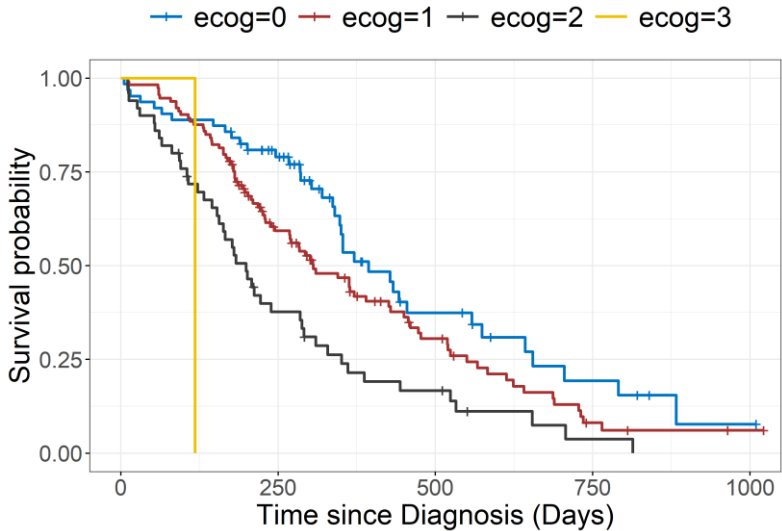
Sensitive dataset

Time	Status	Sex	Age	ECOG	Disease
80	1	Male	27	1	Lung cancer
809	0	Female	33	3	Lung cancer
332	1	Male	66	2	Lung cancer
467	0	Female	56	3	Lung cancer
123	1	Female	87	1	Lung cancer
4	1	NA	18	3	Lung cancer

Time	Status	Sex	Age	ECOG	Disease
80	1			1	Lung cancer
700-900	0			3	Lung cancer
332	1			2	Lung cancer
427-511	0			3	Lung cancer
123	1			1	Lung cancer
4	1			3	Lung cancer



Lung cancer



Time	Status	Sex	Age	ECOG	Disease
80	1			1	Lung cancer
750-870	0			3	Lung cancer
332	1			2	Lung cancer
427-511	0			3	Lung cancer
123	1			1	Lung cancer
4	1			3	Lung cancer

Time	Status	Sex	Age	ECOG	Disease
80	1	Male			Lung cancer
700-900	0	Female			Lung cancer
332	1	Male			Lung cancer
427-511	0	Female			Lung cancer
123	1	Female			Lung cancer

Time	Status	Sex	Age	ECOG	Disease
80	1	Male		1	Lung cancer
750-870	0			3	Lung cancer
700-900	0	Female			Lung cancer
332	1	Male		2	Lung cancer
467-580	0			3	Lung cancer
427-511	0	Female			Lung cancer
123	1	Female		1	Lung cancer
4	1			3	Lung cancer

Time	Status	Sex	Age	ECOG	Disease
80	1	Male	27	1	Lung cancer
809	0	Female	33	3	Lung cancer
332	1	Male	66	2	Lung cancer
467	0	Female	56	3	Lung cancer
123	1	Female	87	1	Lung cancer
4	1	NA	18	3	Lung cancer

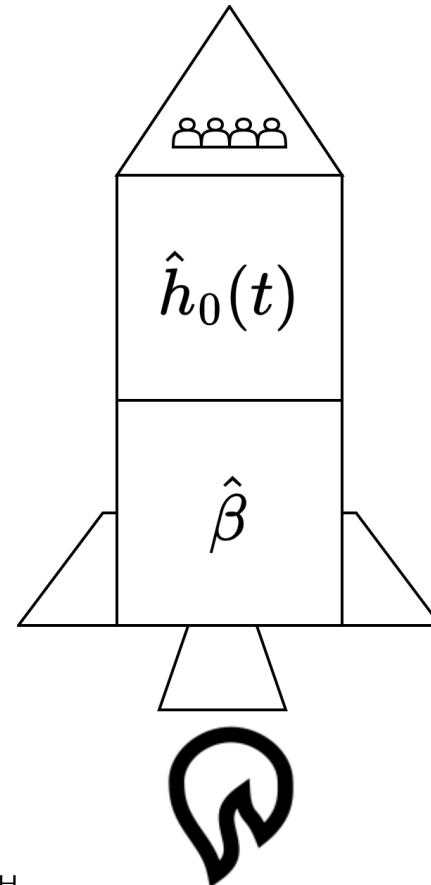
Aim

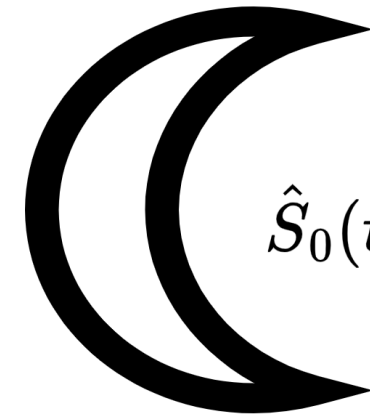
- ▶ To introduce new **federated learning** methodologies for survival prediction without sharing **sensitive information**.
- ▶ To enable **anonymous** survival predictions without exchanging **event-times**



Methods – Federated Cox

- Cox model
- Enable survival predictions
 - Two-step process
 1. Coefficient estimates
 2. Estimation of the baseline hazard




$$\hat{S}_0(t)^{\exp(X\hat{\beta})}$$

Methods – Coefficient estimation

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

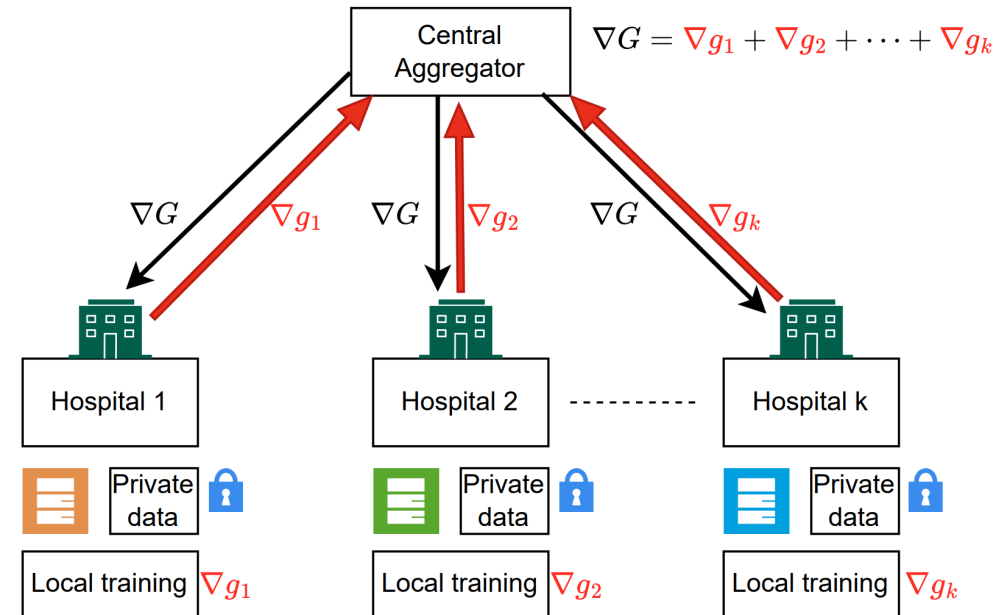
► FL
$$L(\beta) = \prod_{s=1}^S \prod_{i=1}^{n_s} L_{s,i}(\beta|(t, \delta, x)) = \prod_{s=1}^S \prod_{i=1}^{n_s} \left(\frac{\exp(x_{s,i}^\top \beta)}{\sum_{j \in \mathcal{R}_i} \exp(x_{s,j}^\top \beta)} \right)^{\delta_i}$$

$$L(\beta) = \prod_{s=1}^S \prod_{i=1}^{n_s} L_{s,i}(\beta|(t, \delta, x)) = \prod_{s=1}^S \prod_{i=1}^{n_s} \left(\frac{\exp(x_{s,i}^\top \beta)}{\sum_{j \in \mathcal{R}_{s,i}} \exp(x_{s,j}^\top \beta)} \right)^{\delta_i}$$

Methods – Coefficient estimation

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

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



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

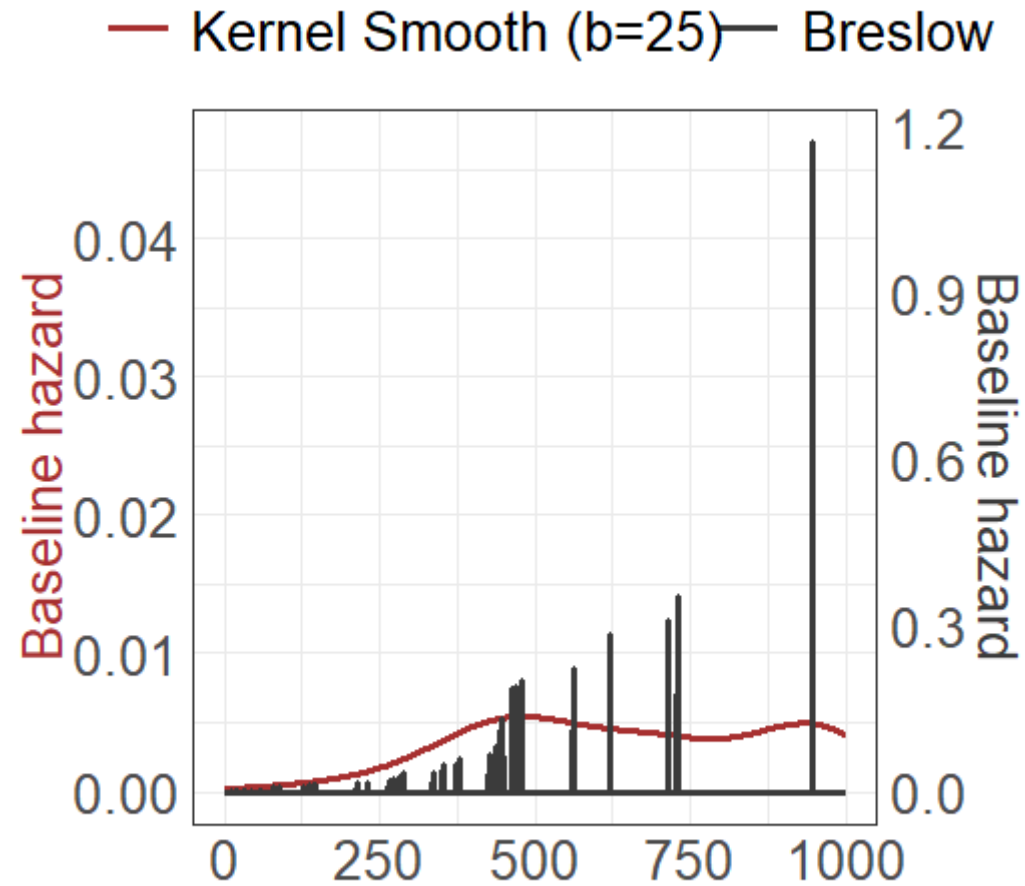
Methods – Survival predictions

$$\hat{S}(t|x_i, \hat{\beta}) = \exp \left(- \int_0^t \hat{h}_0(u) \exp(x_i^T \hat{\beta}) du \right) = \hat{S}_0(t)^{\exp(x_i^T \hat{\beta})},$$

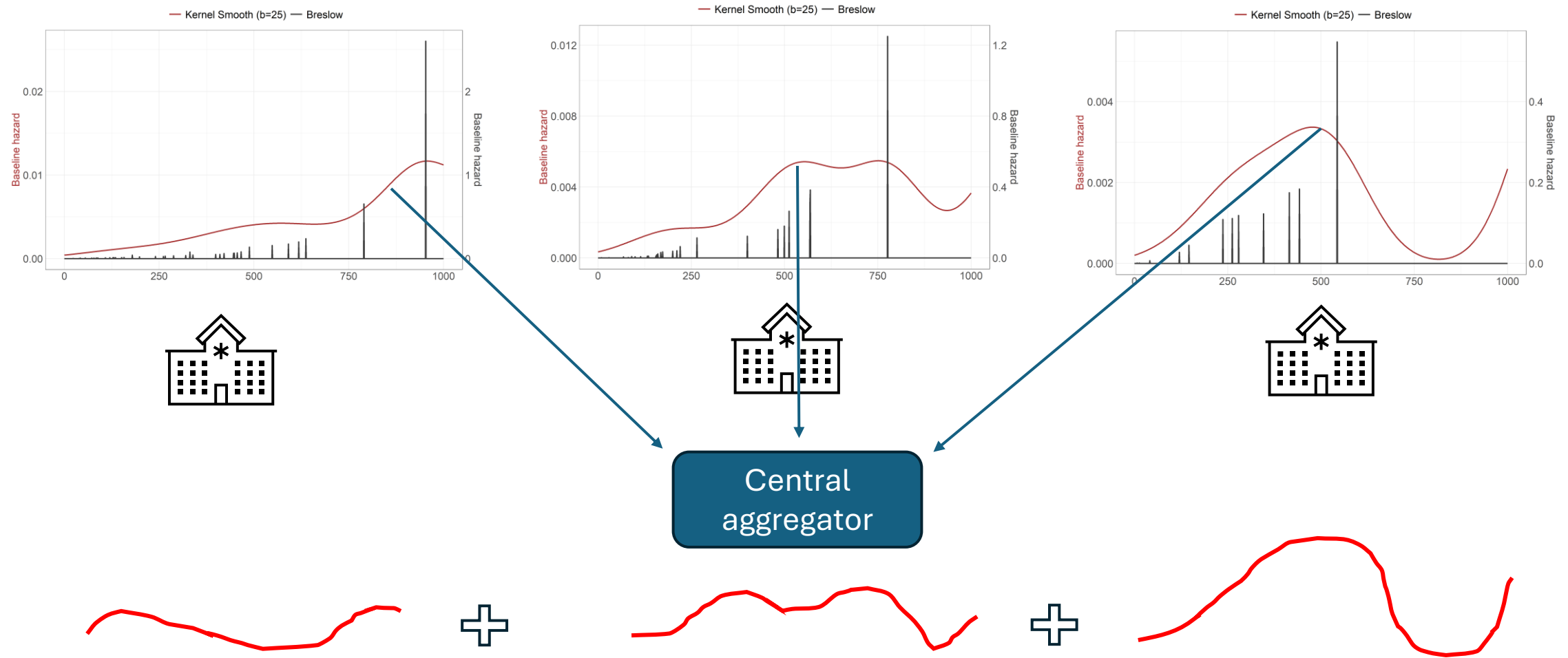
where $\hat{S}_0(t) = \exp(-\hat{H}_0(t))$

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

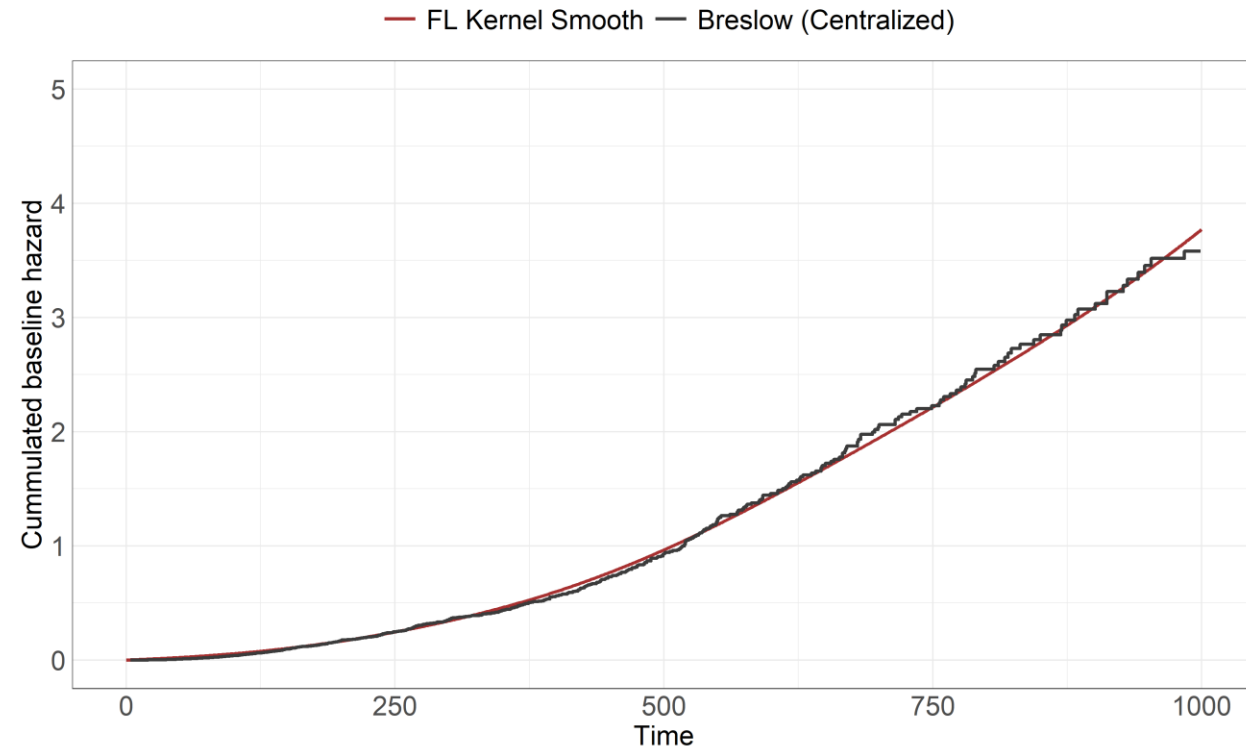
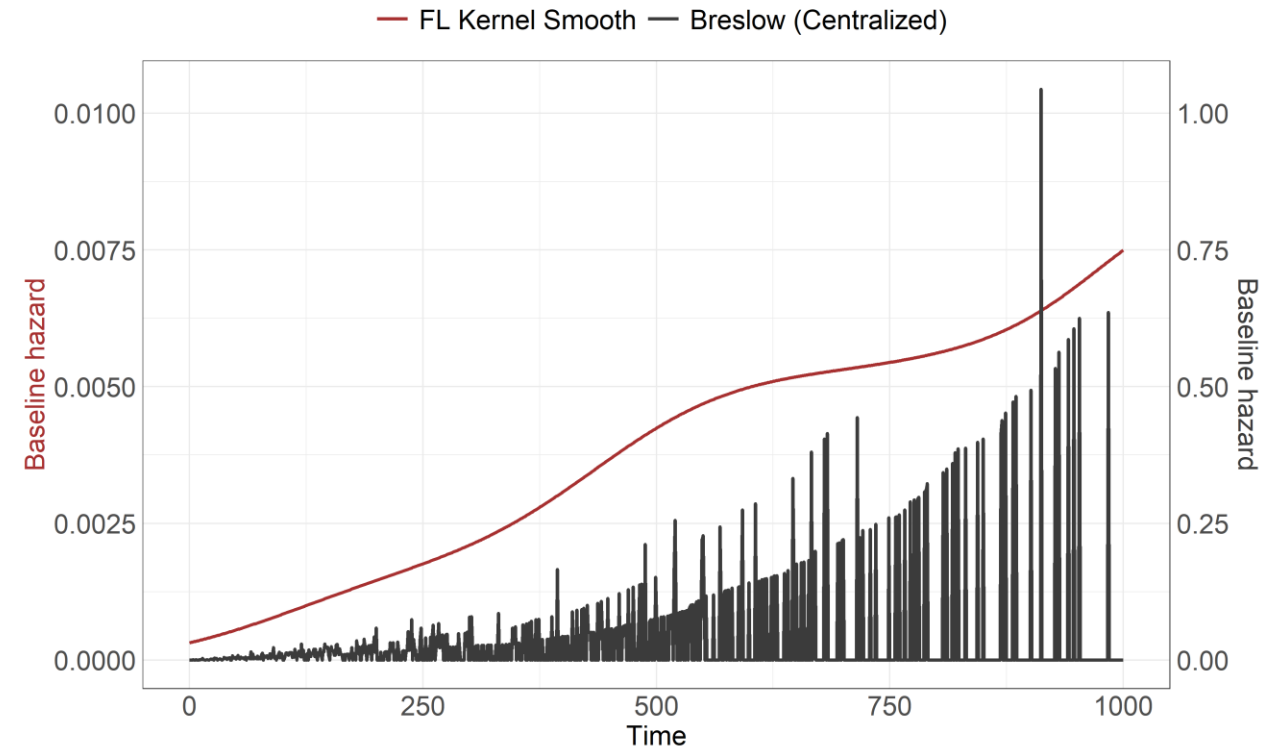
Methods – Baseline hazard



Methods – Federated baseline hazard



Methods – Federated baseline hazard



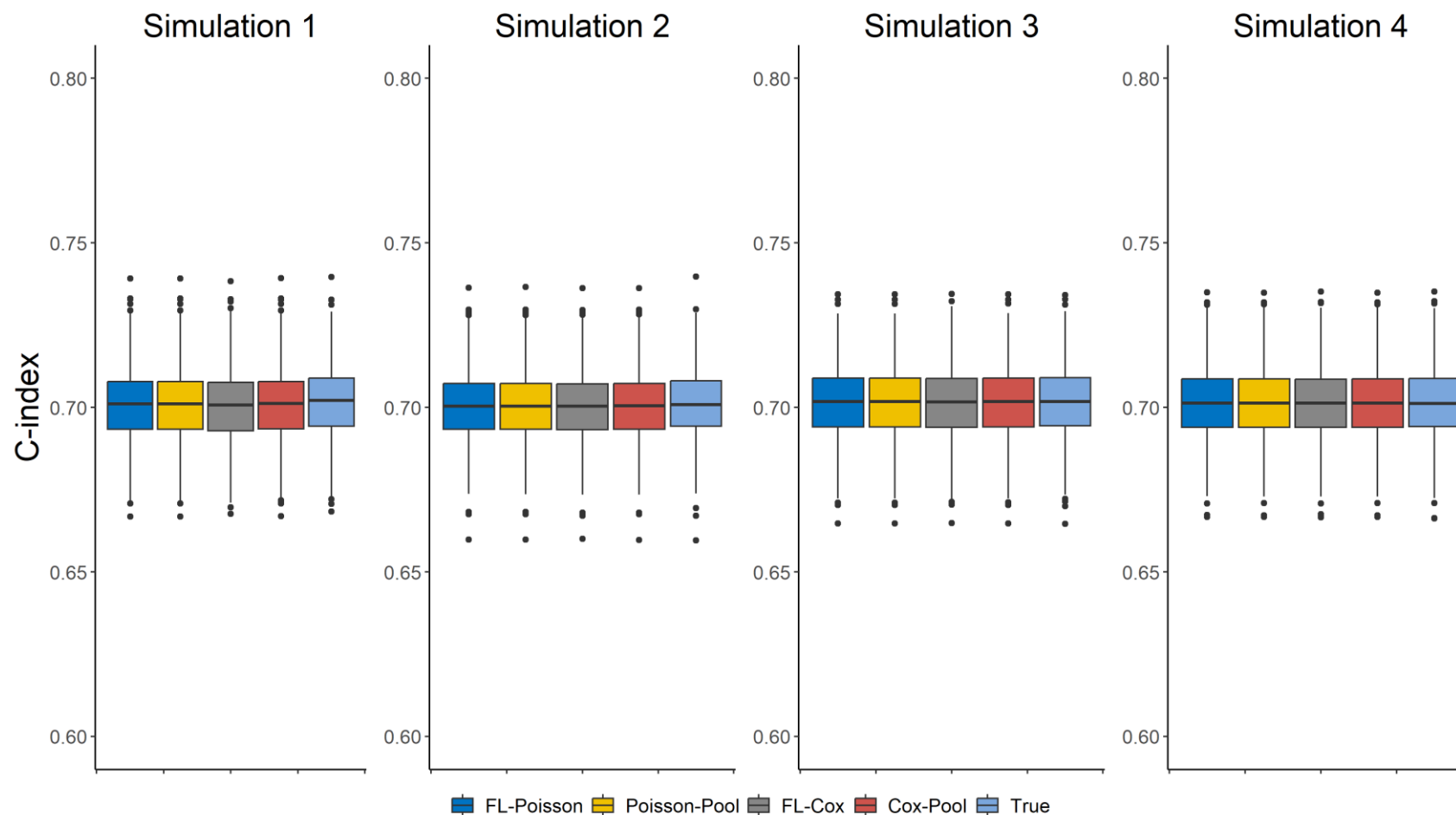
Results - Simulations

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	$n_s \in \{20, 30, \dots, 210\}$	$n_s \in \{20, 30, \dots, 210\}$
$T_{C_{max}}$ (Years)	5	$T_{C_{max}} \in \{1, 2, \dots, 10\}$	5	$T_{C_{max}} \in \{1, 2, \dots, 20\}$
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



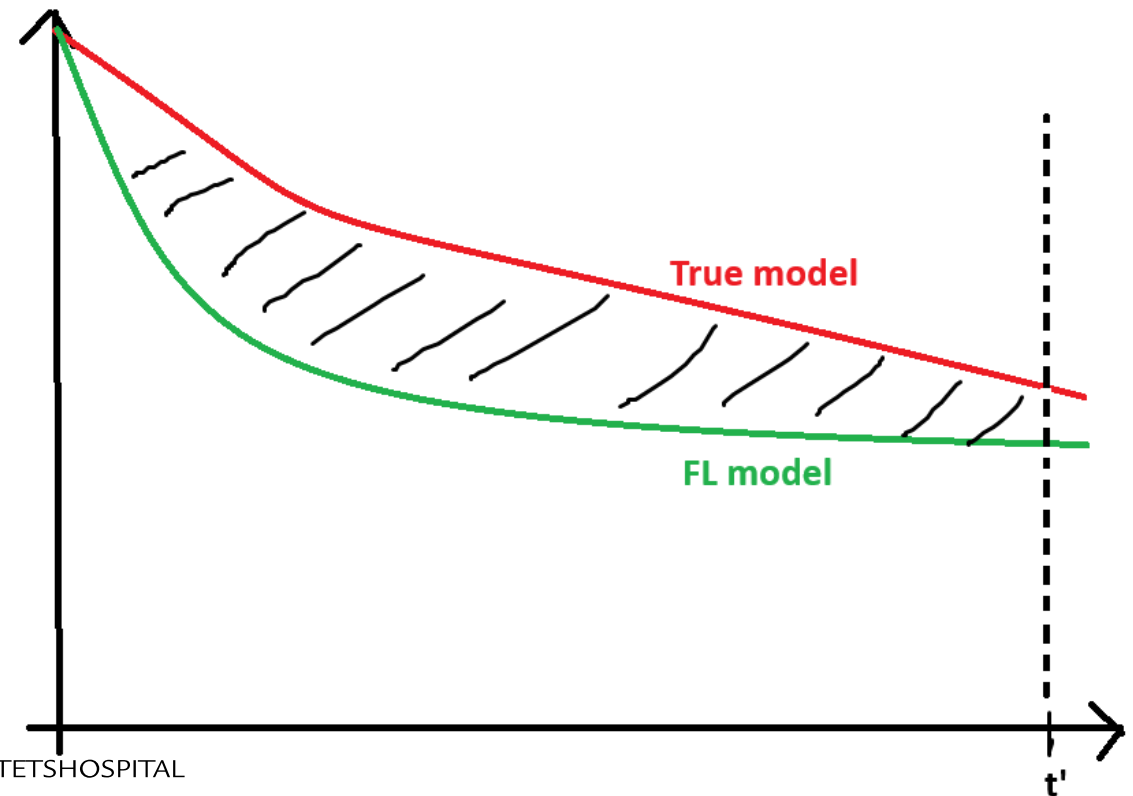
Results – C index



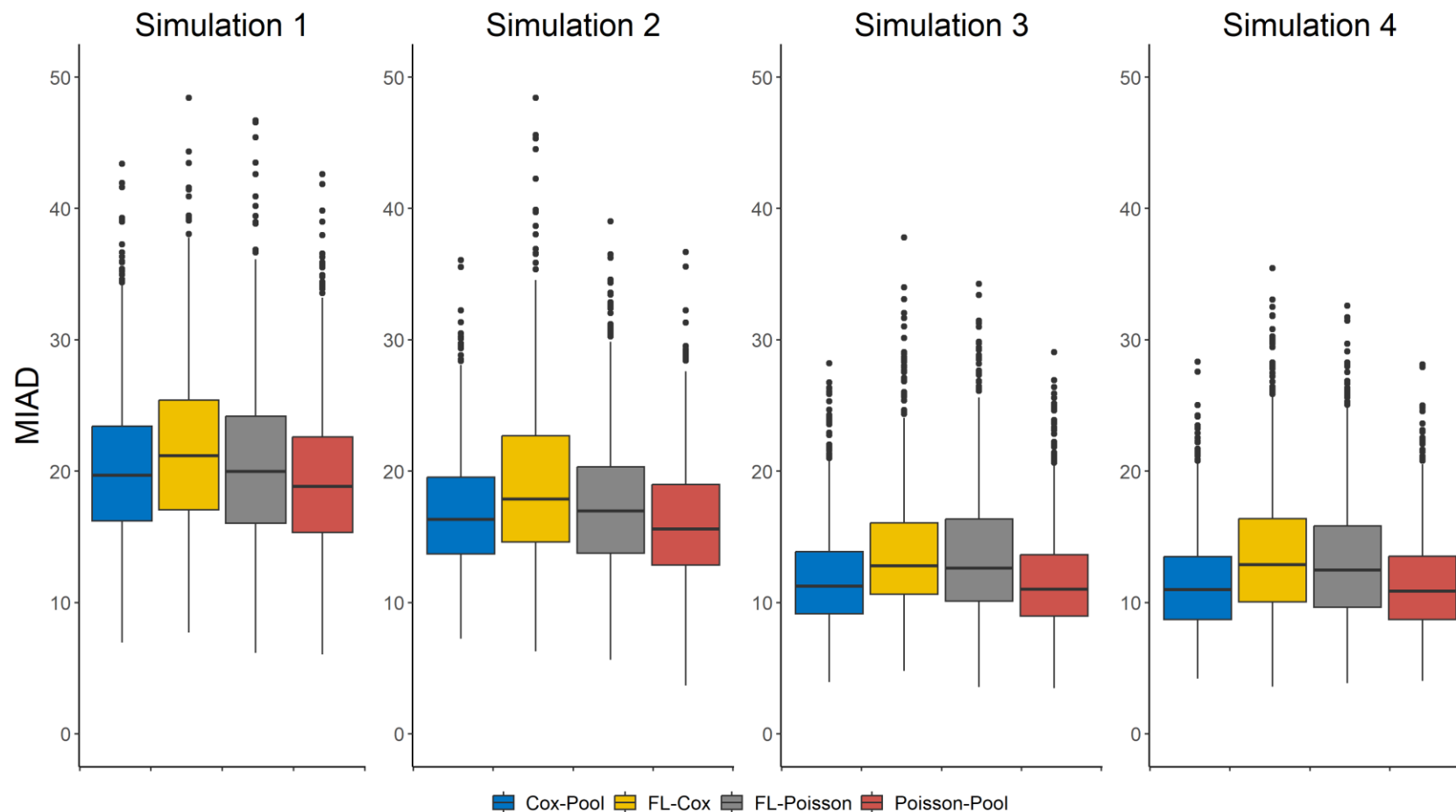
Results - Performance

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

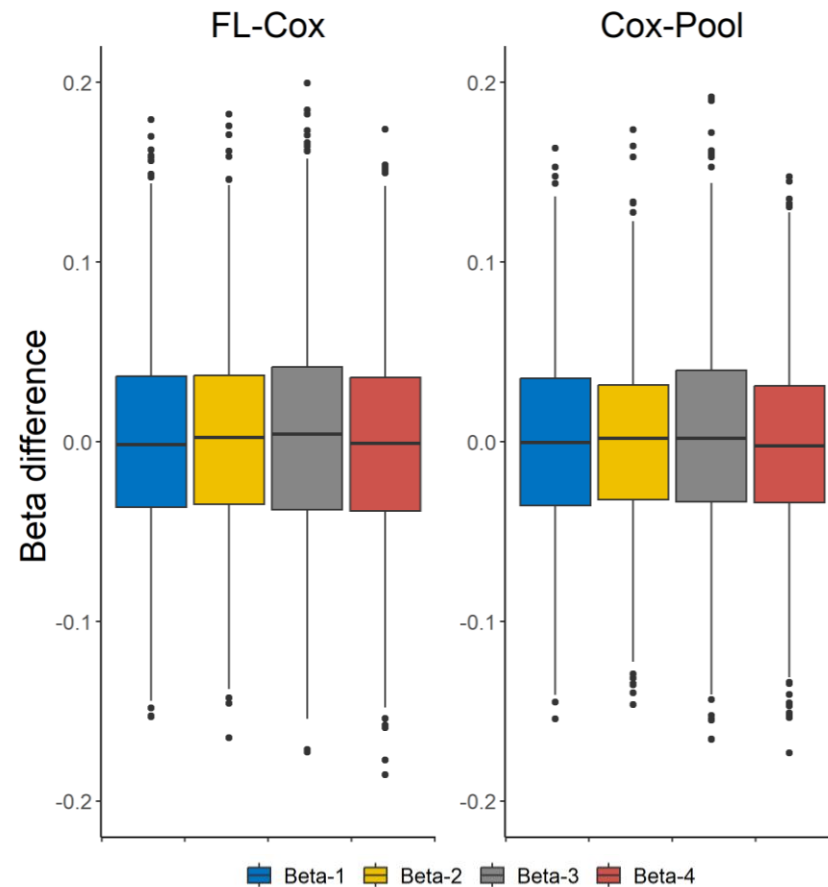
$$MIAD = \frac{1}{n} \sum_{i=1}^n ISD(t'|x_i)$$



Results – MIAD



Results – Coefficient estimation



Results – Real-world experiment

- Real-World advanced-stage Hodgkin Lymphoma
- Total 707 patients
- 70/30 train/test split
- Train dataset split into 8 separate sites
- Predicted 5-year survival after diagnosis

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

Conclusion

- It is **possible** to train survival prediction models without sharing **event-times** with **minimal loss** in predictive accuracy for the Cox model
- The method allows for the possibility of **training** predictions models in **small data subsets** (e.g., rare diseases)
- But what kind of **infrastructure** is needed to make this possible in the **real-world**?

Thank you for your time - Questions

Contact information

Rasmus Rask Kragh Jørgensen, Postdoc
Center for Clinical Data Science, AAU

E-mail: Rasmus.rask@rn.dk/Rasmusrkj@dcm.aau.dk

LinkedIn: [in/rasmusraskkj](https://www.linkedin.com/in/rasmusraskkj)

