Biostatistics (2021), **0**, 0, *pp.* S1–S13
doi:10.1093/biostatistics/output

Supplementary material to *Distributed Cox Proportional Hazards Regression Using Summary-level Information*

DONGDONG LI[1,*], WENBIN LU[2], DI SHU[3,4,5], SENGWEE TOH[1], RUI WANG[1,6]

[1] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA USA*
[2] *Department of Statistics, North Carolina State University, Raleigh, NC, USA*
[3] *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
[4] *Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA*
[5] *Center for Pediatric Clinical Effectiveness, Children's Hospital of Philadelphia, Philadelphia, PA, USA*
[6] *Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA*

S1. Additional Details for Section 2 of the paper

S1.1 *Proof*

Proof of Theorem 2.1. Denote the true parameter value as $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\beta}'_{10}, \ldots, \boldsymbol{\beta}'_{K0}, \boldsymbol{\beta}'_0)'$, where $\boldsymbol{\beta}_{k0}$, $k = 1, \ldots, K$ is the true parameter value within each site. It is assumed that $\boldsymbol{\beta}_{10} = \ldots = \boldsymbol{\beta}_{K0} = \boldsymbol{\beta}_0$. For ease of derivation we will keep the notation of $\boldsymbol{\beta}_{k0}$ for now, $k = 1, \ldots, K$. The set of parameters $\boldsymbol{\theta} \triangleq (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_K, \boldsymbol{\beta}')'$ is estimated by $\hat{\boldsymbol{\theta}}^{(1)} \triangleq (\hat{\boldsymbol{\beta}}'_1, \ldots, \hat{\boldsymbol{\beta}}'_K, \hat{\boldsymbol{\beta}}^{(1)\prime})'$, the solution to the $K+1$ estimating equations $\boldsymbol{U}^*_{\boldsymbol{\beta}_k} = \mathbf{0}$ for $k = 1, \ldots, K$ and $\boldsymbol{U}_{\boldsymbol{\theta}} = \mathbf{0}$. Under regularity conditions (Cox and Hinkley, 1974, p.281), $\sqrt{n}(\hat{\boldsymbol{\beta}}'_1 - \boldsymbol{\beta}'_{10}, \ldots, \hat{\boldsymbol{\beta}}'_K - \boldsymbol{\beta}'_{K0}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{I}^{*-1})$ and $\boldsymbol{I}^*$ is partitioned

*To whom correspondence should be addressed. E-mail: dongdong_li@hphci.harvard.edu

into

$$\boldsymbol{I}^* = \begin{bmatrix} \boldsymbol{I}^*_{11} & \dots & \boldsymbol{I}^*_{1K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{I}^*_{K1} & \dots & \boldsymbol{I}^*_{KK} \end{bmatrix}$$

where $n\boldsymbol{I}^*_{ij}$ is the variance-covariance matrix between $\boldsymbol{U}^*_{\boldsymbol{\beta}_i}$ and $\boldsymbol{U}^*_{\boldsymbol{\beta}_j}$ for $i, j = 1, \dots, K$, since the K sites are independent, $\boldsymbol{I}^*_{ij} = \boldsymbol{0}$ for $i \neq j$.

The information matrix, denoted by $\boldsymbol{I}$, for the $K+1$ score functions ($\boldsymbol{U}^*_{\boldsymbol{\beta}_k}$ for $k = 1, \dots, K$, and $\boldsymbol{U}_{\boldsymbol{\theta}}$) can be partitioned into $(K+1) \times (K+1)$ blocks

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{I}_{11} & \dots & \boldsymbol{I}_{1K} & \boldsymbol{I}_{1(K+1)} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{I}_{K1} & \dots & \boldsymbol{I}_{KK} & \boldsymbol{I}_{K(K+1)} \\ \boldsymbol{I}_{(K+1)1} & \dots & \boldsymbol{I}_{(K+1)K} & \boldsymbol{I}_{(K+1)(K+1)} \end{bmatrix}$$

Taylor expanding the score function $\boldsymbol{U}^*_{\boldsymbol{\beta}_k}$ around $\boldsymbol{\beta}_{k0}$ and evaluating it at $\hat{\boldsymbol{\beta}}_k$:

$$\boldsymbol{U}^*_{\boldsymbol{\beta}_k}(\hat{\boldsymbol{\beta}}_k) = \boldsymbol{0} = \boldsymbol{U}^*_{\boldsymbol{\beta}_k}(\boldsymbol{\beta}_{k0}) - \boldsymbol{A}^*_{\boldsymbol{\beta}_k}(\boldsymbol{\beta}_{k0})(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k0}) + \boldsymbol{\mathcal{O}}(\sqrt{n_k})$$

where

$$\boldsymbol{A}^*_{\boldsymbol{\beta}_k}(\boldsymbol{\beta}_{k0}) = -\frac{\partial \boldsymbol{U}^*_{\boldsymbol{\beta}_k}}{\partial \boldsymbol{\beta}_k}\bigg|_{\boldsymbol{\beta}_k = \boldsymbol{\beta}_{k0}}.$$

Similarly,

$$\begin{aligned} \boldsymbol{U}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\beta}}^{(1)}) = \boldsymbol{0} = \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_{10}, \dots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0) - \\ \boldsymbol{B}_1(\boldsymbol{\beta}_{10}, \dots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) - \\ \dots - \\ \boldsymbol{B}_K(\boldsymbol{\beta}_{10}, \dots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_{K0}) - \\ \boldsymbol{B}(\boldsymbol{\beta}_{10}, \dots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0) + \\ \boldsymbol{\mathcal{O}}(\sqrt{n}) \end{aligned}$$

where

$$\boldsymbol{B}_k(\boldsymbol{\beta}_{10}, \dots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0) = -\frac{\partial \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}$$

for $k = 1, \ldots, K$, and

$$\boldsymbol{B}(\boldsymbol{\beta}_{10}, \ldots, \boldsymbol{\beta}_{K0}, \boldsymbol{\beta}_0) = -\frac{\partial \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

By the law of large numbers, as $n_k \to \infty$, $\boldsymbol{A}^*_{\boldsymbol{\beta}_k}/n_k$, $\boldsymbol{B}_k/n_k$, and $\boldsymbol{B}/n$ converge to $\boldsymbol{I}^*_{kk}$, $\boldsymbol{I}_{(K+1)k}$, and $\boldsymbol{I}_{(K+1)(K+1)}$, respectively, for $k = 1, \ldots, K$. Therefore

$$n^{-\frac{1}{2}} \begin{bmatrix} \boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_{10}) \\ \vdots \\ \boldsymbol{U}^*_{\boldsymbol{\beta}_K}(\boldsymbol{\beta}_{K0}) \\ \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) \end{bmatrix} \approx \sqrt{n}\boldsymbol{P} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} \\ \vdots \\ \hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_{K0} \\ \hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0 \end{bmatrix}$$

where

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I}^*_{11} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & & \ldots & \boldsymbol{I}^*_{KK} & \boldsymbol{0} \\ \boldsymbol{I}_{(K+1)1} & & \ldots & \boldsymbol{I}_{(K+1)K} & \boldsymbol{I}_{(K+1)(K+1)} \end{bmatrix}$$

Next we show that $\mathrm{cov}(\boldsymbol{U}^*_{\boldsymbol{\beta}_k}, \boldsymbol{U}_{\boldsymbol{\theta}}) \approx \boldsymbol{0}$. It suffices to show the case of $K = 2$. Let Data_k to denote data from the $k-$th site. The covariance

$$\mathrm{cov}(\boldsymbol{U}^*_{\boldsymbol{\beta}_1}, \boldsymbol{U}_{\boldsymbol{\theta}}) = \mathbb{E}\big[\boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_1)\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\big] = \mathbb{E}\big[\mathbb{E}\big(\boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_1)\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\big|\mathrm{Data}_2\big)\big]$$
$$= \mathbb{E}\big[\mathbb{E}\boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_1)\mathbb{E}\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathrm{Data}_2)\big] = \boldsymbol{0}$$

since $\mathbb{E}\boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_1) = \boldsymbol{0}$. Similarly the covariance between $\boldsymbol{U}^*_{\boldsymbol{\beta}_2}$ and $\boldsymbol{U}_{\boldsymbol{\theta}}$ is $\boldsymbol{0}$.

By the Central Limit Theorem, $n^{-1/2}(\boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta}_{10}), \ldots, \boldsymbol{U}^*_{\boldsymbol{\beta}_K}(\boldsymbol{\beta}_{K0}), \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0)) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{H})$ where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{I}^*_{11} & \ldots & \boldsymbol{I}^*_{1K} & \boldsymbol{0} \\ \vdots & & \vdots * \vdots & \\ \boldsymbol{I}^*_{K1} & \ldots & \boldsymbol{I}^*_{KK} & \boldsymbol{0} \\ \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{I}_{(K+1)(K+1)} \end{bmatrix}$$

Therefore $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P}'^{-1})$, and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}^{(1)}$ is the bottom right corner block, which is

$$\Sigma^{(1)} = \boldsymbol{I}^{-1}_{(K+1)(K+1)} + \boldsymbol{I}^{-1}_{(K+1)(K+1)}\Big[\sum_{k=1}^{K} \boldsymbol{I}_{(K+1)k}\boldsymbol{I}^{*-1}_{kk}\boldsymbol{I}'_{(K+1)k}\Big]\boldsymbol{I}^{-1}_{(K+1)(K+1)}$$

One can use the observed information to estimate $\boldsymbol{I}_{(K+1)(K+1)}$, $\boldsymbol{I}_{(K+1)k}$ and $\boldsymbol{I}^*_{kk}$. The corresponding observed information are $\boldsymbol{I}(\hat{\boldsymbol{\theta}}^{(1)}) \triangleq -\partial \boldsymbol{U}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{(1)})/\partial \boldsymbol{\theta}$, $\boldsymbol{I}_k(\hat{\boldsymbol{\theta}}^{(1)}) \triangleq -\partial \boldsymbol{U}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{(1)})/\partial \boldsymbol{\beta}_k$, and

$\boldsymbol{I}^*_{kk}(\hat{\boldsymbol{\beta}}_k) \triangleq -\partial \boldsymbol{U}^*_{\boldsymbol{\beta}_k}(\hat{\boldsymbol{\beta}}_k)/\partial\boldsymbol{\beta}_k$, respectively. Note that the $\boldsymbol{I}$ may not be a symmetric matrix, and one may use the negation of approximated Hessian matrix $-\boldsymbol{H}^*(\hat{\boldsymbol{\theta}}^{(1)})$ as the estimator for $\boldsymbol{I}_{(K+1)(K+1)}$. Thus the variance estimator in (2.6) follows. Note that with the assumption $\boldsymbol{\beta}_{10} = \ldots = \boldsymbol{\beta}_{K0} = \boldsymbol{\beta}_0$, $\boldsymbol{I}_{(K+1)(K+1)} = -\mathbb{E}_{\boldsymbol{\beta}_0}\partial\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}$ is identical to $\boldsymbol{I}^* = -\mathbb{E}_{\boldsymbol{\beta}_0}\partial\boldsymbol{U}^*_{\beta}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}$, and that $\boldsymbol{I}_{(K+1)k} = -\mathbb{E}_{\boldsymbol{\theta}_0}\partial\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_k = \mathbf{0}$, since

$$\begin{aligned}
&-\partial\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}' \\
&= \sum_{j=1}^{d}\Bigg\{\frac{\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}}{\sum_{k=1}^K\left[\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+\sum_{l\in R_j(k)}\boldsymbol{X}'e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_k)\right]} - \\
&\frac{\sum_{k=1}^K\left[\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+\sum_{l\in R_j(k)}\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_k)\right]\left[\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}'e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right]}{\left(\sum_{k=1}^K\left[\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+\sum_{l\in R_j(k)}\boldsymbol{X}'e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_k)\right]\right)^2}\Bigg\}\Bigg|_{\boldsymbol{\beta}_1=\ldots=\boldsymbol{\beta}_K=\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= \sum_{j=1}^{d}\left[\frac{\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}}{\sum_{k=1}^K\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}} - \frac{\left[\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right]\left[\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right]'}{\left(\sum_{k=1}^K\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right)^2}\right] \\
&= -\partial\boldsymbol{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}',
\end{aligned}$$

and

$$\begin{aligned}
&-\partial\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_k' \\
&= \sum_{j=1}^{d}\Bigg\{\frac{\sum_{l\in R_j(k)}\left[\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+\left(\boldsymbol{X}_l\boldsymbol{X}_l'\boldsymbol{X}_l e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right)(\boldsymbol{\beta}-\boldsymbol{\beta}_k)'-\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}\right]}{\sum_{k=1}^K\left[\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+(\sum_{l\in R_j(k)}\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l})(\boldsymbol{\beta}-\boldsymbol{\beta}_k)\right]} - \\
&\frac{\left(\sum_{k=1}^K\sum_{l\in R_j(k)}[\boldsymbol{X}_l e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_k)]\right)\left(\sum_{k=1}^K\sum_{l\in R_j(k)}[\boldsymbol{X}_l\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_k)]\right)}{\left(\sum_{k=1}^K\sum_{l\in R_j(k)}e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l}+(\sum_{l\in R_j(k)}\boldsymbol{X}_l' e^{\boldsymbol{\beta}_k'\boldsymbol{X}_l})(\boldsymbol{\beta}-\boldsymbol{\beta}_k)\right)^2}\Bigg\}\Bigg|_{\boldsymbol{\beta}_1=\ldots=\boldsymbol{\beta}_K=\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= \mathbf{0}
\end{aligned}$$

Therefore, with $\boldsymbol{\beta}_{10} = \ldots = \boldsymbol{\beta}_{K0} = \boldsymbol{\beta}$, $\Sigma^{(1)} = \boldsymbol{I}^{*-1}(\boldsymbol{\beta}_0)$ as given in (2.5). This completes the proof. □

Proof of Theorem 2.2. The proof for Theorem 2.2 follows similar argument by replacing the score functions $\boldsymbol{U}_{\boldsymbol{\theta}}$ to $\boldsymbol{S}_{\boldsymbol{\theta}}$, $\boldsymbol{U}^*_{\boldsymbol{\beta}_k}$ to $\boldsymbol{S}^*_{\boldsymbol{\beta}_k}$ and $\boldsymbol{U}^*_{\boldsymbol{\beta}}$ to $\boldsymbol{S}^*_{\boldsymbol{\beta}}$. □

S1.2 Robust Sandwich Variance

We derive robust variance estimator for $\hat{\theta}^{(1)}$ for the unstratified model. The estimating equations are equation (2.3) and the estimating equations (2.4) for the K sites. The formula of the proposed robust sandwich variance estimator is $\hat{\Sigma}_{rob}^{(1)} = \boldsymbol{S}^{(1)\prime}\boldsymbol{B}^{(1)}\boldsymbol{S}^{(1)}$, where

$$\boldsymbol{S}^{(1)} = \begin{bmatrix} -\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}}^{(1)})\boldsymbol{I}_1^{-1}(\hat{\boldsymbol{\theta}}^{(1)})\hat{\boldsymbol{I}}_{11}^{*-1}(\hat{\boldsymbol{\beta}}_1) \\ \vdots \\ -\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}}^{(1)})\boldsymbol{I}_K^{-1}(\hat{\boldsymbol{\theta}}^{(1)})\hat{\boldsymbol{I}}_{KK}^{*-1}(\hat{\boldsymbol{\beta}}_K) \\ \hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}}^{(1)}) \end{bmatrix}$$

and

$$\boldsymbol{B}^{(1)} = \begin{bmatrix} \sum_{i\in\Omega_1}\omega_i(\hat{\boldsymbol{\beta}}_1)^{\otimes 2} & \dots & \boldsymbol{0} & \sum_{i\in\Omega_1}\omega_i(\hat{\boldsymbol{\beta}}_1)\nu_i(\hat{\boldsymbol{\beta}}^{(1)})' \\ \vdots & & \vdots & \vdots \\ \boldsymbol{0} & \dots & \sum_{i\in\Omega_K}\omega_i(\hat{\boldsymbol{\beta}}_K)^{\otimes 2} & \sum_{i\in\Omega_K}\omega_i(\hat{\boldsymbol{\beta}}_K)\nu_i(\hat{\boldsymbol{\beta}}^{(1)})' \\ \sum_{i\in\Omega_1}\nu_i(\hat{\boldsymbol{\beta}}^{(1)})\omega_i(\hat{\boldsymbol{\beta}}_1)' & \dots & \sum_{i\in\Omega_K}\nu_i(\hat{\boldsymbol{\beta}}^{(1)})\omega_i(\hat{\boldsymbol{\beta}}_K)' & \sum_{i\in\Omega}\nu_i(\hat{\boldsymbol{\beta}}^{(1)})^{\otimes 2} \end{bmatrix}$$

where Ω_k is the index set for site K, and $\omega_i(\hat{\boldsymbol{\beta}}_k) = \delta_i\Big[\boldsymbol{X}_i - \frac{\sum_{l\in R_i}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}}{\sum_{l\in R_i} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}}\Big] - \sum_{j\in D_k}\Big[\frac{I(i\in R_j)e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_i}}{\sum_{l\in R_j} e^{\hat{\boldsymbol{\beta}}_k\boldsymbol{X}_l}}\Big(\boldsymbol{X}_i - \frac{\sum_{l\in R_j}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}}{\sum_{l\in R_j} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}}\Big)\Big]$, which is the estimates of the independent and identically distributed (iid) term derived as in Lin and Wei (1989), and for $i \in \Omega_k$,

$$\begin{aligned}\nu_i(\hat{\boldsymbol{\beta}}^{(1)}) =& \delta_i\Big[\boldsymbol{X}_i - \frac{\sum_{k=1}^K\sum_{l\in R_i(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l} + \sum_{k=1}^K\sum_{l\in R_i(k)}\boldsymbol{X}_l\boldsymbol{X}_l' e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)}{\sum_{k=1}^K\sum_{l\in R_i(k)} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l} + \sum_{k=1}^K\sum_{l\in R_i(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)}\Big] \\ & - \sum_{j=1}^d\Big[\frac{I(i\in R_j)\{e^{\hat{\boldsymbol{\beta}}_k\boldsymbol{X}_i} + \boldsymbol{X}_i e^{\hat{\boldsymbol{\beta}}_k\boldsymbol{X}_i}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)\}}{\sum_{k=1}^K\sum_{l\in R_j(k)} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l} + \sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)} \times \\ & \Big(\boldsymbol{X}_i - \frac{\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l} + \sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)}{\sum_{k=1}^K\sum_{l\in R_j(k)} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l} + \sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)}\Big)\Big],\end{aligned}$$

which can be viewed as a Taylor approximation of the iid terms. Each site can calculate its $\sum_{i\in\Omega_k}\omega_i(\hat{\boldsymbol{\beta}}_k)^{\otimes 2}$ and send it to the analysis center. To calculate the last row in the matrix $\boldsymbol{B}^{(1)}$, the center needs to send back to each site the summed terms $\sum_{k=1}^K\sum_{l\in R_j(k)} e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}$, $\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}' e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)$, and $\sum_{k=1}^K\sum_{l\in R_j(k)}\boldsymbol{X}_l\boldsymbol{X}_l' e^{\hat{\boldsymbol{\beta}}_k'\boldsymbol{X}_l}(\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_k)$ for $j = 1, \dots, d$, together with the $\hat{\boldsymbol{\beta}}^{(1)}$. Thus, two additional file transfers are required if one wants to evaluate the robust variance estimator.

S1.3 *Summary-level Information*

Table S1 shows the summary level information required for the unstratified and stratified distributed methods.

Table S1: File transfers for unstratified and stratified distributed analysis.

	Unstratified		**Stratified**	
No. of transfers	3		1	
Transfer 1	(site to center): Information on observed times †		(not required)	
Transfer 2	(center to site): Information on observed times †		(not required)	
Transfer 3	**Summary-level statistics**	count	**Summary-level statistics from each site**	count
Scalars	$\sum_{l\in R_j(k)} e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d	$\sum_{l\in R_j(k)} e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d_k
Vectors	$\sum_{j\in D_k} \boldsymbol{X}_{i(j)}$	1	$\sum_{j\in D_k} \boldsymbol{X}_{i(j)}$	1
	$\hat{\boldsymbol{\beta}}_k$	1	$\hat{\boldsymbol{\beta}}_k$	1
	$\sum_{l\in R_j(k)} \boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d	$\sum_{l\in R_j(k)} \boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d_k
	$\sum_{l\in R_j(k)} \boldsymbol{X}_l \boldsymbol{X}_l' \boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$ ‡	d	$\sum_{l\in R_j(k)} \boldsymbol{X}_l \boldsymbol{X}_l' \boldsymbol{X}_l e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$ ‡	d_k
Matrices	$\sum_{l\in R_j(k)} \boldsymbol{X}_l \boldsymbol{X}_l' e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d	$\sum_{l\in R_j(k)} \boldsymbol{X}_l \boldsymbol{X}_l' e^{\hat{\boldsymbol{\beta}}_k' \boldsymbol{X}_l}$	d_k

†: There are three methods for transfer 1 and transfer 2:
(1) Sites send observed original or monotonically transformed failure times to center. Center sends merged times to site.
(2) Sites send observed (failure or censoring) times to center. Center sends the overall ranks of the respective site's observed times and the ranks of all failure times to site.
(3) Sites send observed (failure or censoring) times to center. Center send the risk sets to each site.
‡: required for variance estimation

S2. Additional Details and Results for Section 3

Definition of confidence interval overlap statistic (cios)

Cios is a measure of the degree of overlap between confidence intervals obtained based on the same regression fit using the distributed data and the pooled data. Let β_1 be the coefficient (e.g., the log hazard ratio for treatment) in the Cox PH regression model, following Karr *and others* (2006), the cios for the estimated $\hat{\beta}_1$ is defined as follows. Let $(L_{\text{dist}}, U_{\text{dist}})$ be the 95% confidence interval for the coefficient $\hat{\beta}_1$ obtained from the distributed (stratified or unstratified) regression analysis, and let $(L_{\text{pool}}, U_{\text{pool}})$ be the corresponding interval for the estimate $\tilde{\beta}_1$ obtained from the corresponding (stratified or unstratified) regression analysis using the pooled data. Let $\hat{f}_{\text{dist}}$ and $\hat{f}_{\text{pool}}$ be the estimated posterior distribution of β_1 computed under the distributed regression and the pooled regression, respectively. In our simulation setting, it was a normal distribution with mean $\hat{\beta}_1$ or $\tilde{\beta}_1$ and variance the respective estimated variance. The cios is defined as $\frac{1}{2}\left[\int_{L_{\text{dist}}}^{U_{\text{dist}}} \hat{f}_{\text{pool}}(t)dt + \int_{L_{\text{pool}}}^{U_{\text{pool}}} \hat{f}_{\text{dist}}(t)dt\right]$. By design, $0 \leqslant \text{cios} \leqslant 0.95$ with effectively no overlap corresponding to $\text{cios} = 0$ and perfect overlap corresponding to $\text{cios} = 0.95$.

Table S2: Simulation parameter specification.

	K	p	n_k	X distributions (for $X_1; \ldots; X_p$)
Table 1, S3, S5	3	3	as shown in tables	B(0.5); N(0,1); B(0.5)
Table 2	3	3	as shown in tables	site1:B(0.5); N(0,1); X_2+N(0,0.04)
				site2:B(0.5); $5X_1$ + N(0, 0.5); N(2, 0.5)
				site3: Ber(0.9); $1/3X_1$ + N(0, 0.04); N(-1, 1.5)
Table 3 and Table S4	3	5	(167, 167, 166); (83, 166, 251)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7)
	3	7	(167, 167, 166); (83, 166, 251)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7); N(0, 0.5); Ber(0.2)
	5	5	$n_k = 100$ for all k; (50, 50, 50, 175, 175)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7)
	10	5	$n_k = 50$ for all k; (25, 25, 25, 25, 25, 75, 75, 75, 75, 75)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7)
	5	7	$n_k = 100$ for all k; (50, 50, 50, 175, 175)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7); N(0, 0.5); Ber(0.2)
	10	7	$n_k = 50$ for all k; (25, 25, 25, 25, 25, 75, 75, 75, 75, 75)	B(0.5); N(0,1); Ber(0.3); N(0, 0.25); Ber(0.7); N(0, 0.5); Ber(0.2)
	all other settings			X_{odd}: B(0.5); X_{even}: N(0,1)

Table S3: Comparing distributed Cox regression (Dist), pooled Cox regression (Pooled), multivariate meta-analysis (MulV), and univariate meta-analysis (UniV) with simulated data of sample sizes $n = 500, 1000, 3000$ under two data generating models: unstratified and stratified model. Sizes for the data-contributing sites n_k are different and provided in the table below. Covariates distributions for all the data-contributing sites are the same. True log HR value is $\beta_{01} = 1/3$. Number of replications = 10000.

	Data Generating Model: unstratified							Data Generating Model: stratified					
Method	Unstratified		Stratified		Meta-Analysis			Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV		Pooled	Dist	Pooled	Dist	MulV	UniV
	$K = 3$; $n_k = (83, 167, 250)$; $p = 3$							$K = 3$; $n_k = (83, 167, 250)$; $p = 3$					
bias	0.0020	0.0135	0.0019	0.0138	0.0011	0.0032	bias	-0.0223	-0.0113	0.0020	0.0140	0.0013	0.0034
cr[1]	0.9490	0.9532	0.9524	0.9530	0.9529	0.9505	cr[1]	0.9414	0.9648	0.9491	0.9514	0.9499	0.9501
mae[2]	0.0789	0.0809	0.0800	0.0822	0.0798	0.0808	mae[2]	0.0827	0.0821	0.0809	0.0830	0.0807	0.0818
cios[3]	0.9500	0.9450	0.9471	0.9417	0.9471	0.9453	cios[3]	0.9215	0.9261	0.9500	0.9452	0.9500	0.9483
sse[4]	0.0991	0.1007	0.1004	0.1023	0.1002	0.1014	sse	0.1011	0.1022	0.1014	0.1032	0.1011	0.1024
ese[5]	0.0992	0.1015	0.1005	0.1038	0.1004	0.1011	ese	0.1000	0.1025	0.1015	0.1045	0.1014	0.1021
rse[6]	0.0984	0.1029					rse	0.0999	0.1113				
	$K = 3$; $n_k = (167, 333, 500)$; $p = 3$							$K = 3$; $n_k = (167, 333, 500)$; $p = 3$					
bias	0.0010	0.0065	0.0009	0.0066	0.0006	0.0017	bias	-0.0241	-0.0180	0.0009	0.0067	0.0006	0.0016
cr	0.9548	0.9551	0.9537	0.9530	0.9539	0.9531	cr	0.9359	0.9613	0.9543	0.9541	0.9544	0.9547
mae	0.0548	0.0553	0.0552	0.0558	0.0551	0.0554	mae	0.0592	0.0580	0.0557	0.0563	0.0556	0.0560
cios	0.9500	0.9478	0.9484	0.9463	0.9484	0.9476	cios	0.9153	0.9227	0.9500	0.9479	0.9500	0.9492
sse	0.0689	0.0693	0.0694	0.0699	0.0693	0.0697	sse	0.0703	0.0706	0.0701	0.0706	0.0700	0.0704
ese	0.0698	0.0705	0.0703	0.0711	0.0702	0.0705	ese	0.0704	0.0712	0.0710	0.0717	0.0709	0.0712
rse	0.0695	0.0709					rse	0.0706	0.0772				
	$K = 3$; $n_k = (500, 1000, 1500)$; $p = 3$							$K = 3$; $n_k = (500, 1000, 1500)$; $p = 3$					
bias	0.0001	0.0018	0.0000	0.0018	-0.0001	0.0002	bias	-0.0250	-0.0221	-0.0001	0.0017	-0.0002	0.0001
cr	0.9492	0.9487	0.9488	0.9476	0.9488	0.9486	cr	0.9051	0.9391	0.9491	0.9493	0.9489	0.9492
mae	0.0321	0.0322	0.0322	0.0323	0.0322	0.0323	mae	0.0385	0.0372	0.0324	0.0325	0.0324	0.0325
cios	0.9500	0.9494	0.9494	0.9488	0.9494	0.9491	cios	0.8854	0.8995	0.9500	0.9494	0.9500	0.9497
sse	0.0404	0.0404	0.0405	0.0406	0.0405	0.0406	sse	0.0408	0.0408	0.0407	0.0408	0.0407	0.0408
ese	0.0401	0.0403	0.0402	0.0404	0.0402	0.0403	ese	0.0405	0.0407	0.0406	0.0408	0.0406	0.0407
rse	0.0401	0.0403					rse	0.0407	0.0441				

cr[1]: coverage rate; mae[2]: mean absolute error; cios[3]: confidence interval overlap statistics
sse[4]: sample standard error; ese[5]: estimated standard error; rse[6]: robust standard error estimate

Table S4: Comparing distributed Cox regression, pooled Cox regression, and meta-analyses with simulated data for a varying number of sites K, site sizes n_k, and number of covariates p with data generated from the unstratified Cox PH models. Sizes for the contributing-sites n_k are provided in the table below. Covariate distributions for all the data-contributing sites are the same. True log HR for treatment is $\beta_{01} = 1/p$. Number of replications = 10000.

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=3, n_k=(20,20,20), p=1, \beta_{01}=1$					
bias	-0.0371	0.0558	-0.0471	0.0452	-0.0668	-0.0668
cr[1]	0.9631	0.9676	0.9651	0.9803	0.9634	0.9634
mae[2]	0.2345	0.2403	0.2473	0.2501	0.2485	0.2485
cios[3]	0.9500	0.9243	0.9358	0.9147	0.9343	0.9343
sse[4]	0.2914	0.2932	0.3045	0.3052	0.3027	0.3027
ese[5]	0.3147	0.3532	0.3328	0.3846	0.3335	0.3335
rse[6]	0.3052	0.3526				
	$K=3, n_k=(40,40,40), p=1, \beta_{01}=1$					
bias	0.0024	0.0568	0.0025	0.0575	-0.0084	-0.0084
cr	0.9561	0.9546	0.9570	0.9614	0.9577	0.9577
mae	0.1722	0.1830	0.1780	0.1891	0.1767	0.1767
cios	0.9500	0.9338	0.9417	0.9268	0.9412	0.9412
sse	0.2161	0.2220	0.2230	0.2294	0.2212	0.2212
ese	0.2201	0.2344	0.2276	0.2464	0.2277	0.2277
rse	0.2163	0.2358				
	$K=3, n_k=(50,50,50), p=1, \beta_{01}=1$					
bias	0.0062	0.0497	0.0055	0.0496	-0.0029	-0.0029
cr	0.9530	0.9531	0.9520	0.9550	0.9525	0.9525
mae	0.1561	0.1651	0.1602	0.1694	0.1590	0.1590
cios	0.9500	0.9371	0.9432	0.9315	0.9428	0.9428
sse	0.1958	0.2019	0.2018	0.2081	0.2002	0.2002
ese	0.1961	0.2057	0.2016	0.2139	0.2016	0.2016
rse	0.1934	0.2068				
	$K=3, n_k=(20,20,20), p=2, \beta_{01}=1/2$					
bias	-0.0558	0.0363	-0.0726	0.0149	-0.0808	-0.0828
cr	0.9625	0.9752	0.9634	0.9844	0.9642	0.9654
mae	0.2368	0.2376	0.2512	0.2480	0.2487	0.2567
cios	0.9500	0.9209	0.9317	0.9101	0.9312	0.9262
sse	0.2920	0.2929	0.3083	0.3078	0.3033	0.3133
ese	0.3147	0.3606	0.3395	0.3971	0.3395	0.3484
rse	0.3016	0.3651				
	$K=3, n_k=(60,60,60), p=2, \beta_{01}=1/2$					
bias	0.0022	0.0377	0.0015	0.0376	-0.0023	0.0007
cr	0.9537	0.9546	0.9533	0.9573	0.9548	0.9522
mae	0.1417	0.1482	0.1459	0.1526	0.1450	0.1471
cios	0.9500	0.9384	0.9423	0.9318	0.9422	0.9403
sse	0.1766	0.1808	0.1819	0.1866	0.1808	0.1835
ese	0.1759	0.1842	0.1816	0.1920	0.1812	0.1828
rse	0.1729	0.1878				

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=3, n_k=(80,80,80), p=2, \beta_{01}=1/2$					
bias	0.0031	0.0296	0.0033	0.0299	0.0004	0.0031
cr	0.9509	0.9537	0.9500	0.9541	0.9506	0.9507
mae	0.1211	0.1256	0.1241	0.1286	0.1234	0.1247
cios	0.9500	0.9414	0.9440	0.9363	0.9440	0.9425
sse	0.1520	0.1549	0.1557	0.1588	0.1549	0.1565
ese	0.1515	0.1564	0.1553	0.1612	0.1551	0.1561
rse	0.1495	0.1588				
	$K=3, n_k=(100,100,100), p=2, \beta_{01}=1/2$					
bias	0.0008	0.0215	0.0008	0.0217	-0.0014	0.0010
cr	0.9466	0.9481	0.9485	0.9486	0.9481	0.9477
mae	0.1103	0.1134	0.1131	0.1161	0.1126	0.1137
cios	0.9500	0.9435	0.9451	0.9393	0.9451	0.9438
sse	0.1376	0.1400	0.1408	0.1433	0.1402	0.1414
ese	0.1351	0.1383	0.1379	0.1416	0.1377	0.1385
rse	0.1336	0.1399				
	$K=3, n_k=(50,50,50), p=3, \beta_{01}=1/3$					
bias	-0.0079	0.0291	-0.0088	0.0286	-0.0119	-0.0070
cr	0.9561	0.9688	0.9580	0.9689	0.9596	0.9554
mae	0.1431	0.1482	0.1484	0.1538	0.1474	0.1537
cios	0.9500	0.9318	0.9414	0.9259	0.9413	0.9353
sse	0.1795	0.1841	0.1858	0.1908	0.1844	0.1920
ese	0.1851	0.2021	0.1917	0.2132	0.1914	0.1960
rse	0.1809	0.2112				
	$K=3, n_k=(100,100,100), p=3, \beta_{01}=1/3$					
bias	0.0006	0.0200	0.0005	0.0203	-0.0009	0.0030
cr	0.9474	0.9511	0.9459	0.9491	0.9475	0.9440
mae	0.1043	0.1078	0.1063	0.1100	0.1058	0.1080
cios	0.9500	0.9415	0.9452	0.9374	0.9451	0.9421
sse	0.1309	0.1342	0.1333	0.1368	0.1328	0.1354
ese	0.1289	0.1338	0.1315	0.1372	0.1313	0.1329
rse	0.1273	0.1371				
	$K=3, n_k=(83,166,251), p=5, \beta_{01}=1/5$					
bias	0.0007	0.0117	0.0006	0.0121	0.0000	0.0026
cr	0.9492	0.9564	0.9488	0.9533	0.9500	0.9484
mae	0.0771	0.0793	0.0778	0.0804	0.0776	0.0794
cios	0.9500	0.9421	0.9471	0.9388	0.9471	0.9435
sse	0.0972	0.0993	0.0981	0.1008	0.0978	0.1002
ese	0.0968	0.1010	0.0979	0.1043	0.0978	0.0993
rse	0.0957	0.1042				

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=3, n_k=(166,167,167), p=7, \beta_{01}=1/7$					
bias	0.0003	0.0110	0.0000	0.0108	-0.0005	0.0023
cr	0.9498	0.9615	0.9512	0.9563	0.9518	0.9486
mae	0.0780	0.0802	0.0790	0.0812	0.0788	0.0823
cios	0.9500	0.9420	0.9470	0.9406	0.9469	0.9408
sse	0.0977	0.0997	0.0989	0.1011	0.0986	0.1030
ese	0.0976	0.1016	0.0988	0.1033	0.0987	0.1010
rse	0.0961	0.1058				
	$K=3, n_k=(83,166,251), p=7, \beta_{01}=1/7$					
bias	-0.0005	0.0098	-0.0008	0.0100	-0.0013	0.0008
cr	0.9507	0.9608	0.9493	0.9554	0.9494	0.9462
mae	0.0775	0.0795	0.0782	0.0806	0.0781	0.0814
cios	0.9500	0.9417	0.9471	0.9383	0.9470	0.9415
sse	0.0973	0.0992	0.0984	0.1007	0.0981	0.1023
ese	0.0976	0.1022	0.0988	0.1059	0.0987	0.1010
rse	0.0962	0.1062				
	$K=5, n_k=(20,20,20,20,20), p=1, \beta_{01}=1$					
bias	-0.0301	0.0850	-0.0378	0.0828	-0.0627	-0.0627
cr	0.9635	0.9740	0.9637	0.9873	0.9594	0.9594
mae	0.1811	0.1972	0.1920	0.2089	0.1939	0.1939
cios	0.9500	0.9075	0.9320	0.8854	0.9295	0.9295
sse	0.2236	0.2287	0.2363	0.2421	0.2334	0.2334
ese	0.2406	0.2848	0.2580	0.3426	0.2587	0.2587
rse	0.2353	0.2887				
	$K=5, n_k=(10,10,10,10,10), p=1, \beta_{01}=1$					
bias	-0.1224	0.0758	-0.1659	0.0286	-0.2034	-0.2034
cr	0.9536	0.9751	0.9525	0.9921	0.9525	0.9525
mae	0.2598	0.2519	0.2856	0.2589	0.2964	0.2964
cios	0.9500	0.8832	0.9200	0.8577	0.9138	0.9138
sse	0.3018	0.3049	0.3198	0.3153	0.3101	0.3101
ese	0.3434	0.5748	0.3847	0.5910	0.3892	0.3892
rse	0.3315	0.4525				
	$K=5, n_k=(30,30,30,30,30), p=1, \beta_{01}=1$					
bias	-0.0052	0.0779	-0.0064	0.0818	-0.0237	-0.0237
cr	0.9592	0.9629	0.9577	0.9688	0.9586	0.9586
mae	0.1505	0.1697	0.1579	0.1787	0.1573	0.1573
cios	0.9500	0.9175	0.9375	0.9007	0.9360	0.9360
sse	0.1888	0.1961	0.1982	0.2067	0.1961	0.1961
ese	0.1958	0.2199	0.2059	0.2521	0.2063	0.2063
rse	0.1928	0.2237				

(to be continued)

cr[1]: coverage rate; mae[2]: mean absolute error; cios[3]: confidence interval overlap statistics
sse[4]: sample standard error; ese[5]: estimated standard error; rse[6]: robust standard error estimate

(continued from the previous page)

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=5, n_k=(40,40,40,40,40), p=1, \beta_{01}=1$					
bias	0.0003	0.0647	-0.0009	0.0674	-0.0135	-0.0135
cr[1]	0.9527	0.9529	0.9516	0.9602	0.9532	0.9532
mae[2]	0.1337	0.1480	0.1386	0.1552	0.1377	0.1377
cios[3]	0.9500	0.9240	0.9402	0.9112	0.9391	0.9391
sse[4]	0.1674	0.1743	0.1738	0.1824	0.1719	0.1719
ese[5]	0.1690	0.1838	0.1758	0.2035	0.1760	0.1760
rse[6]	0.1670	0.1864				
	$K=5, n_k=(50,50,50,50,50), p=1, \beta_{01}=1$					
bias	0.0038	0.0565	0.0031	0.0589	-0.0069	-0.0069
cr	0.9540	0.9504	0.9531	0.9562	0.9541	0.9541
mae	0.1190	0.1314	0.1231	0.1370	0.1220	0.1220
cios	0.9500	0.9281	0.9420	0.9176	0.9412	0.9412
sse	0.1497	0.1561	0.1546	0.1628	0.1530	0.1530
ese	0.1509	0.1610	0.1560	0.1744	0.1561	0.1561
rse	0.1495	0.1629				
	$K=5, n_k=(40,40,40,40,40), p=2, \beta_{01}=1/2$					
bias	-0.0055	0.0551	-0.0084	0.0557	-0.0154	-0.0112
cr	0.9498	0.9633	0.9520	0.9671	0.9542	0.9514
mae	0.1320	0.1437	0.1391	0.1515	0.1378	0.1402
cios	0.9500	0.9237	0.9365	0.9085	0.9362	0.9329
sse	0.1654	0.1707	0.1738	0.1800	0.1715	0.1755
ese	0.1664	0.1834	0.1756	0.2055	0.1752	0.1780
rse	0.1635	0.1904				
	$K=5, n_k=(20,20,20,20,20), p=2, \beta_{01}=1/2$					
bias	-0.0450	0.0608	-0.0570	0.0505	-0.0669	-0.0668
cr	0.9613	0.9777	0.9622	0.9884	0.9649	0.9652
mae	0.1830	0.1923	0.2000	0.2037	0.1977	0.2050
cios	0.9500	0.9069	0.9262	0.8825	0.9257	0.9198
sse	0.2244	0.2294	0.2428	0.2471	0.2374	0.2464
ese	0.2383	0.2880	0.2616	0.3470	0.2616	0.2697
rse	0.2310	0.2974				
	$K=5, n_k=(60,60,60,60,60), p=2, \beta_{01}=1/2$					
bias	0.0002	0.0422	0.0007	0.0447	-0.0039	0.0004
cr	0.9504	0.9564	0.9519	0.9569	0.9530	0.9508
mae	0.1076	0.1157	0.1122	0.1213	0.1112	0.1134
cios	0.9500	0.9314	0.9406	0.9202	0.9404	0.9378
sse	0.1351	0.1392	0.1405	0.1455	0.1393	0.1422
ese	0.1351	0.1436	0.1404	0.1543	0.1401	0.1416
rse	0.1335	0.1482				

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=5, n_k=(80,80,80,80,80), p=2, \beta_{01}=1/2$					
bias	0.0027	0.0343	0.0026	0.0354	-0.0008	0.0024
cr	0.9520	0.9505	0.9504	0.9489	0.9516	0.9514
mae	0.0936	0.0991	0.0962	0.1021	0.0956	0.0970
cios	0.9500	0.9360	0.9427	0.9283	0.9425	0.9409
sse	0.1170	0.1200	0.1205	0.1239	0.1198	0.1215
ese	0.1166	0.1216	0.1202	0.1274	0.1200	0.1209
rse	0.1156	0.1246				
	$K=5, n_k=(100,100,100,100,100), p=2, \beta_{01}=1/2$					
bias	0.0035	0.0283	0.0034	0.0290	0.0008	0.0035
cr	0.9505	0.9499	0.9491	0.9484	0.9494	0.9484
mae	0.0838	0.0884	0.0855	0.0904	0.0850	0.0860
cios	0.9500	0.9391	0.9439	0.9327	0.9438	0.9424
sse	0.1048	0.1074	0.1072	0.1102	0.1066	0.1079
ese	0.1041	0.1073	0.1067	0.1112	0.1065	0.1072
rse	0.1033	0.1096				
	$K=5, n_k=(100,100,100,100,100), p=3, \beta_{01}=1/3$					
bias	0.0020	0.0252	0.0018	0.0262	0.0001	0.0043
cr	0.9520	0.9568	0.9512	0.9532	0.9521	0.9496
mae	0.0792	0.0841	0.0809	0.0865	0.0805	0.0828
cios	0.9500	0.9366	0.9443	0.9302	0.9442	0.9401
sse	0.0990	0.1021	0.1012	0.1050	0.1007	0.1036
ese	0.0992	0.1043	0.1016	0.1089	0.1015	0.1029
rse	0.0984	0.1080				
	$K=5, n_k=(50,50,50,175,175), p=3, \beta_{01}=1/3$					
bias	0.0016	0.0238	0.0017	0.0261	0.0000	0.0035
cr	0.9519	0.9578	0.9516	0.9603	0.9517	0.9503
mae	0.0793	0.0832	0.0808	0.0861	0.0804	0.0826
cios	0.9500	0.9366	0.9445	0.9236	0.9444	0.9410
sse	0.0992	0.1018	0.1013	0.1050	0.1008	0.1034
ese	0.0992	0.1052	0.1015	0.1173	0.1013	0.1028
rse	0.0984	0.1086				
	$K=5, n_k=(100,100,100,100,100), p=5, \beta_{01}=1/5$					
bias	0.0005	0.0221	0.0000	0.0232	-0.0012	0.0040
cr	0.9498	0.9690	0.9495	0.9607	0.9500	0.9453
mae	0.0787	0.0840	0.0805	0.0867	0.0800	0.0848
cios	0.9500	0.9287	0.9443	0.9237	0.9442	0.9360
sse	0.0982	0.1022	0.1006	0.1055	0.1000	0.1058
ese	0.0967	0.1065	0.0990	0.1136	0.0989	0.1019
rse	0.0957	0.1147				

	Unstratified		Stratified		Meta-Analysis	
	Pooled	Dist	Pooled	Dist	MulV	UniV
	$K=10, n_k=(25,25,25,25,25,75,75,75,75,75), p=3, \beta_{01}=1/3$					
bias	-0.0003	0.0457	-0.0006	0.0506	-0.0043	0.0029
cr	0.9532	0.9685	0.9502	0.9752	0.9540	0.9467
mae	0.0784	0.0903	0.0825	0.0967	0.0814	0.0862
cios	0.9500	0.9093	0.9382	0.8661	0.9380	0.9300
sse	0.0983	0.1033	0.1033	0.1100	0.1019	0.1082
ese	0.0992	0.1152	0.1040	0.1677	0.1038	0.1070
rse	0.0982	0.1229				
	$K=10, n_k=(50,50,50,50,50,50,50,50,50,50), p=5, \beta_{01}=1/5$					
bias	-0.0075	0.0333	-0.0083	0.0381	-0.0113	-0.0068
cr	0.9568	0.9923	0.9551	0.9916	0.9571	0.9434
mae	0.0758	0.0852	0.0798	0.0922	0.0790	0.0888
cios	0.9500	0.8895	0.9383	0.8563	0.9379	0.9211
sse	0.0948	0.1013	0.0995	0.1091	0.0981	0.1107
ese	0.0967	0.1253	0.1013	0.1693	0.1012	0.1082
rse	0.0954	0.1412				
	$K=10, n_k=(25,25,25,25,25,75,75,75,75,75), p=5, \beta_{01}=1/5$					
bias	-0.0023	0.0367	-0.0033	0.0384	-0.0055	-0.0015
cr	0.9449	0.9780	0.9439	0.9870	0.9429	0.9319
mae	0.0794	0.0912	0.0827	0.0950	0.0811	0.0912
cios	0.9500	0.8924	0.9391	0.8465	0.9386	0.9220
sse	0.1002	0.1081	0.1046	0.1132	0.1026	0.1157
ese	0.0967	0.1239	0.1012	0.1885	0.1011	0.1077
rse	0.0955	0.1388				
	$K=10, n_k=(50,50,50,50,50,50,50,50,50,50), p=7, \beta_{01}=1/7$					
bias	-0.0064	0.0314	-0.0078	0.0338	-0.0096	-0.0136
cr	0.9339	0.9839	0.9387	0.9887	0.9371	0.9419
mae	0.0818	0.0894	0.0838	0.0936	0.0823	0.0984
cios	0.9500	0.8895	0.9383	0.8609	0.9373	0.9092
sse	0.1035	0.1095	0.1058	0.1142	0.1038	0.1223
ese	0.0975	0.1279	0.1022	0.1672	0.1022	0.1133
rse	0.0957	0.1440				
	$K=10, n_k=(25,25,25,25,25,75,75,75,75,75), p=7, \beta_{01}=1/7$					
bias	-0.0162	0.0217	-0.0163	0.0194	-0.0182	-0.0212
cr	0.9495	1.0000	0.9394	1.0000	0.9495	0.9394
mae	0.0747	0.0872	0.0828	0.0950	0.0797	0.0917
cios	0.9500	0.8892	0.9407	0.8593	0.9404	0.9186
sse	0.0959	0.1050	0.1036	0.1119	0.1001	0.1143
ese	0.0975	0.1256	0.1021	0.1721	0.1021	0.1125
rse	0.0956	0.1431				

cr[1]: coverage rate; mae[2]: mean absolute error; cios[3]: confidence interval overlap statistics
sse[4]: sample standard error; ese[5]: estimated standard error; rse[6]: robust standard error estimate

Table S5. Comparing distributed Cox regression with different orders of Taylor expansion with data generated from the unstratified Cox PH models. Sample size $n = 3000$. Taylor 1-4 represents the first-order to fourth-order Taylor expansion. The sizes of the data-contributing sites are approximately the same. Covariates distributions of the data-contributing sites are the same. The true log HR for treatment $\beta_{01} = 1/3$. Number of replications = 10000.

Method	Unstratified Pooled	Unstratified Distributed				Stratified Pooled	Stratified Distributed			
		Taylor 1	Taylor 2	Taylor 3	Taylor 4		Taylor 1	Taylor 2	Taylor 3	Taylor 4
bias	0.0035	0.0146	0.0029	0.0037	0.0034	0.0033	0.0145	0.0025	0.0035	0.0033
cr[1]	0.9542	0.9556	0.9457	0.9521	0.9516	0.9523	0.9530	0.9525	0.9522	0.9523
mae[2]	0.0785	0.0805	0.0785	0.0785	0.0785	0.0796	0.0816	0.0796	0.0796	0.0796
cios[3]	0.9500	0.9455	0.9493	0.9498	0.9498	0.9470	0.9429	0.9469	0.9470	0.9470
sse[4]	0.0985	0.0999	0.0984	0.0985	0.0985	0.0999	0.1015	0.0999	0.0999	0.0999
ese[5]	0.0992	0.1012	0.0994	0.0992	0.0992	0.1005	0.1026	0.1006	0.1005	0.1005
rse[6]	0.0984	0.1025	0.0963	0.0985	0.0983					

[1]: coverage rate; [2]: mean absolute error; [3]: confidence interval overlap statistics
sse[4]: sample standard error; ese[5]: estimated standard error; rse[6]: robust standard error estimate

S3. Additional Details and Results for Section 4

Table S6. Comparing distributed Cox regression, pooled Cox regression, and meta-analyses with real data under partition 1. Results are presented as logHR with estimated standard error in brackets.

	Unstratified Pooled	Unstratified Distributed	Stratified Pooled	Stratified Distributed	Univariate Meta-Analysis	Multivariate Meta-Analysis
treatment	-0.359 (0.1007)	-0.366 (0.1043)	-0.359 (0.1007)	-0.366 (0.1045)	-0.361 (0.1012)	-0.362 (0.1011)
age	-0.165 (0.0541)	-0.170 (0.0569)	-0.165 (0.0541)	-0.170 (0.0570)	-0.163 (0.0541)	-0.167 (0.0540)
comorbidity score	0.137 (0.0304)	0.144 (0.0334)	0.137 (0.0304)	0.144 (0.0333)	0.131 (0.0310)	0.142 (0.0307)
diagnosis of cancer	-0.226 (0.1931)	-0.164 (0.2151)	-0.224 (0.1930)	-0.164 (0.2143)	-0.190 (0.1943)	-0.182 (0.1938)
depression	-0.254 (0.1110)	-0.243 (0.1202)	-0.255 (0.1110)	-0.244 (0.1203)	-0.242 (0.1117)	-0.248 (0.1115)
diabetes	-0.049 (0.1103)	-0.042 (0.1152)	-0.048 (0.1103)	-0.041 (0.1151)	-0.051 (0.1108)	-0.050 (0.1106)
eating disorder	0.097 (0.1269)	0.094 (0.1312)	0.098 (0.1269)	0.095 (0.1312)	0.105 (0.1273)	0.101 (0.1272)
gastroesophageal reflux disease	0.322 (0.1049)	0.324 (0.1056)	0.323 (0.1049)	0.325 (0.1057)	0.313 (0.1054)	0.323 (0.1052)
hypertension	0.180 (0.1183)	0.185 (0.1166)	0.180 (0.1183)	0.184 (0.1166)	0.176 (0.1185)	0.184 (0.1182)
non-alcoholic fatty liver disease	-0.264 (0.1280)	-0.276 (0.1336)	-0.264 (0.1280)	-0.276 (0.1336)	-0.236 (0.1285)	-0.268 (0.1280)
number of emergency department visits	-0.500 (0.2305)	-0.510 (0.2469)	-0.496 (0.2306)	-0.507 (0.2463)	-0.511 (0.2347)	-0.524 (0.2338)
number of unique generic medications.	0.714 (0.2189)	0.736 (0.2352)	0.710 (0.2189)	0.732 (0.2349)	0.733 (0.2241)	0.748 (0.2232)

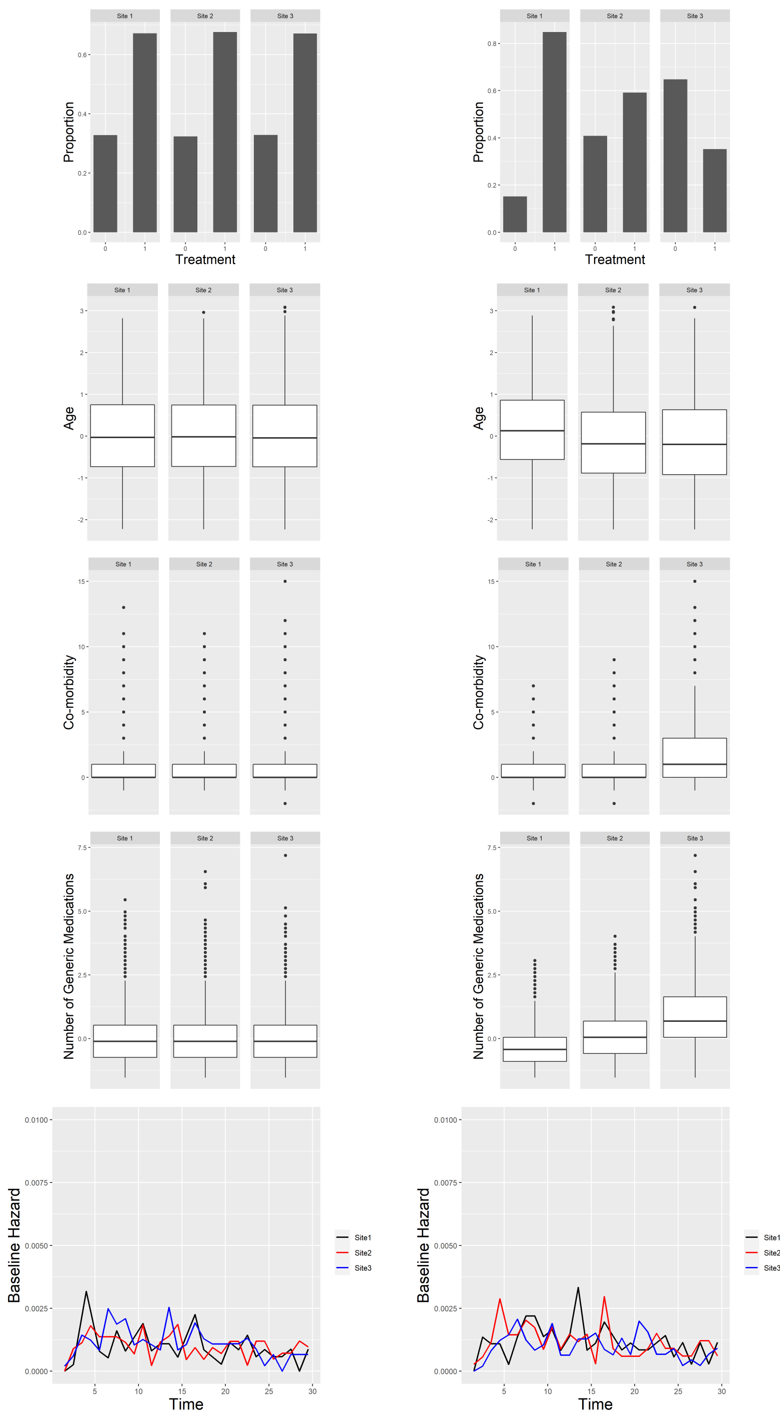
Table S7. Comparing distributed Cox regression, pooled Cox regression, and meta-analyses with real data under partition 2. Results are presented as logHR with estimated standard error in brackets.

	Unstratified Pooled	Unstratified Distributed	Stratified Pooled	Stratified Distributed	Univariate Meta-Analysis	Multivariate Meta-Analysis
treatment	-0.359 (0.1007)	-0.361 (0.1063)	-0.338 (0.1240)	-0.353 (0.1272)	-0.318 (0.1302)	-0.347 (0.1210)
age	-0.165 (0.0541)	-0.167 (0.0587)	-0.157 (0.0616)	-0.163 (0.0649)	-0.139 (0.0656)	-0.154 (0.0612)
comorbidity score	0.137 (0.0304)	0.139 (0.0342)	0.132 (0.0355)	0.139 (0.0379)	0.124 (0.0380)	0.133 (0.0348)
diagnosis of cancer	-0.226 (0.1931)	-0.209 (0.2071)	-0.216 (0.1974)	-0.207 (0.2111)	-0.129 (0.2027)	-0.184 (0.1978)
depression	-0.254 (0.1110)	-0.245 (0.1171)	-0.241 (0.1206)	-0.239 (0.1257)	-0.168 (0.1270)	-0.231 (0.1209)
diabetes	-0.049 (0.1103)	-0.059 (0.1177)	-0.049 (0.1103)	-0.060 (0.1179)	-0.036 (0.1109)	-0.046 (0.1100)
eating disorder	0.097 (0.1269)	0.099 (0.1351)	0.090 (0.1288)	0.094 (0.1366)	0.075 (0.1296)	0.104 (0.1281)
gastroesophageal reflux disease	0.322 (0.1049)	0.313 (0.1068)	0.304 (0.1246)	0.303 (0.1236)	0.239 (0.1336)	0.298 (0.1245)
hypertension	0.180 (0.1183)	0.203 (0.1281)	0.174 (0.1214)	0.203 (0.1308)	0.082 (0.1304)	0.171 (0.1219)
non-alcoholic fatty liver disease	-0.264 (0.1280)	-0.267 (0.1475)	-0.254 (0.1347)	-0.266 (0.1532)	-0.155 (0.1495)	-0.227 (0.1349)
number of emergency department visits	-0.500 (0.2305)	-0.497 (0.2395)	-0.518 (0.2364)	-0.508 (0.2444)	-0.580 (0.2413)	-0.522 (0.2375)
number of unique generic medications.	0.714 (0.2189)	0.705 (0.2267)	0.723 (0.2198)	0.713 (0.2271)	0.731 (0.2252)	0.725 (0.2215)

Table S8: Covariate distribution of the sites under two partitions. Results are presented as mean with standard deviation in brackets.

	partition 1			partition 2		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
treatment	0.67 (0.470)	0.68 (0.468)	0.67 (0.470)	0.85 (0.359)	0.59 (0.492)	0.35 (0.478)
age	0.00 (0.990)	0.01 (1.006)	-0.01 (1.002)	0.14 (0.955)	-0.14 (1.007)	-0.13 (1.053)
comorbidity score	0.65 (1.480)	0.66 (1.466)	0.64 (1.484)	0.25 (1.056)	0.61 (1.320)	1.76 (2.016)
diagnosis of cancer	0.08 (0.273)	0.08 (0.272)	0.08 (0.268)	0.10 (0.297)	0.06 (0.242)	0.06 (0.238)
depression	0.29 (0.456)	0.32 (0.467)	0.32 (0.466)	0.36 (0.481)	0.26 (0.441)	0.26 (0.441)
diabetes	0.35 (0.478)	0.36 (0.480)	0.36 (0.481)	0.32 (0.467)	0.35 (0.478)	0.48 (0.500)
eating disorder	0.17 (0.375)	0.16 (0.371)	0.18 (0.382)	0.13 (0.341)	0.20 (0.401)	0.22 (0.412)
gastroesophageal reflux disease	0.59 (0.492)	0.57 (0.494)	0.58 (0.494)	0.38 (0.486)	0.74 (0.441)	0.84 (0.371)
hypertension	0.65 (0.476)	0.66 (0.475)	0.66 (0.474)	0.61 (0.487)	0.67 (0.471)	0.75 (0.432)
non-alcoholic fatty liver disease	0.21 (0.409)	0.20 (0.402)	0.19 (0.395)	0.22 (0.415)	0.18 (0.388)	0.18 (0.385)
number of emergency department visits	0.01 (1.003)	0.00 (0.988)	-0.01 (1.009)	-0.40 (0.762)	0.10 (0.881)	0.88 (1.127)
number of unique generic medications	0.00 (0.999)	0.00 (0.992)	-0.01 (1.008)	-0.40 (0.740)	0.09 (0.875)	0.89 (1.164)

Fig. S1: Data application: Plots of distribution of treatment, age, co-morbidity score, and number of generic medications and plots of baseline hazard estimates with two data partitions. First column: partition 1. Second column: partition 2.



S4. Illustration Using A Simulated Data Set

This section presents the analysis results from a simulated data set. A detailed illustration for how to use the code to reproduce it can be found on `https://github.com/dli-stats/distributed_cox_paper_repro`.

Table S9: Simulated data analysis for illustrative purpose: comparing distributed Cox regression, pooled Cox regression, and meta-analyses withreal data. Results are presented as logHR with estimated standard error in brackets.

	Unstratified Pooled	Unstratified Distributed	Stratified Pooled	Stratified Distributed	Univariate Meta-Analysis	Multivariate Meta-Analysis
treatment	-0.343(0.0802)	-0.340(0.081)	-0.344(0.0802)	-0.341(0.0811)	-0.344(0.0802)	-0.341(0.0803)
age	-0.483(0.2206)	-0.490(0.2305)	-0.483(0.2206)	-0.490(0.2304)	-0.486(0.2208)	-0.512(0.2214)
comorbidity score	0.092(0.0965)	0.089(0.0978)	0.091(0.0965)	0.088(0.0979)	0.089(0.0966)	0.087(0.0967)
diagnosis of cancer	-0.217(0.1437)	-0.197(0.1507)	-0.217(0.1436)	-0.197(0.1507)	-0.201(0.1438)	-0.200(0.1442)
depression	-0.170(0.0865)	-0.169(0.0886)	-0.169(0.0865)	-0.168(0.0886)	-0.168(0.0866)	-0.166(0.0867)
diabetes	-0.094(0.0829)	-0.095(0.0834)	-0.095(0.0829)	-0.095(0.0834)	-0.095(0.0829)	-0.092(0.0831)
eating disorder	0.087(0.0965)	0.086(0.0976)	0.086(0.0965)	0.085(0.0976)	0.085(0.0965)	0.086(0.0967)
gastroesophageal reflux disease	0.261(0.0807)	0.251(0.0843)	0.261(0.0807)	0.252(0.0843)	0.256(0.081)	0.252(0.0811)
hypertension	0.076(0.0857)	0.077(0.0863)	0.076(0.0857)	0.077(0.0864)	0.074(0.0858)	0.075(0.0859)
non-alcoholic fatty liver disease	-0.098(0.1007)	-0.089(0.1037)	-0.097(0.1007)	-0.089(0.1037)	-0.09(0.1009)	-0.107(0.1012)
number of emergency department visits	-0.479(0.0409)	-0.479(0.0421)	-0.479(0.0409)	-0.479(0.0421)	-0.48(0.0409)	-0.479(0.041)
number of unique generic medications.	1.059(0.2226)	1.066(0.2297)	1.059(0.2226)	1.066(0.2295)	1.063(0.223)	1.092(0.2235)

References

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

Karr, A. F, Kohnen, C. N, Oganian, A, Reiter, J. P and Sanil, A. P. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *The American statistician* **60**(3), 224–232.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association* **84**(408), 1074–1078.

[*Received August 1, 202x; revised October 1, 202x; accepted for publication November 1, 202x*]