

3 **Supplementary Tables**

4 We provide three notation glossaries: the first described general notation conventions, the second
5 describes quantities that pertain to our regression setting, and the last describes quantities that pertain to the vertical setting and the proposed algorithm.

Table S1. Glossary for general notation conventions

Random variable in \mathbb{R}	A	Uppercase Non-italic
Random vector in \mathbb{R}^p	\mathbf{A}	Uppercase Non-italic bold
Scalar in \mathbb{R}	a	Lowercase Italic
Vector in \mathbb{R}^p	\mathbf{a}	Lowercase Italic bold
Vector in \mathbb{R}^p with all components equal to 1	$\mathbf{1}_p$	-
Matrix in $\mathbb{R}^{n \times p}$	\mathbf{A}	Uppercase Italic bold
Identity matrix in $\mathbb{R}^{n \times n}$	\mathbf{I}_n	-
Gradient of $f(\boldsymbol{\theta})$ (column vector)	$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$	$\nabla_{\boldsymbol{\theta}}^2$ for Hessian
$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) _{\boldsymbol{\theta}=\mathbf{a}}$	$\nabla_{\boldsymbol{\theta}} f(\mathbf{a})$	$\nabla_{\boldsymbol{\theta}}^2 f(\mathbf{a})$ for Hessian
$\max_{1 \leq j \leq p} a_j $	$\ \mathbf{a}\ _{\infty}$	Infinite norm
$\sum_{j=1}^p a_j $	$\ \mathbf{a}\ _1$	$\ell 1$ -norm
$\sqrt{\sum_{j=1}^p a_j^2}$	$\ \mathbf{a}\ _2$	$\ell 2$ -norm
Diagonal matrix with entries of \mathbf{a} on diagonal	$\text{diag}(\mathbf{a})$	Dimension $p \times p$ for $\mathbf{a} \in \mathbb{R}^p$
Quantity \cdot at iteration t (step count)	$\cdot_{(t)}$	Starts with $\cdot_{(0)}$

6

Table S2. Glossary for quantities that pertain to the regression settings

Covariate vector for i th individual	$\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]^{\top}$
Covariate vector for i th individual with intercept	$[1, \mathbf{x}_i^{\top}]^{\top}$
Covariate matrix in $\mathbb{R}^{n \times p}$	$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^{\top} \\ \vdots \\ \mathbf{x}_n^{\top} \end{bmatrix}$
True (unknown) parameters	$\beta_{0*}, \boldsymbol{\beta}_*$
Exact MLE of the parameter	$\hat{\beta}_0, \hat{\boldsymbol{\beta}}$
Exact penalized estimate of the parameter	$\hat{\beta}_0^{\lambda}, \hat{\boldsymbol{\beta}}^{\lambda}$
Estimate obtained via numerical approximation	$\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}, \tilde{\beta}_0^{\lambda}, \tilde{\boldsymbol{\beta}}^{\lambda}$
Log-likelihood	$\ell_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \log(\cdot)$
Penalized log-likelihood	$l_n^{\lambda}(\boldsymbol{\beta})$
Mean of the j th column in covariate matrix	$u_{n,j}$
Standard deviation of the j th column in covariate matrix	$s_{n,j}$
Fisher information matrix	$\mathcal{I}(\beta_0, \boldsymbol{\beta})$

Table S3. Glossary for quantities specific to the vertical setting

Number of covariate-nodes	K
Number of covariates at covariate-node k	$p^{(k)}$
Covariate matrix at covariate-node k	$\mathbf{X}^{(k)}$
Centered and scaled covariate matrix at covariate-node k	$\mathbf{X}_{\text{cs}}^{(k)}$
Mean and s.d. of the j th column in $\mathbf{X}^{(k)}$	$u_{n,j}^{(k)}, s_{n,j}^{(k)}$
Gram matrix at covariate-node k	$\mathbf{K}^{(k)} := \mathbf{X}_{\text{cs}}^{(k)} (\mathbf{X}_{\text{cs}}^{(k)})^\top$
Dual parameter estimates (numerical approx.)	$\tilde{\alpha}^\lambda$
Penalized estimate associated with covariate-node k (numerical approx.)	$\tilde{\beta}_j^{\lambda(k)}$
Standard errors associated with covariate-node k (numerical approx.)	$\tilde{\sigma}_j^{\lambda(k)}$
Matrix in null-space of $\mathbf{K}^{(k)}$	$\mathcal{N}^{(k)}$
intermediate quantities	$\tilde{\mathbf{c}}^{\lambda(k)}, \tilde{\mathbf{S}}$

7 In the followings, we provide the detailed estimates and standard errors obtained for the analysis
8 using real health data pertaining first to CKD, and second to the MIMIC-IV database.

Table S4. Logistic regression model parameter estimates obtained with kidney failure at two years after baseline as outcome

Covariate-node	Variable	Vertical Partition $\tilde{\beta}^\lambda(\tilde{\sigma}^\lambda)$	Centralized $\hat{\beta}(\hat{\sigma})$
1	Age	-0.2069(0.1609)	-0.2069(0.1609)
1	Sex (male)	-0.0509(0.0594)	-0.0509(0.0594)
1	eGFR	-0.7731(0.5337)	-0.7732(0.5337)
1	uACR	-0.0076(0.0270)	-0.0076(0.0270)
2	Serum albumin	-0.6463(0.4388)	-0.6464(0.4389)
2	Serum bicarbonate	-0.0997(0.0806)	-0.0997(0.0806)
2	Serum calcium	0.3463(0.5243)	0.3463(0.5243)
2	Serum phosphate	0.5202(0.2887)	0.5202(0.2887)

A log-transformation was applied to uACR due to a skewed distribution¹.

The estimates for the centralized setting were computed using the *glm* function in R with a non-penalized logistic regression model.

Table S5. Logistic regression model parameter estimates obtained with MIMIC-IV database and *Death* as outcome

Covariate-node	Variable	Vertical Partition $\tilde{\beta}^{\lambda}(\tilde{\sigma}^{\lambda})$	Centralized $\hat{\beta}(\hat{\sigma})^1$
1	Age	0.0272(0.0023)	0.0272(0.0023)
1	Myocardial infarction	0.4052(0.0828)	0.4052(0.0828)
1	Renal disease	0.1867(0.0751)	0.1867(0.0751)
1	Malignant cancer	0.8814(0.0716)	0.8815(0.0716)
2	Triage temperature	-0.0951(0.0276)	-0.0951(0.0276)
2	Triage sbp	-0.0033(0.0015)	-0.0033(0.0015)
2	Triage dbp	-0.0097(0.0025)	-0.0097(0.0025)
2	Triage respiratory rate > 20	0.4752(0.0856)	0.4752(0.0856)
3	First-day urine output	-0.0005(0.0000)	-0.0005(0.0000)
3	First-day heart rate mean	0.0319(0.0020)	0.0319(0.0020)
3	First-day glucose mean	0.0025(0.0006)	0.0025(0.0006)

The estimates for the centralized setting were computed using the *glm* function in R with a non-penalized logistic regression model.

Supplementary Notes

The code for the implementation of the algorithm using R is available at: https://github.com/OpenLHS/Distrib_analysis/tree/main/Vertically_distributed_analysis/logistic_regression_nonpenalized. It includes an automated example with simulated data. The folder also includes a basic implementation of the tool that supports the privacy assessment for the response-node individual-level data.

15 Supplementary Methods - Details for the derivation of the dual optimization problem

16 Recall from the manuscript that to estimate $(\beta_{0\star}, \beta_\star)$, we consider the solutions $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$ of the
 17 ridge-penalized logistic regression problem

$$\max_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p} \left(l_n^\lambda(\beta_0, \beta) = \ell_n(\beta_0, \beta) - \frac{\lambda}{2} \left[\left(\beta_0 + \sum_{j=1}^p \beta_j \mu_{n,j} \right)^2 + \sum_{j=1}^p \beta_j^2 s_{n,j}^2 \right] \right),$$

where

$$\ell_n(\beta_0, \beta) = n^{-1} \sum_{i=1}^n \log \left[\frac{1}{1 + \exp\{-y_i(\beta_0 + \mathbf{x}_i^\top \beta)\}} \right].$$

18 Since, for all $i \in \{1, \dots, n\}$ we have

$$\begin{aligned} \beta_0 + \mathbf{x}_i^\top \beta &= \beta_0 + \sum_{j=1}^p x_{ij} \beta_j \\ &= \left(\beta_0 + \sum_{j=1}^p \mu_{n,j} \beta_j \right) + \sum_{j=1}^p \left(\frac{x_{ij} - \mu_{n,j}}{s_{n,j}} \right) (\beta_j s_{n,j}) \\ &= (\beta_0 + \hat{\boldsymbol{\mu}}^\top \beta) + \mathbf{x}_{i,\text{cs}}^\top \hat{\boldsymbol{\Sigma}} \beta, \end{aligned}$$

19 where we have introduced

$$\hat{\boldsymbol{\mu}} = [\mu_{n,1}, \dots, \mu_{n,p}]^\top, \quad \hat{\boldsymbol{\Sigma}} = \text{diag}([s_{n,1}, \dots, s_{n,p}]^\top), \quad (\text{S1})$$

20 it follows upon adopting the re-parametrization $(\beta_0^\circ, \beta^\circ) \equiv (\beta_0 + \hat{\boldsymbol{\mu}}^\top \beta, \hat{\boldsymbol{\Sigma}} \beta)$ that

$$\begin{aligned} &\max_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p} l_n^\lambda(\beta_0, \beta) \\ &= \max_{\beta_0^\circ \in \mathbb{R}, \beta^\circ \in \mathbb{R}^p} \left(\check{l}_n^\lambda(\beta_0^\circ, \beta^\circ) := \check{\ell}_n(\beta_0^\circ, \beta^\circ) - \frac{\lambda}{2} \sum_{j=0}^p (\beta_j^\circ)^2 \right), \\ &\quad \text{with } \check{\ell}_n(\beta_0^\circ, \beta^\circ) := -n^{-1} \sum_{i=1}^n \log [1 + \exp\{-y_i(\beta_0^\circ + \mathbf{x}_{i,\text{cs}}^\top \beta^\circ)\}]. \end{aligned}$$

21 Following the derivations in² (see Appendix A therein), since the concave dual conjugate of
 22 $x \mapsto (1 + e^{-x})^{-1}$ is $y \mapsto y \log y + (1 - y) \log(1 - y)$ for $y \in (0, 1)$, we obtain that the Fenchel dual
 23 of $\check{l}_n^\lambda(\beta_0^\circ, \beta^\circ)$ is the following minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\alpha} \in (0,1)^n} \left(\frac{1}{n} \sum_{i=1}^n \left\{ (1 - \alpha_i) \log(1 - \alpha_i) + \alpha_i \log(\alpha_i) \right\} \right. \\ &\quad \left. + \frac{1}{2\lambda n^2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_{i,\text{cs}}^\top \mathbf{x}_{j,\text{cs}} + 1) \right) = \min_{\boldsymbol{\alpha} \in (0,1)^n} J^\lambda(\boldsymbol{\alpha}). \end{aligned}$$

24 Its unique maximizer $\hat{\boldsymbol{\alpha}}^\lambda = (\hat{\alpha}_1^\lambda, \dots, \hat{\alpha}_n^\lambda)^\top \in (0, 1)^n$ then satisfies

$$\begin{bmatrix} \check{\beta}_0^\lambda \\ \check{\boldsymbol{\beta}}^\lambda \end{bmatrix} = (n\lambda)^{-1} \sum_{i=1}^n y_i \hat{\alpha}_i^\lambda \begin{bmatrix} 1 \\ \mathbf{x}_{i,\text{cs}} \end{bmatrix}.$$

Since the bijective nature of the reparametrization implies

$$\begin{bmatrix} \hat{\beta}_0^\lambda \\ \hat{\beta}^\lambda \end{bmatrix} = \begin{bmatrix} \check{\beta}_0^\lambda - \hat{\mu}^\top \hat{\beta}^\lambda \\ \hat{\Sigma}^{-1} \check{\beta}^\lambda \end{bmatrix}, \quad (S2)$$

we derive that

$$\hat{\beta}_0^\lambda = (n\lambda)^{-1} \sum_{i=1}^n y_i \hat{\alpha}_i^\lambda - \hat{\mu}^\top \hat{\beta}^\lambda, \quad \hat{\beta}^\lambda = \hat{\Sigma}^{-1} \left((n\lambda)^{-1} \sum_{i=1}^n y_i \hat{\alpha}_i^\lambda \mathbf{x}_{i,cs} \right).$$

By additionally noting that $\check{\ell}_n^\lambda(\beta_0, \beta)$ is strongly concave, its maximum is unique and is achieved at the point $(\check{\beta}_0^\lambda, \check{\beta}^\lambda)$ that satisfies

$$n^{-1} \sum_{i=1}^n \frac{y_i}{1 + \exp\{y_i(\check{\beta}_0^\lambda + \mathbf{x}_i^\top \check{\beta}^\lambda)\}} \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix} = \lambda \begin{bmatrix} \check{\beta}_0^\lambda \\ \check{\beta}^\lambda \end{bmatrix},$$

we can use the triangle inequality and derive

$$\max_{0 \leq j \leq p} |\check{\beta}_j^\lambda| \leq \lambda^{-1} \left(n^{-1} \sum_{i=1}^n (\|\mathbf{x}_i\|_\infty + 1) \right). \quad (S3)$$

Therefore, the primal maximization problem can be restricted over a compact search space, and the relation between the dual and primal parameters implies that the dual maximization problem can also be restricted over a compact search space.

Supplementary Methods - Equivalence to the standard non-penalized log-likelihood and computation of standard errors

Equivalence between parameter estimates

In this section, we formally establish, under appropriate conditions, the equivalence between the penalized estimators $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$, which maximize $\ell_n^\lambda(\beta_0, \beta)$, and the unpenalized maximum likelihood estimators. To ensure this equivalence, the conditions must guarantee that maximum likelihood estimates exist and are unique. It is well known³ that if the columns of \mathbf{X} are linearly independent and also linearly independent of the vector $\mathbf{1}_n$, then the Hessian $\nabla_{\beta_0, \beta}^2 \ell_n(\beta_0, \beta)$ is strictly negative definite, which implies that the log-likelihood function $\ell_n(\beta_0, \beta)$ is strictly concave. In this case, if a maximizer exists for the problem $\max_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p} \ell_n(\beta_0, \beta)$, then it must be unique and coincide with a stationary point of $\nabla_{\beta_0, \beta} \ell_n(\beta_0, \beta)$. The existence of such a solution is guaranteed when the response vector \mathbf{y} is not separable⁴. Specifically, \mathbf{y} is said to be separable if there exists (β_0, β) such that $y_i(\beta_0 + \mathbf{x}_i^\top \beta) > 0$ for all $i \in 1, \dots, n$. We will assume these conditions throughout the following discussion. Let $(\hat{\beta}_0, \hat{\beta})$ denote the unique maximizer of $\ell_n(\beta_0, \beta)$.

In what follows, for any positive definite matrix \mathbf{A} , let $\iota_{\min}(\mathbf{A})$ denote its smallest eigen value. Also, let $M \geq 1$ be a constant such that $\max_{1 \leq i \leq n} \|\mathbf{x}_{i,cs}\|_\infty \leq M$.

To establish the statistical equivalence between $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$ and $(\hat{\beta}_0, \hat{\beta})$, we start by proving an inequality that involves the reparametrized version of $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$ introduced in Supplementary Methods - *Details for the derivation of the dual optimization problem*, namely $(\check{\beta}_0^\lambda, \check{\beta}^\lambda) = \arg \max_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p} \check{\ell}_n^\lambda(\beta_0, \beta)$, and a corresponding reparametrized version of the maximum likelihood estimates $(\hat{\beta}_0, \hat{\beta})$. Specifically, by reparametrisation arguments analogous to those in the previous Supplementary Methods, it follows that $\max_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p} \ell_n(\beta_0, \beta) = \max_{\beta_0^\circ \in \mathbb{R}, \beta^\circ \in \mathbb{R}^p} \check{\ell}_n(\beta_0^\circ, \beta^\circ)$, with

$\check{\ell}_n$ defined in the previous Supplementary Methods. Moreover, $\check{\ell}_n$ has a unique maximizer $(\check{\beta}_0, \check{\beta})$ satisfying

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \check{\beta}_0 - \hat{\mu}^\top \hat{\beta} \\ \hat{\Sigma}^{-1} \check{\beta} \end{bmatrix}.$$

To derive the inequality, we adopt some notational simplifications. We will occasionally write $\mathbf{z}_i = [1, \mathbf{x}_i^\top]^\top$ and $\mathbf{z}_{i,\text{cs}} = [1, \mathbf{x}_{i,\text{cs}}^\top]^\top$. We will also use the shorthand $\hat{\theta}^\lambda$, $\hat{\theta}$, $\check{\theta}^\lambda$, and $\check{\theta}$ to denote, respectively, $[\hat{\beta}_0^\lambda, (\hat{\beta}^\lambda)^\top]^\top$, $[\hat{\beta}_0, \hat{\beta}^\top]^\top$, $[\check{\beta}_0^\lambda, (\check{\beta}^\lambda)^\top]^\top$, and $[\check{\beta}_0, \check{\beta}^\top]^\top$. In this notation, the inequality we shall prove below is the following:

$$\|\check{\theta}^\lambda - \check{\theta}\|_\infty \leq \frac{(p+1)\lambda\|\check{\theta}\|_\infty}{\lambda + \omega\{(p+2)^2\|\check{\theta}\|_\infty\}\iota_{\min}(n^{-1}\sum_{i=1}^n \mathbf{z}_{i,\text{cs}}\mathbf{z}_{i,\text{cs}}^\top)}, \quad (\text{S4})$$

where $\omega(x) := e^x/(1+e^x)^2$.

Since $\check{\theta}^\lambda$ and $\check{\theta}$ are the unique stationary points of $\check{\ell}_n^\lambda$ and $\check{\ell}_n$, respectively, they satisfy the corresponding first-order conditions, that is, they cancel the gradients of these functions. Hence, we have

$$\begin{aligned} \lambda\check{\theta}^\lambda &= n^{-1} \sum_{i=1}^n y_i \left[\frac{1}{1 + \exp(y_i \mathbf{z}_{i,\text{cs}}^\top \check{\theta}^\lambda)} - \frac{1}{1 + \exp(y_i \mathbf{z}_{i,\text{cs}}^\top \check{\theta})} \right] \mathbf{z}_{i,\text{cs}} \\ &= n^{-1} \sum_{i=1}^n y_i \left[\frac{1}{1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + \Delta)\}} - \frac{1}{1 + \exp(y_i \mathbf{z}_{i,\text{cs}}^\top \check{\theta})} \right] \mathbf{z}_{i,\text{cs}}, \end{aligned}$$

where we have introduced $\Delta = \check{\theta}^\lambda - \check{\theta}$. Since

$$\frac{1}{1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + \Delta)\}} - \frac{1}{1 + \exp(y_i \mathbf{z}_{i,\text{cs}}^\top \check{\theta})} = - \int_0^1 \frac{\exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}}{[1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}]^2} dt y_i \mathbf{z}_{i,\text{cs}}^\top \Delta,$$

we further deduce that

$$\begin{aligned} \lambda\check{\theta}^\lambda &= - \int_0^1 \left(n^{-1} \sum_{i=1}^n \frac{\exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}}{[1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}]^2} \mathbf{z}_{i,\text{cs}} \mathbf{z}_{i,\text{cs}}^\top \right) dt \Delta \\ &= - \int_0^1 \left(n^{-1} \sum_{i=1}^n \frac{\exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}}{[1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}]^2} \mathbf{z}_{i,\text{cs}} \mathbf{z}_{i,\text{cs}}^\top \right) dt (\check{\theta}^\lambda - \check{\theta}). \end{aligned}$$

From straightforward algebra manipulations, the latter equation implies

$$\begin{aligned} \check{\theta}^\lambda - \check{\theta} &= -\lambda \left\{ \int_0^1 \left(n^{-1} \sum_{i=1}^n \frac{\exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}}{[1 + \exp\{y_i \mathbf{z}_{i,\text{cs}}^\top (\check{\theta} + t\Delta)\}]^2} \mathbf{z}_{i,\text{cs}} \mathbf{z}_{i,\text{cs}}^\top \right) dt + \lambda \mathbf{I}_{p+1} \right\}^{-1} \check{\theta} \\ &= -\lambda \left\{ \int_0^1 \mathcal{I}(\check{\theta} + t\Delta) dt + \lambda \mathbf{I}_{p+1} \right\}^{-1} \check{\theta}, \end{aligned}$$

where $\mathcal{I}(\check{\theta} + t\Delta)$ is the Fisher information matrix at (3) computed at parameter $\check{\theta} + t\Delta$.

This provides the bound

$$\|\check{\theta}^\lambda - \check{\theta}\|_\infty \leq \frac{(p+1)\lambda\|\check{\theta}\|_\infty}{\lambda + \iota_{\min}\{\int_0^1 \mathcal{I}(\check{\theta} + t\Delta) dt\}}.$$

Since the latter inequality trivially implies $\|\Delta\|_\infty = \|\check{\theta}^\lambda - \check{\theta}\|_\infty \leq (p+1)\|\check{\theta}\|_\infty$, which further implies $\max_{1 \leq i \leq n} |y_i \mathbf{z}_{i,\text{cs}}^\top (\hat{\theta} + t\Delta)| \leq M(p+1)\{\|\check{\theta}\|_\infty + (p+1)\|\check{\theta}\|_\infty\} \leq M(p+2)^2\|\check{\theta}\|_\infty$, we deduce the bound in (S4) from the fact that $\omega(x) = e^x/(1+e^x)^2$ is even, and strictly decreasing for all $x > 0$.

From this bound, we will derive a bound for $\|\hat{\theta}^\lambda - \hat{\theta}\|_\infty$ using the relationship

$$\begin{aligned} \begin{bmatrix} \hat{\beta}_0^\lambda - \hat{\beta}_0 \\ \hat{\beta}^\lambda - \hat{\beta} \end{bmatrix} &= \begin{bmatrix} \check{\beta}_0^\lambda - \check{\beta}_0 - \hat{\mu}^\top (\hat{\beta}^\lambda - \hat{\beta}) \\ \hat{\Sigma}^{-1}(\check{\beta}^\lambda - \check{\beta}) \end{bmatrix} \\ &= \begin{bmatrix} \check{\beta}_0^\lambda - \check{\beta}_0 - \hat{\mu}^\top \hat{\Sigma}^{-1}(\check{\beta}^\lambda - \check{\beta}) \\ \hat{\Sigma}^{-1}(\check{\beta}^\lambda - \check{\beta}) \end{bmatrix}. \end{aligned}$$

From the above relationship, it can readily be deduced that

$$\|\hat{\theta}^\lambda - \hat{\theta}\|_\infty \leq \left\{1 + \frac{p\|\hat{\mu}\|_\infty}{\iota_{\min}(\hat{\Sigma})}\right\} \|\check{\theta}^\lambda - \check{\theta}\|_\infty.$$

Using (S4), this directly implies that

$$\begin{aligned} \|\hat{\theta}^\lambda - \hat{\theta}\|_\infty &\leq \frac{\left\{1 + \frac{p\|\hat{\mu}\|_\infty}{\iota_{\min}(\hat{\Sigma})}\right\} (p+1)\lambda \|\check{\theta}\|_\infty}{\lambda + \omega\{(p+2)^2\|\check{\theta}\|_\infty\} \iota_{\min}(n^{-1} \sum_{i=1}^n \mathbf{z}_{i,\text{cs}} \mathbf{z}_{i,\text{cs}}^\top)} \\ &\leq \frac{\left\{1 + \frac{p\|\hat{\mu}\|_\infty}{\iota_{\min}(\hat{\Sigma})}\right\}^2 (p+1)\lambda \|\hat{\theta}\|_\infty}{\lambda + \omega\left[(p+2)^2\left\{1 + \frac{p\|\hat{\mu}\|_\infty}{\iota_{\min}(\hat{\Sigma})}\right\} \|\hat{\theta}\|_\infty\right] \iota_{\min}(n^{-1} \sum_{i=1}^n \mathbf{z}_{i,\text{cs}} \mathbf{z}_{i,\text{cs}}^\top)}. \end{aligned}$$

Under the assumption that the \mathbf{x}_i 's have finite marginal means and variances, and that their variance-covariance matrix is invertible, together with the condition $\hat{\theta}$ have finite components, the above expression ensures that taking λ sufficiently small yields the equivalence—up to a negligible remainder—between $\hat{\theta}^\lambda$ and $\hat{\theta}$.

Standard errors

The standard deviation of $\hat{\beta}_j$, given by $\sqrt{\text{Var}(\hat{\beta}_j)} = n^{-1/2}[(\mathbb{E}\{\mathcal{I}(\beta_{0\star}, \beta_\star)\})^{-1}]_{jj}$ is commonly estimated using the following quantity:

$$\hat{\sigma}_j := \frac{1}{\sqrt{n}} \left(\left[(\mathcal{I}(\hat{\beta}_0, \hat{\beta}))^{-1} \right]_{jj} \right)^{1/2}.$$

Based on the previous subsection, which implies that $\max(|\hat{\beta}_0^\lambda - \hat{\beta}_0|, \|\hat{\beta}^\lambda - \hat{\beta}\|_\infty)$ can be made arbitrarily small by choosing λ small, the proof of the consistency of our standard error computation procedure follows from the derivations provided in the following section, which show that $[(\mathcal{I}(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda))^{-1}]_{jj} = [(\mathcal{I}^\lambda)^{-1}]_{jj}/s_{n,j}^2$, with \mathcal{I}^λ defined at the beginning of Methods - *Methodology for computing standard errors of parameter estimates in VALORIS* (in main manuscript).

Computation of standard errors

Recall from above the definition of $(\check{\beta}_0^\lambda, \check{\beta}^\lambda)$.

For $j \in \{1, \dots, p\}$, one obtains from the relationship between $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$ and $(\check{\beta}_0^\lambda, \check{\beta}^\lambda)$ that

$$[\{\mathcal{I}(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)\}^{-1}]_{jj} = [\{-\nabla_{\beta_0, \beta}^2 \check{\ell}_n(\check{\beta}_0^\lambda, \check{\beta}^\lambda)\}^{-1}]_{jj}/s_{n,j}^2,$$

92 where

$$\begin{aligned}
-\nabla_{\beta_0, \beta}^2 \check{\ell}_n(\check{\beta}_0^\lambda, \check{\beta}^\lambda) &= \frac{1}{n} \sum_{i=1}^n \frac{\exp\{y_i(\check{\beta}_0^\lambda + \mathbf{x}_{i,cs}^\top \check{\beta}^\lambda)\}}{[1 + \exp\{y_i(\check{\beta}_0^\lambda + \mathbf{x}_{i,cs}^\top \check{\beta}^\lambda)\}]^2} \begin{bmatrix} 1 & \mathbf{x}_{i,cs}^\top \\ \mathbf{x}_{i,cs} & \mathbf{x}_{i,cs} \mathbf{x}_{i,cs}^\top \end{bmatrix} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}}{[1 + \exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}]^2} \begin{bmatrix} 1 & \mathbf{x}_{i,cs}^\top \\ \mathbf{x}_{i,cs} & \mathbf{x}_{i,cs} \mathbf{x}_{i,cs}^\top \end{bmatrix}.
\end{aligned}$$

93 Now, recall that, for each $k \in \{1, \dots, K\}$, the vector $\hat{\mathbf{c}}^{\lambda(k)}$ defined in (7) satisfies
94 $\hat{\mathbf{c}}^{\lambda(k)} = \mathbf{X}_{cs}^{(k)} \text{diag}(s_{n,1}^{(k)}, \dots, s_{n,p(k)}^{(k)}) \hat{\beta}^{\lambda(k)} = \mathbf{X}^{(k)} \hat{\beta}^{\lambda(k)} - (\sum_{j=1}^{p(k)} \hat{\beta}_j^{\lambda(k)} \mu_j^{(k)}) \mathbf{1}_n$, and that the response-
95 node has access to $\hat{\mathbf{c}}^{\lambda(1)}, \dots, \hat{\mathbf{c}}^{\lambda(K)}$. Since the response-node can also compute $(n\lambda)^{-1} \sum_{i=1}^n \hat{\alpha}_i^\lambda y_i =$
96 $\hat{\beta}_0^\lambda + \sum_{j=1}^n \hat{\beta}_j^\lambda \mu_{n,j}$ (recall the expression given in (6)), it is therefore able to compute

$$\hat{\beta}_0^\lambda \mathbf{1}_n + \mathbf{X} \hat{\beta}^\lambda = \hat{\beta}_0^\lambda \mathbf{1}_n + \sum_{k=1}^K \mathbf{X}^{(k)} \hat{\beta}^{\lambda(k)} = ((n\lambda)^{-1} \sum_{i=1}^n \hat{\alpha}_i^\lambda y_i) \mathbf{1}_n + \sum_{k=1}^K \hat{\mathbf{c}}^{\lambda(k)}.$$

97 Then, upon defining $\hat{\mathbf{V}}^\lambda$, whose diagonal entries $[\hat{\mathbf{V}}^\lambda]_{jj}$ satisfying

$$\begin{aligned}
[\hat{\mathbf{V}}^\lambda]_{jj} &= \frac{\exp\left[y_j \left\{ (n\lambda)^{-1} \sum_{i=1}^n \hat{\alpha}_i^\lambda y_i + \sum_{k=1}^K \hat{c}_j^{\lambda(k)} \right\}\right]}{\left(1 + \exp\left[y_j \left\{ (n\lambda)^{-1} \sum_{i=1}^n \hat{\alpha}_i^\lambda y_i + \sum_{k=1}^K \hat{c}_j^{\lambda(k)} \right\}\right]\right)^2} \\
&= \frac{\exp\{y_j(\hat{\beta}_0^\lambda + \mathbf{x}_j^\top \hat{\beta}^\lambda)\}}{[1 + \exp\{y_j(\hat{\beta}_0^\lambda + \mathbf{x}_j^\top \hat{\beta}^\lambda)\}]^2},
\end{aligned}$$

98 the matrix $-\nabla_{\beta_0, \beta}^2 \check{\ell}_n(\check{\beta}_0^\lambda, \check{\beta}^\lambda)$ can be computed as

$$\begin{aligned}
\mathcal{I}^\lambda &= n^{-1} \begin{bmatrix} \mathbf{1}_n & \mathbf{X}_{cs}^{(1)} & \dots & \mathbf{X}_{cs}^{(K)} \end{bmatrix}^\top \hat{\mathbf{V}}^\lambda \begin{bmatrix} \mathbf{1}_n & \mathbf{X}_{cs}^{(1)} & \dots & \mathbf{X}_{cs}^{(K)} \end{bmatrix} \\
&= n^{-1} \begin{bmatrix} \mathbf{1}_n^\top \hat{\mathbf{V}}^\lambda \mathbf{1}_n & \mathbf{1}_n^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(1)} & \dots & \mathbf{1}_n^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(K)} \\ (\mathbf{X}_{cs}^{(1)})^\top \hat{\mathbf{V}}^\lambda \mathbf{1}_n & (\mathbf{X}_{cs}^{(1)})^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(1)} & \dots & (\mathbf{X}_{cs}^{(1)})^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{X}_{cs}^{(K)})^\top \hat{\mathbf{V}}^\lambda \mathbf{1}_n & (\mathbf{X}_{cs}^{(K)})^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(1)} & \dots & (\mathbf{X}_{cs}^{(K)})^\top \hat{\mathbf{V}}^\lambda \mathbf{X}_{cs}^{(K)} \end{bmatrix}.
\end{aligned}$$

99 **Supplementary Methods - Box-constrained optimization algorithm and stopping criteria**

100 **Two-metric projected Newton algorithm**

101 The convexity of the dual problem to solve at the response-node ensures that a unique solution
102 exists on the domain of the objective function. The algorithm used to solve the problem should allow
103 sufficient descent to reach an adequate approximation of this unique solution. Since the components
104 of α are bounded by a compact set included in the open set $(0, 1)$ (see previous Supplementary
105 Methods), an algorithm adequate for box-constrained convex optimization problems had to be
106 selected. While many methods exist for box-constrained optimization⁵, the chosen method should
107 allow to reach convergence with sufficient precision given the potentially small magnitude of the
108 dual parameter α while still offering efficient computation when the dimension of the dual is high.
109 We used the Two-Metric Projected Newton method suggested by Bertsekas⁶, applicable because
110 we previously argued that the dual parameter estimates lie in a compact parameter space. We refer

to⁷ for an extensive description of the method and convergence details. Briefly, all components of the estimate $\hat{\alpha}_{(t)}^\lambda$ at step t at a boundary of the search domain and for which the gradient would pull the search toward the opposite side of the search domain are updated through gradient descent projected in the domain, while all other components are updated using Newton descent projected in the domain. The update is therefore $\hat{\alpha}_{(t+1)}^\lambda = \text{Proj}[\hat{\alpha}_{(t)}^\lambda - \theta \mathbf{D}_{(t)} \nabla_{\alpha} J^\lambda(\hat{\alpha}_{(t)}^\lambda)]$, where $\mathbf{D}_{(t)}$ depends of the component as described before and $\text{Proj}[\cdot]$ denotes the projection under the Euclidean norm. The step size θ is selected through backtracking line search (Armijo rule) along projection arc detailed in^{5,7}. An initial admissible estimate has to be provided, which was set at $\hat{\alpha}_{(0)}^\lambda = [0.1, \dots, 0.1]^\top$.

Stopping criteria

The error entailed by the approximation of $\hat{\alpha}^\lambda$ in the chosen algorithm should ideally be low enough such that it preserve the asymptotic properties derived for the primal estimate. We notice that λ holds a scaling role over the dual parameter α when it comes to retrieving the associated primal parameter β . A restriction in function of λ consequently needs to be imposed in the estimation of the dual parameter to preserve the asymptotic properties of the primal parameters. We derived a stopping criteria for the dual estimation that ensures the asymptotic properties of the primal parameter hold.

We first fix $\epsilon > 0$ and consider $\tilde{\alpha}^\lambda := \tilde{\alpha}_\epsilon^\lambda \in (0, 1)^n$ such that

$$\|\nabla_{\alpha} J^\lambda(\tilde{\alpha}^\lambda)\|_2 \leq \frac{2\lambda}{\sqrt{p+1}} \left(p(n-1) + n \right)^{-1/2} \epsilon.$$

For $i \in \{1, \dots, n\}$ let $x_{0i} = 1$ and $\check{\beta}_j^\lambda = (\lambda n)^{-1} \sum_{i=1}^n y_i \tilde{\alpha}_i^\lambda x_{ij, \text{cs}}$ and $\check{\beta}_j^\lambda = (\lambda n)^{-1} \sum_{i=1}^n y_i \hat{\alpha}_i^\lambda x_{ij, \text{cs}}$ for $j \in \{0, \dots, p\}$.

Using the fact that $y_i \in \{-1, 1\}$ for all $i \in \{1, \dots, n\}$, we derive that

$$\begin{aligned} \max_{j \in \{0, \dots, p\}} |\check{\beta}_j^\lambda - \check{\beta}_j^\lambda| &= \max_{j \in \{0, \dots, p\}} \left| (\lambda n)^{-1} \sum_{i=1}^n x_{ij, \text{cs}} y_i (\tilde{\alpha}_i^\lambda - \hat{\alpha}_i^\lambda) \right| \\ &\leq (\lambda n)^{-1} \sum_{j=0}^p \sum_{i=1}^n |x_{ij, \text{cs}}| |\tilde{\alpha}_i^\lambda - \hat{\alpha}_i^\lambda| \\ &\leq (\lambda n)^{-1} \|\tilde{\alpha}^\lambda - \hat{\alpha}^\lambda\|_2 \sum_{j=0}^p \left(\sum_{i=1}^n x_{ij, \text{cs}}^2 \right)^{1/2} \\ &\leq \sqrt{p+1} (\lambda n)^{-1} \|\tilde{\alpha}^\lambda - \hat{\alpha}^\lambda\|_2 \left(\sum_{j=0}^p \sum_{i=1}^n x_{ij, \text{cs}}^2 \right)^{1/2} \\ &= \sqrt{p+1} (\lambda n)^{-1} \|\tilde{\alpha}^\lambda - \hat{\alpha}^\lambda\|_2 \left(p(n-1) + n \right)^{1/2}. \end{aligned} \tag{S5}$$

To obtain the second-to-last line, we used Cauchy-Schwartz inequality, and to obtain the one-to-last line, we used the fact that for any positive a_0, \dots, a_p we have $\sum_{j=0}^p a_j \leq \sqrt{p+1} (\sum_{j=0}^p a_j^2)^{1/2}$. Finally, the last line is derived from the fact that $\sum_{i=1}^n x_{ij, \text{cs}}^2 = n-1$ for all $j \in \{1, \dots, p\}$.

Now observe that, using standard vector calculus manipulations, the Hessian matrix of $J^\lambda(\alpha)$

136 can be expressed as

$$\begin{aligned} \nabla_{\alpha}^2 J^{\lambda}(\alpha) = & (\lambda n^2)^{-1} \text{diag}(\mathbf{y}) \left(\sum_{k=1}^K \mathbf{\kappa}^{(k)} + \mathbf{1}_n \mathbf{1}_n^{\top} \right) \text{diag}(\mathbf{y}) \\ & + n^{-1} \text{diag} \left\{ [(\alpha_1(1 - \alpha_1))^{-1}, \dots, (\alpha_n(1 - \alpha_n))^{-1}]^{\top} \right\}. \end{aligned}$$

137 In the equation above, the matrix in the first term of the right-hand side of the equality is semi-
 138 positive definite, since for any vector $\alpha \in \mathbb{R}^n$, $\alpha^{\top} \text{diag}(\mathbf{y}) \left(\sum_{k=1}^K \mathbf{\kappa}^{(k)} + \mathbf{1}_n \mathbf{1}_n^{\top} \right) \text{diag}(\mathbf{y}) \alpha \geq 0$. As
 139 the matrix $n^{-1} \text{diag}\{[(\alpha_1(1 - \alpha_1))^{-1}, \dots, (\alpha_n(1 - \alpha_n))^{-1}]^{\top}\}$ is positive definite for all $\alpha \in (0, 1)^n$,
 140 with $(\alpha_i(1 - \alpha_i))^{-1} \geq 4$ for all $i \in \{1, \dots, n\}$, it follows that $\nabla_{\alpha}^2 J^{\lambda}(\alpha)$ is strongly convex, with
 141 strong convexity parameter $m = 4n^{-1}$, since it follows from the last discussion that the matrix

$$\nabla_{\alpha}^2 J^{\lambda}(\alpha) - \frac{m}{2} \mathbf{I},$$

142 is positive definite.

143 This allows us to conclude as in e.g.⁸, Section 9.1.2, p.459, that it holds for all $\alpha \in (0, 1)^n$ that

$$\|\alpha - \hat{\alpha}^{\lambda}\|_2 \leq \frac{2}{m} \|\nabla_{\alpha} J^{\lambda}(\alpha)\|_2 = \frac{n}{2} \|\nabla_{\alpha} J^{\lambda}(\alpha)\|_2.$$

144 Combining this result with the inequality derived at (S5), we obtain

$$\max_{j \in \{0, \dots, p\}} |\tilde{\beta}_j^{\lambda} - \check{\beta}_j^{\lambda}| \leq \frac{\sqrt{p+1}}{2\lambda} \|\nabla_{\alpha} J^{\lambda}(\tilde{\alpha}^{\lambda})\|_2 \left(p(n-1) + n \right)^{1/2}.$$

145 Now combining this result with the fact that

$$\|\nabla_{\alpha} J^{\lambda}(\tilde{\alpha}^{\lambda})\|_2 \leq \frac{2\lambda}{\sqrt{p+1}} \left(p(n-1) + n \right)^{-1/2} \epsilon,$$

146 it directly follows that $\max_{j \in \{0, \dots, p\}} |\tilde{\beta}_j^{\lambda} - \check{\beta}_j^{\lambda}| < \epsilon$.

147 It remains that we wish to establish the bound over $\max_{j \in \{0, \dots, p\}} |\tilde{\beta}_j^{\lambda} - \hat{\beta}_j^{\lambda}|$, where $(\tilde{\beta}_0^{\lambda}, \tilde{\beta}^{\lambda})$ denotes
 148 a version of $(\hat{\beta}_0^{\lambda}, \hat{\beta}^{\lambda})$ computed based on $\tilde{\alpha}^{\lambda}$ instead of $\hat{\alpha}^{\lambda}$. As shown in previous sections, the
 149 maximizer $(\hat{\beta}_0^{\lambda}, \hat{\beta}^{\lambda})$ of $l_n^{\lambda}(\beta_0, \beta)$ satisfies

$$\begin{bmatrix} \hat{\beta}_0^{\lambda} \\ \hat{\beta}^{\lambda} \end{bmatrix} = \begin{bmatrix} \check{\beta}_0^{\lambda} - \hat{\mu}^{\top} \check{\beta}^{\lambda} \\ \hat{\Sigma}^{-1} \check{\beta}^{\lambda} \end{bmatrix},$$

150 where $\hat{\mu}$ and $\hat{\Sigma}$ are defined in (S1).

151 Therefore, $(\tilde{\beta}_0^{\lambda}, \tilde{\beta}^{\lambda})$ satisfies

$$\begin{bmatrix} \tilde{\beta}_0^{\lambda} \\ \tilde{\beta}^{\lambda} \end{bmatrix} = \begin{bmatrix} \check{\beta}_0^{\lambda} - \hat{\mu}^{\top} \tilde{\beta}^{\lambda} \\ \hat{\Sigma}^{-1} \tilde{\beta}^{\lambda} \end{bmatrix},$$

152 Therefore, if $\tilde{\alpha}^{\lambda}$ is a point such that

$$\|\nabla_{\alpha} J^{\lambda}(\tilde{\alpha}^{\lambda})\|_2 \leq \frac{2\lambda}{\sqrt{p+1}} \left(p(n-1) + n \right)^{-1/2} \epsilon,$$

153 it follows that $\max_{j \in \{0, \dots, p\}} |\tilde{\beta}_j^{\lambda} - \check{\beta}_j^{\lambda}| < \epsilon$. Then, for $j \in \{1, \dots, p\}$, we have

$$|\hat{\beta}_j^{\lambda} - \tilde{\beta}_j^{\lambda}| \leq s_{n,j}^{-1} |\check{\beta}_j^{\lambda} - \tilde{\beta}_j^{\lambda}| \leq s_{n,j}^{-1} \epsilon.$$

154 The inequality over $\|\nabla_{\alpha} J^{\lambda}(\tilde{\alpha}^{\lambda})\|_2$ can therefore be used as a stopping criteria, and the asymptotic
 155 properties are preserved under $\epsilon = o(n^{-1/2})$.

Supplementary Methods - Privacy-preserving properties

Theoretical details : When only parameter estimates and their standard errors are available to the response-node, without the intermediate quantities

Recall from Methods - *Methodology for computing standard errors of parameter estimates in VALORIS* (in main manuscript) that $\hat{\sigma}_j^\lambda = \sqrt{[(\mathcal{I}^\lambda + \eta \mathbf{I}_{p+1})^{-1}]_{j+1,j+1}/s_{n,j}}$, and where from Supplementary Methods - *Computation of standard errors*, \mathcal{I}^λ can be expressed as

$$\mathcal{I}^\lambda = \frac{1}{n} \sum_{i=1}^n \frac{\exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}}{[1 + \exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}]^2} \begin{bmatrix} 1 & \mathbf{x}_{i,\text{cs}}^\top \\ \mathbf{x}_{i,\text{cs}} & \mathbf{x}_{i,\text{cs}} \mathbf{x}_{i,\text{cs}}^\top \end{bmatrix}.$$

Also recall that $(\hat{\beta}_0^\lambda, \hat{\beta}^\lambda)$ solve the maximization problem defined in (5) in the manuscript, which therefore implies that $\hat{\beta}^\lambda$ satisfies

$$n^{-1} \sum_{i=1}^n \frac{y_i}{1 + \exp\{y_i(\hat{\beta}_0^\lambda + \hat{\mu}^\top \hat{\beta}^\lambda + \mathbf{x}_{i,\text{cs}}^\top \hat{\Sigma} \hat{\beta}^\lambda)\}} \mathbf{x}_{i,\text{cs}} = \lambda \hat{\Sigma} \hat{\beta}^\lambda,$$

where we have introduced

$$\hat{\mu} = [\mu_{n,1}, \dots, \mu_{n,p}]^\top, \quad \hat{\Sigma} = \text{diag}([s_{n,1}, \dots, s_{n,p}]^\top),$$

see Supplementary Methods - *Details for the derivation of the dual optimization problem* for details.

Suppose, without loss of generality, that the response-node also holds covariate data and is labeled as covariate-node $k = 1$, and consider the setting where only two nodes participate in the analysis: the response-node (also acting as a covariate-node) and a single additional covariate-node. The extension to scenarios involving more than two covariate-nodes follows analogously.

To analyze the privacy risk entailed when the set $\{(\hat{\beta}_j^{\lambda(2)}, \hat{\sigma}_j^{\lambda(2)}) : j \in \{1, \dots, p^{(2)}\}\}$ of parameter estimates and their associated standard errors—together with the response vector \mathbf{y} and the covariate data it holds—is the only information available to the response-node, we adopt a deliberately more adverse scenario: for the purposes of this analysis only, we assume that, in addition to the $\hat{\beta}_j^{\lambda(2)}$'s, the response-node has access to the full matrix \mathcal{J}^λ defined as

$$\mathcal{J}^\lambda := \hat{\Sigma}^{-2} \left(\frac{1}{n} \sum_{i=1}^n \frac{\exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}}{[1 + \exp\{y_i(\hat{\beta}_0^\lambda + \mathbf{x}_i^\top \hat{\beta}^\lambda)\}]^2} \mathbf{x}_{i,\text{cs}} \mathbf{x}_{i,\text{cs}}^\top + \eta \mathbf{I}_p \right),$$

rather than solely to the $\hat{\sigma}_j^{\lambda(2)}$'s, which satisfy $\hat{\sigma}_j^{\lambda(2)} = \sqrt{[(\mathcal{J}^\lambda)^{-1}]_{jj}}$. This conservative assumption simplifies the mathematical derivations and yields an upper bound on the potential privacy risk.

When at least one continuous covariate is held by a covariate-node located outside the response-node, then candidate matrices for $\mathbf{X}_{\text{cs}}^{(2)}$ are column-centered and scaled matrix $\mathbf{A} = [\mathbf{a}_1^\top, \dots, \mathbf{a}_n^\top]^\top$ satisfying

$$n^{-1} \sum_{i=1}^n \frac{y_i}{1 + \exp[y_i\{b + (\mathbf{x}_{i,\text{cs}}^{(1)})^\top \hat{\Sigma}^{(1)} \hat{\beta}^{\lambda(1)} + \mathbf{a}_i^\top \Delta \hat{\beta}^{\lambda(2)}\}]} \begin{bmatrix} \mathbf{x}_{i,\text{cs}}^{(1)} \\ \mathbf{a}_i \end{bmatrix} = \lambda \begin{bmatrix} \hat{\Sigma}^{(1)} \hat{\beta}^{\lambda(1)} \\ \Delta \hat{\beta}^{\lambda(2)} \end{bmatrix}, \quad (\text{S6})$$

and

$$\begin{aligned} n^{-1} \sum_{i=1}^n \frac{\exp\{b + (\mathbf{x}_{i,\text{cs}}^{(1)})^\top \hat{\Sigma}^{(1)} \hat{\beta}^{\lambda(1)} + \mathbf{a}_i^\top \Delta \hat{\beta}^{\lambda(2)}\}}{[1 + \exp\{b + (\mathbf{x}_{i,\text{cs}}^{(1)})^\top \hat{\Sigma}^{(1)} \hat{\beta}^{\lambda(1)} + \mathbf{a}_i^\top \Delta \hat{\beta}^{\lambda(2)}\}]^2} \begin{bmatrix} \mathbf{x}_{i,\text{cs}}^{(1)} (\mathbf{x}_{i,\text{cs}}^{(1)})^\top & \mathbf{x}_{i,\text{cs}}^{(1)} \mathbf{a}_i^\top \\ \mathbf{a}_i (\mathbf{x}_{i,\text{cs}}^{(1)})^\top & \mathbf{a}_i \mathbf{a}_i^\top \end{bmatrix} \\ = \begin{bmatrix} \hat{\Sigma}^{(1)} & \mathbf{0} \\ \mathbf{0} & \Delta \end{bmatrix}^2 \mathcal{J}^\lambda - \eta \mathbf{I}_p, \end{aligned} \quad (\text{S7})$$

with $\hat{\Sigma}^{(1)} := \text{diag}([s_{n,1}^{(1)}, \dots, s_{n,p(1)}^{(1)}])$, where $\Delta := \text{diag}([\delta_1, \dots, \delta_{p(2)}]^\top)$ is the vector of unknown standard deviations, and b denotes the unknown intercept.

Assume temporarily that the only unknowns in the system of equations above are those associated with the (or one of the, if multiple exist) continuous variables held by covariate-node $k = 2$, and that all other entries of \mathbf{A} are fixed and equal to those of $\mathbf{X}_{\text{cs}}^{(2)}$. In this case, the system comprises $n + 2$ real-valued unknowns: the n entries of the continuous variable in \mathbf{A} , one candidate intercept b , and one candidate standard deviation corresponding to the continuous variable in Δ . On the other hand, the system imposes $p + p(p + 1)/2$ equality constraints: p equations from the first-order optimality condition in (S6), and $p(p + 1)/2$ from the symmetry-reduced second-order condition in (S7) (since \mathcal{J}^λ is symmetric, the $p(p - 1)/2$ off-diagonal constraints are not independent). In addition, two further constraints are imposed to ensure that the column of \mathbf{A} associated with the continuous variable is centered and scaled (i.e., has mean zero and variance one). Thus, the total number of equations is $p + p(p + 1)/2 + 2$, while the number of unknowns remains $n + 2$.

When the inequality $p + p(p + 1)/2 < n$ holds, the number of unknowns exceeds the number of independent constraints. Since a solution lying in the interior of the feasible set exists (i.e., the configuration defined by $\mathbf{X}_{\text{cs}}^{(k)}$), the constraint set defines a smooth manifold of positive dimension in a neighborhood of that point. Therefore, the system admits infinitely many solutions when $p + p(p + 1)/2 < n$, provided that the component of $\hat{\beta}^{\lambda(2)}$ associated with the continuous variable is nonzero. This argument can be repeated for each continuous covariate held by covariate-node $k = 2$, showing that each associated column admits infinitely many admissible candidate configurations.

Now assume that, in addition to a continuous covariate, $\mathbf{X}_{\text{cs}}^{(2)}$ also includes a centered and scaled binary covariate. Consider the case where the entries of \mathbf{A} corresponding to this binary covariate match those of $\mathbf{X}_{\text{cs}}^{(2)}$, except for two entries—one originally positive and one originally negative in $\mathbf{X}_{\text{cs}}^{(2)}$ —whose signs are flipped to preserve the column's mean and variance (such a pair always exists under the assumption that \mathbf{X}_{cs} is not colinear with $\mathbf{1}_n$). As above, assume that all other entries of \mathbf{A} are equal to those of $\mathbf{X}_{\text{cs}}^{(2)}$, except for those in the column corresponding to the (or one of the, if multiple exist) continuous covariate. Now interpret the equations in (S6) and (S7) as defining a system in which the unknowns are the entries of \mathbf{A} associated with the continuous covariate, its associated standard error, and the unknown intercept. Because the constraints vary smoothly with respect to the continuous covariate entries of \mathbf{A} , and provided that the vector \mathbf{y} contains at least two entries equal to 1 and at least two equal to -1 , the system admits at least one solution whenever n is larger than $p + p(p + 1)/2$. This is because, as the constraints behave smoothly, small changes in the continuous values can compensate for small mismatches elsewhere, allowing the system to adjust without violating the structure required by the observed quantities.

It follows that, for any entry in a column corresponding to a binary covariate in $\mathbf{X}_{\text{cs}}^{(2)}$, there exists an admissible candidate matrix in which that entry takes a different value.

Privacy assessment for the response vector - Empirical criterion

We now provide details regarding the ability of covariate-node k to retrieve the response-vector \mathbf{y} from $\hat{\mathbf{c}}^{\lambda(k)}$. Recall that covariate-node k has access to $\mathcal{K}^{(k)} = \mathbf{X}_{\text{cs}}^{(k)}(\mathbf{X}_{\text{cs}}^{(k)})^\top$ and to η , and that

$$\mathbb{S}(\hat{\mathbf{c}}^{\lambda(k)}) = \{\mathbf{y}^\dagger \in \{-1, 1\}^n : \hat{\mathbf{c}}^{\lambda(k)} = (n\lambda)^{-1} \mathcal{K}^{(k)} \text{diag}(\boldsymbol{\alpha}) \mathbf{y}^\dagger \text{ for some } \boldsymbol{\alpha} \in (0, 1)^n\}.$$

Since we have assumed that at least one continuous covariate is held outside of covariate-node k , $\boldsymbol{\alpha}$ can be treated as a vector with real-valued entries. Moreover, since we have assumed that $\mathbf{X}^{(k)}$ has full-column rank, as $n > p^{(k)}$, the null-space of $\mathcal{K}^{(k)}$ has dimension $n - p^{(k)} > 0$. Letting \mathbf{W} denote an $n \times (n - p^{(k)})$ matrix of linearly independent columns spanning the null-space of $\mathcal{K}^{(k)}$,

any solution \mathbf{x}_0 satisfying $\hat{\mathbf{c}}^{\lambda(k)} = (n\lambda)^{-1}\mathbf{K}^{(k)}\mathbf{x}_0$ can be expressed as $\mathbf{x}_0 = \text{diag}(\hat{\boldsymbol{\alpha}}^\lambda)\mathbf{y} + \mathbf{W}\mathbf{b}$, for $\mathbf{b} \in \mathbb{R}^{n-p^{(k)}}$, the solution space $\mathbb{S}(\hat{\mathbf{c}}^{\lambda(k)})$ can be re-expressed as

$$\mathbb{S}(\hat{\mathbf{c}}^{\lambda(k)}) = \{\mathbf{y}^\dagger \in \{-1, 1\}^n : \mathbf{y}^\dagger = \text{sign}\{\text{diag}(\hat{\boldsymbol{\alpha}}^\lambda)\mathbf{y} + \mathbf{W}\mathbf{b}\}, \text{ with } \text{diag}(\hat{\boldsymbol{\alpha}}^\lambda)\mathbf{y} + \mathbf{W}\mathbf{b} \in (-1, 1)^n\},$$

where, in the above equation, the function $\text{sign}(\cdot)$, when applied to a vector, is understood component-wise: it returns -1 for each negative entry and 1 for each positive entry. To derive this expression, we also used the fact that any $\mathbf{y}^\dagger \in \{-1, 1\}^n$ satisfying $\hat{\mathbf{c}}^{\lambda(k)} = (n\lambda)^{-1}\mathbf{K}^{(k)}\text{diag}(\boldsymbol{\alpha})\mathbf{y}^\dagger$ for some $\boldsymbol{\alpha} \in (0, 1)^n$ satisfies $\mathbf{y}^\dagger = \text{sign}(\text{diag}(\boldsymbol{\alpha})\mathbf{y}^\dagger)$.

Using these derivations, an empirical criterion was derived to verify if, using the quantities available at the covariate-nodes, every entry of the response-node's data can be flipped while still leading to an admissible candidates for the response vector. This criterion is described in Algorithm S1 to support numerical implementation.

This criterion can be verified at the response-node for any covariate-node k not co-located at the response-node.

Algorithm S1 Empirical criterion for the privacy assessment of the response vector \mathbf{y} at covariate-node k

Input: Gram matrix $\mathbf{K}^{(k)}$ from covariate-node k , response vector \mathbf{y} and dual numerical estimate $\tilde{\boldsymbol{\alpha}}^\lambda$.

Output: Number of entries of the vector \mathbf{y} that could be flipped.

Procedure:

1. Generate \mathbf{W} in the null-space of $\mathbf{K}^{(k)}$.
 2. For every $i \in \{1, \dots, n\}$, verify if $\exists \mathbf{x}_0$ such that $\text{sign}(x_{0i}) \neq \text{sign}(y_i)$, where $\mathbf{x}_0 = \text{diag}(\tilde{\boldsymbol{\alpha}}^\lambda)\mathbf{y} + \mathbf{W}\mathbf{b} \in (-1, 1)^n$.
 3. Count the number of entries y_i that satisfied the condition.
-

Based on numerical simulations, the empirical criterion is likely to be satisfied when the sample size is sufficiently large relative to the number of covariates at node k —for example, when $n \geq 100$ and $p^{(k)} \leq 10$.

Supplementary Methods - Conservative Scenario: Response-Node Retains Intermediate Quantities

In real applications, a covariate-node may not assume that the response-node has executed the algorithm as intended. By analogy with the use of the term *honest-but-curious* in the main text, this situation can be viewed as involving a *malicious* adversary⁹, where the malicious behavior consists solely of not performing step 4 of the algorithm—that is, without collusion with other parties.

In the following, we consider the conservative scenario in which the response-node has access to the local Gram matrices, in addition to the other disclosed estimates. The ability of the response-node to attempt reverse-engineering then depends on whether the covariate-nodes disclose p -values, or parameter estimates together with their standard errors. To analyze this ability, let

$$\mathbb{S}(\mathbf{K}^{(k)}) = \left\{ \mathbf{A} \in \mathcal{M}_{n, p^{(k)}}(\mathbb{R}) \left| \begin{array}{l} \mathbf{A}\mathbf{A}^\top = \mathbf{K}^{(k)}, \mathbf{A}^\top \mathbf{1}_n = 0, \\ \text{and } \text{diag}_{\text{vec}}(\mathbf{A}^\top \mathbf{A}) = (n-1)\mathbf{1}_{p^{(k)}} \end{array} \right. \right\}.$$

The set $\mathbb{S}(\mathbf{K}^{(k)})$ consists of all candidate matrices for $\mathbf{X}_{\text{cs}}^{(k)}$ from which node k 's local Gram matrix $\mathbf{K}^{(k)}$ could equivalently be computed. When $p^{(k)} = 1$, we obtain directly that $\mathbb{S}(\mathbf{K}^{(k)}) = \{\mathbf{X}_{\text{cs}}^{(k)}, -\mathbf{X}_{\text{cs}}^{(k)}\}$. For $p^{(k)} \geq 2$, any sign-permutation matrix \mathbf{P}_π^\pm —that is, any matrix with exactly one nonzero entry in each row and each column, with each nonzero entry equal to either 1 or -1 —

can be shown to satisfy $\mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\pi^\pm \in \mathbb{S}(\mathcal{K}^{(k)})$, so that $\mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\pi^\pm$ is always an admissible candidate for $\mathbf{X}_{\text{cs}}^{(k)}$.

Note on the separation of binary and continuous covariates

In the privacy analysis, we analyze binary and continuous covariates separately. Proceeding in this way yields to conservative privacy risk assessments, since, for instance, the local Gram matrices $\mathcal{K}^{(k,\text{bin})} = \mathbf{X}^{(k,\text{bin})}(\mathbf{X}^{(k,\text{bin})})^\top$ and $\mathcal{K}^{(k,\text{cont})} = \mathbf{X}^{(k,\text{cont})}(\mathbf{X}^{(k,\text{cont})})^\top$ are not transmitted separately to the response-node; instead, only their sum, $\mathcal{K}^{(k)} = \mathcal{K}^{(k,\text{bin})} + \mathcal{K}^{(k,\text{cont})}$, is disclosed. We proceed in this way because it remains unclear, at this stage, under which conditions the above decomposition is unique, and thus whether there is a risk that the response-node could uniquely recover $\mathcal{K}^{(k,\text{bin})}$ and $\mathcal{K}^{(k,\text{cont})}$ from the aggregate matrix $\mathcal{K}^{(k)}$.

When only the local Gram matrix and parameter estimates are available to the response-node

When, in addition to $\mathcal{K}^{(k)}$, the response node has access to a selected subset $J \subseteq \{1, \dots, p^{(k)}\}$ of the components of the estimated parameters $\hat{\beta}^{\lambda(k)}$ (which may include all components), reverse-engineering $\mathbf{X}_{\text{cs}}^{(k)}$ would require the response-node to search for a matrix $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ that satisfies

$$s_{n,j}^{(k)} \hat{\beta}_j^{\lambda(k)} = (n\lambda)^{-1} \sum_{i=1}^n [\mathbf{A}]_{ij} \hat{\alpha}_i^\lambda y_i \quad \text{for all } j \in J,$$

where the values $s_{n,j}^{(k)}$ are not available to the response-node.

When $\mathbf{X}_{\text{cs}}^{(k)}$ is computed from continuous covariate data, successful reverse-engineering of node k 's data would require the response-node to identify a matrix \mathbf{A} such that there exists a set of constants $\{a_j : j \in J\}$ with strictly positive entries ($a_j > 0$ for all j) satisfying

$$a_j \hat{\beta}_j^{\lambda(k)} = (n\lambda)^{-1} \sum_{i=1}^n [\mathbf{A}]_{ij} \hat{\alpha}_i^\lambda y_i \quad \text{for all } j \in J. \quad (\text{S8})$$

Now for any $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$, if \mathbf{A} does not satisfy (S8), there exists a set of indices $J' \subseteq J$ such that $(\hat{\beta}_j^{\lambda(k)} n\lambda)^{-1} \sum_{i=1}^n [\mathbf{A}]_{ij} \hat{\alpha}_i^\lambda y_i < 0$ for all $j \in J'$. Since flipping the signs of all entries in any given column of a matrix in $\mathbb{S}(\mathcal{K}^{(k)})$ yields another matrix that also belongs to $\mathbb{S}(\mathcal{K}^{(k)})$, it follows that the matrix $\mathbf{A}' = \mathbf{A} \prod_{j \in J'} \mathbf{P}_j^\pm$ still belongs to $\mathbb{S}(\mathcal{K}^{(k)})$ while now meeting the constraint given by (S8). We draw the following conclusions:

- Case $p^{(k)} = 1$: In this case, as $\mathbb{S}(\mathcal{K}^{(k)}) = \{\mathbf{X}_{\text{cs}}^{(k)}, -\mathbf{X}_{\text{cs}}^{(k)}\}$, for any $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$, only one of \mathbf{A} and $\mathbf{A}' = \mathbf{A} \mathbf{P}_1^\pm$ will meet (S8). Centered and scaled individual-level data could be retrieved.
- Case $p^{(k)} \geq 2$: Given that matrices $\mathbf{A} = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\pi^\pm$ are in $\mathbb{S}(\mathcal{K}^{(k)})$, it follows that at least $\mathbf{A} = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\pi$, where \mathbf{P}_π denotes a permutation matrix (i.e., either the identity matrix, or the matrix that permutes the columns) will meet constraint (S8).

In the binary-covariates case, as the binary nature must be preserved, it can be shown that candidate matrices for $\mathbf{X}_{\text{cs}}^{(k)}$ are all of the form $\mathbf{A} = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\pi^\pm$. Therefore, if the response-node is able to identify a single admissible candidate in $\mathbb{S}(\mathcal{K}^{(k)})$, it can compute all candidates in $\mathbb{S}(\mathcal{K}^{(k)})$. The binary nature of the covariates involved at node k implies that the entries $\mathbf{X}_{\text{cs}}^{(k)}$ have the form

$$[\mathbf{X}_{\text{cs}}^{(k)}]_{ij} = \frac{x_{ij}^{(k)} - \mu_j}{\sqrt{\mu_j(1 - \mu_j)}} \sqrt{\frac{n-1}{n}}, \quad x_{ij}^{(k)} \in \{0, 1\}, \quad \mu_j \in \left\{ \frac{1}{n}, \dots, \frac{n-1}{n} \right\}.$$

287 This implies that the proportion of positive entries in the j th column of $\mathbf{X}_{\text{cs}}^{(k)}$ is equal to $\mu_{n,j}^{(k)}$, and
 288 the standard deviation of the covariate in column j can therefore be computed as

$$\sqrt{\frac{n}{n-1}} \sqrt{\left\{ n^{-1} \sum_{i=1}^n \mathbb{I}(x_{ij,\text{cs}}^{(k)} > 0) \right\} \left\{ 1 - n^{-1} \sum_{i=1}^n \mathbb{I}(x_{ij,\text{cs}}^{(k)} > 0) \right\}},$$

289 where $\mathbb{I}(B)$ is the indicator function taking value 1 if B is true, and 0 otherwise. Then, recalling
 290 equation (S8), if $\hat{\beta}_j^{(k)}$ is disclosed for all $j \in J$, the response-node can narrow its search for $\mathbf{X}_{\text{cs}}^{(k)}$ by
 291 identifying which admissible candidate $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ satisfies the following equality:

$$\sqrt{\frac{n}{n-1}} \sqrt{\mu_j^A (1 - \mu_j^A)} = (\hat{\beta}_j^{\lambda(k)} n \lambda)^{-1} \sum_{i=1}^n A_{ij} \hat{\alpha}_i^\lambda y_i \quad \text{for all } j \in J, \quad (\text{S9})$$

292 with $\mu_j^A = n^{-1} \sum_{i=1}^n \mathbb{I}([\mathbf{A}]_{ij} > 0)$. This typically yields a set of admissible candidates with a unique
 293 possibility for every column $j \in J$ of $\mathbf{X}_{\text{cs}}^{(k)}$, irrespective of the cardinality of J .

294 Therefore, in the conservative non-colluding malicious adversary setting, disclosing parameter
 295 estimates for binary covariates entails a risk of data leakage.

296 **When only the local Gram matrix, parameter estimates and their standard errors are** 297 **available to the response-node**

298 We now turn to the solution space of candidate matrices \mathbf{A} when coefficient estimates indexed
 299 by $J \subseteq \{1, \dots, p^{(k)}\}$ are disclosed, and for a subset of these, $J_{\text{sd}} \subseteq J$, the corresponding standard
 300 errors are also released. We restrict attention to the case of continuous covariates, since, as shown
 301 previously, the disclosure of parameter estimates for binary covariates entails a risk of data leakage.

302 Recall the expression of $\hat{\sigma}_j^{\lambda(k)}$ for $j \in \{1, \dots, p^{(k)}\}$, given by

$$s_{n,j}^{(k)} \hat{\sigma}_j^{\lambda(k)} = \left\{ n^{-1} \eta^{-1} - n^{-1} \eta^{-2} [(\mathbf{X}_{\text{cs}}^{(k)})^\top \hat{\mathbf{S}}^{-1} \mathbf{X}_{\text{cs}}^{(k)}]_{jj} \right\}^{1/2}. \quad (\text{S10})$$

303 While the values $s_{n,j}^{(k)}$ are not known to the response-node, for $j \in J$ the coefficients are $\hat{\beta}_j^{(k)}$.
 304 Moreover, from (S8), a candidate $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ implicitly determines a corresponding candidate
 305 value for $s_{n,j}^{(k)}$, given by $a_j = (\hat{\beta}_j^{\lambda(k)} n \lambda)^{-1} \sum_{i=1}^n [\mathbf{A}]_{ij} \hat{\alpha}_i^\lambda y_i$. Consequently, reverse-engineering $\mathbf{X}_{\text{cs}}^{(k)}$
 306 requires the response-node to search for $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ that satisfies (S8) for all $j \in J$, as well as the
 307 following for all $j \in J_{\text{sd}}$:

$$n^{-1} \eta^{-1} - n^{-1} \eta^{-2} \sum_{m=1}^n \sum_{i=1}^n [\mathbf{A}]_{ij} [\mathbf{A}]_{mj} [\hat{\mathbf{S}}^{-1}]_{im} = \left(\frac{\hat{\sigma}_j^{\lambda(k)}}{\hat{\beta}_j^{\lambda(k)} n \lambda} \sum_{i=1}^n [\mathbf{A}]_{ij} \hat{\alpha}_i^\lambda y_i \right)^2. \quad (\text{S11})$$

308 We next show that, when $p^{(k)} \geq 4$ and $|J_{\text{sd}}| \leq (p^{(k)} - 1)(p^{(k)} - 2)/2 - 1$ and $\mathbf{X}_{\text{cs}}^{(k)}$ has full column
 309 rank, the set of candidate matrices $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ that simultaneously satisfy (S8) for all $j \in J$ and
 310 (S11) for all $j \in J_{\text{sd}}$ is infinite.

311 To do so, for $\boldsymbol{\theta} \in (-\pi, \pi]^{p^{(k)}(p^{(k)}-1)/2}$, let $\mathbf{P}_{\boldsymbol{\theta}}$ be defined as:

$$\mathbf{P}_{\boldsymbol{\theta}} = \prod_{i=1}^{p^{(k)}-1} \prod_{j=i+1}^{p^{(k)}} \mathcal{G}(\theta_{(i-1)(p^{(k)}-i/2)+(j-i)}; i, j), \quad (\text{S12})$$

where, for $1 \leq i < j \leq p^{(k)}$ and $\theta \in (-\pi, \pi]$, $\mathbf{g}(\theta; i, j) \in \mathcal{M}_{p^{(k)} \times p^{(k)}}(\mathbb{R})$ denotes the Givens rotation matrix (see e.g.¹⁰) with entries $[\mathbf{g}(\theta; i, j)]_{\ell\ell'}$ defined as

$$[\mathbf{g}(\theta; i, j)]_{\ell\ell'} = \begin{cases} \cos(\theta) & \text{if } (\ell, \ell') \in \{(i, i), (j, j)\}, \\ -\sin(\theta) & \text{if } (\ell, \ell') = (i, j), \\ \sin(\theta) & \text{if } (\ell, \ell') = (j, i), \\ 1 & \text{if } (\ell, \ell') \in \{(1, 1), \dots, (p^{(k)}, p^{(k)})\} \setminus \{(i, i), (j, j)\}, \\ 0 & \text{elsewhere.} \end{cases} \quad (\text{S13})$$

In this notation, the set $\{\mathbf{P}_\theta : \theta \in (-\pi, \pi]^{p^{(k)}(p^{(k)}-1)/2}\}$ is exactly the set of all rotation matrices of size $p^{(k)} \times p^{(k)}$, from which the full set of orthogonal matrices can be generated by additionally allowing reflections.

Consider $\mathbf{g}(\theta) = [g_1(\theta), \dots, g_{p^{(k)}-1}(\theta)]$, where, for $j \in \{1, \dots, p^{(k)} - 1\}$,

$$g_j(\theta) = \sum_{\ell=1}^{p^{(k)}-1} \sum_{\ell'=\ell+1}^{p^{(k)}} [\mathbf{P}_\theta]_{\ell j} [\mathbf{P}_\theta]_{\ell' j} \tau_{\ell\ell'}^{(k)},$$

where $\tau_{\ell\ell'}^{(k)} = n^{-1} \sum_{i=1}^n x_{i\ell, \text{cs}}^{(k)} x_{i\ell', \text{cs}}^{(k)}$. Since $\mathbf{A}_\theta = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\theta$ is centered and scaled matrix if and only if $\mathbf{g}(\theta) = \mathbf{0}$, it can be readily verified that that

$$\left\{ \mathbf{A} = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\theta \mathbf{P}_\pi^\pm \mid \theta \in (-\pi, \pi]^{p^{(k)}(p^{(k)}-1)/2} \text{ and } \mathbf{g}(\theta) = \mathbf{0} \right\} = \mathbb{S}(\mathcal{K}^{(k)}).$$

Moreover, such \mathbf{A}_θ satisfies (S11) for all $j \in J_{\text{sd}}$ if and only if $\mathbf{h}_{J_{\text{sd}}}(\theta) = \mathbf{0}$, where $\mathbf{h}(\theta) = [h_1(\theta), \dots, h_{p^{(k)}}(\theta)]$ with

$$h_j(\theta) = \frac{1}{n\eta} - \frac{1}{n\eta^2} \sum_{\ell=1}^{p^{(k)}} \sum_{\ell'=1}^{p^{(k)}} [\mathbf{P}_\theta]_{\ell j} [\mathbf{P}_\theta]_{\ell' j} [(\mathbf{X}_{\text{cs}}^{(k)})^\top \widehat{\mathbf{S}}^{-1} \mathbf{X}_{\text{cs}}^{(k)}]_{\ell\ell'} - \left(\frac{\widehat{\sigma}_j^{\lambda(k)}}{\widehat{\beta}_j^{\lambda(k)}} \sum_{\ell=1}^{p^{(k)}} [\mathbf{P}_\theta]_{\ell j} s_{n,\ell}^{(k)} \widehat{\beta}_\ell^{(k)} \right)^2.$$

Now note that the cardinality of the set $\{\theta \in (-\pi, \pi]^{p^{(k)}(p^{(k)}-1)/2} : \mathbf{g}(\theta) = \mathbf{0} \text{ and } \mathbf{h}_{J_{\text{sd}}}(\theta) = \mathbf{0}\}$ is infinite. This follows directly from the smoothness of $\mathbf{g}(\theta)$, $\mathbf{h}_{J_{\text{sd}}}(\theta)$, the fact that if $|J_{\text{sd}}| \leq (p^{(k)} - 1)(p^{(k)} - 2)/2 - 1$ the dimension of θ exceeds the number of equations, the existence of a solution $\theta_0 \in (-\pi, \pi)^{p^{(k)}(p^{(k)}-1)/2}$ satisfying this solution set (i.e., the solution $\theta = \mathbf{0}$, which corresponds to $\mathbf{P}_\theta = \mathbf{I}_{p^{(k)}}$), and an application of the Implicit Function Theorem.

Since for any $\mathbf{A}_\theta = \mathbf{X}_{\text{cs}}^{(k)} \mathbf{P}_\theta$ with $\mathbf{g}(\theta) = \mathbf{0}$ and $\mathbf{h}_{J_{\text{sd}}}(\theta) = \mathbf{0}$ we have $\mathbf{A}_\theta \in \mathbb{S}(\mathcal{K}^{(k)})$, and as all matrices of the form $\mathbf{A}_\theta \mathbf{P}^\pm$ satisfy both $\mathbf{A}_\theta \mathbf{P}^\pm \in \mathbb{S}(\mathcal{K}^{(k)})$ and (S11), the proof that the set of candidate matrices $\mathbf{A} \in \mathbb{S}(\mathcal{K}^{(k)})$ that simultaneously satisfy (S8) for all $j \in J$ and (S11) for all $j \in J_{\text{sd}}$ is infinite follows from the fact that for any θ there always exists a \mathbf{P}^\pm such that $\mathbf{A}_\theta \mathbf{P}^\pm$ satisfies (S8).

It follows that the targeted estimates and standard errors can be disclosed by the covariate-node while ensuring Privacy Level II.

When only the local Gram matrix and two-sided p-values are available to the response-node

We now study the solution space of candidate matrices \mathbf{A} for $\mathbf{X}_{\text{cs}}^{(k)}$ when a subset $J \subseteq \{1, \dots, p^{(k)}\}$ of two-sided p-values $\rho_{\text{val},j}^{\lambda(k)}$ are available to the response-node, in addition to the local Gram matrix

338 $\mathcal{K}^{(k)}$. Given the one-to-one relationship between $\rho_{\text{val},j}^{\lambda(k)}$ and $|\beta_j^{\lambda(k)}|/\hat{\sigma}_j^{\lambda(k)}$, and using equations (S8)
 339 and (S10), for a candidate matrix \mathbf{A} for $\mathbf{X}_{\text{cs}}^{(k)}$ to be such that the disclosed quantities $\mathcal{K}^{(k)}$ and
 340 $\rho_{\text{val},J}^{\lambda(k)} := \{\rho_{\text{val},j}^{\lambda(k)} : j \in J\}$ could have been equivalently computed from either \mathbf{A} or $\mathbf{X}_{\text{cs}}^{(k)}$, the
 341 response-node must identify a matrix \mathbf{A} such that

$$(\rho_{\text{val},j}^{\lambda(k)})^2 \left(n^{-1}\eta^{-1} - n^{-1}\eta^{-2} \sum_{m=1}^n \sum_{i=1}^n [\mathbf{A}]_{ij} [\mathbf{A}]_{mj} [\hat{\mathbf{S}}^{-1}]_{im} \right) = \left(\frac{1}{n\lambda} \sum_{i=1}^n A_{ij} \hat{\alpha}_i^{\lambda} y_i \right)^2. \quad (\text{S14})$$

342 It can be easily verified that $\{\mathbf{X}_{\text{cs}}^{(k)}, -\mathbf{X}_{\text{cs}}^{(k)}\}$ are admissible candidates that respect equation
 343 (S14).

344 In Figure S1, we illustrated the conditions under which a covariate-node holding only continuous
 345 covariates can disclose p -values, parameter estimates and standard errors while achieving Privacy
 346 Level II. The privacy-preserving properties for the case of a covariate-node k holding only continuous
 347 covariates are illustrated in Figure S1.

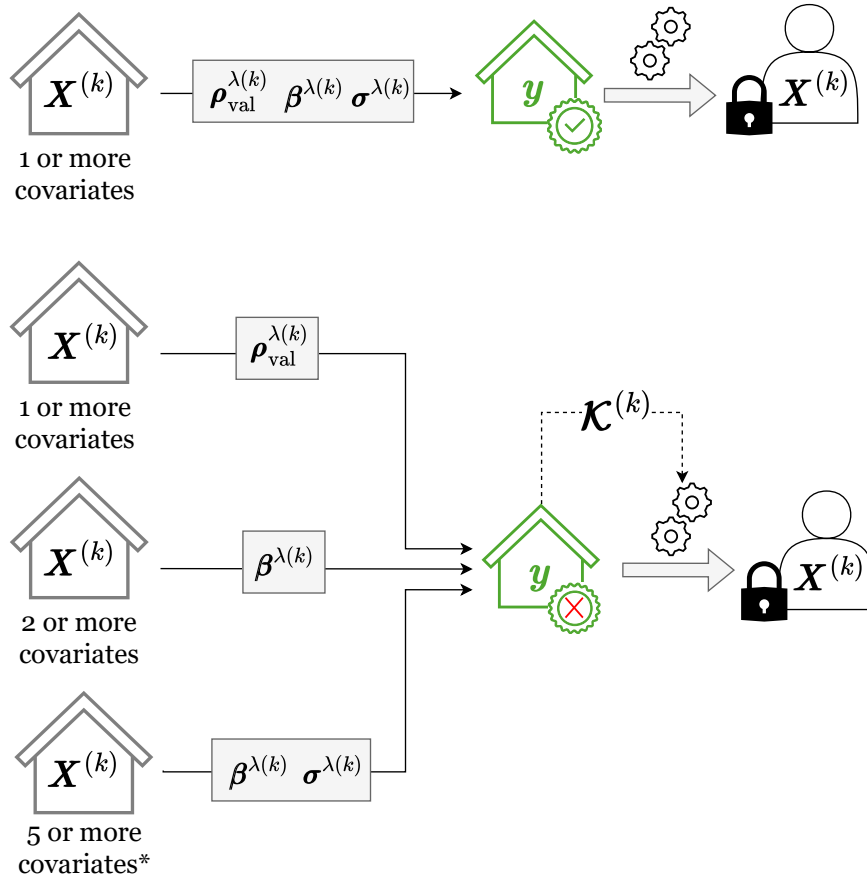


Fig. S1. Privacy-preserving properties for a covariate-node k holding continuous covariates only. The response-nodes are assumed to have cleared intermediate quantities in the upper figure, and it is assumed to have not cleared intermediate quantities in the lower figure. *If $\mathbf{X}^{(k)}$ holds 4 continuous covariates, Privacy Level II is guaranteed when at most two pairs of estimates and standard errors are disclosed. If $\mathbf{X}^{(k)}$ holds 5 or more continuous covariates, Privacy Level II is guaranteed all pairs are disclosed.

References

1. Tangri, N. *et al.* A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure. *JAMA* **305**, 1553–1559, DOI: [10.1001/jama.2011.451](https://doi.org/10.1001/jama.2011.451) (2011).
2. Li, Y., Jiang, X., Wang, S., Xiong, H. & Ohno-Machado, L. VERTIcal grid lOgistic regression (VERTIGO). *J. Am. Med. Informatics Assoc.* **23**, 570–579, DOI: [10.1093/jamia/ocv146](https://doi.org/10.1093/jamia/ocv146) (2016).
3. Wedderburn, R. W. M. On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. *Biometrika* **63**, 27–32, DOI: [10.2307/2335080](https://doi.org/10.2307/2335080) (1976).
4. Albert, A. & Anderson, J. A. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10 (1984).
5. Bertsekas, D. P. *Nonlinear Programming* (Athena Scientific, Belmont, MA, USA, 1999).
6. Bertsekas, D. P. Projected Newton Methods for Optimization Problems with Simple Constraints. *SIAM J. on Control. Optim.* **20**, 221–246, DOI: [10.1137/0320018](https://doi.org/10.1137/0320018) (1982).
7. Schmidt, M., Kim, D. & Sra, S. Projected Newton-type Methods in Machine Learning. In *Optimization for Machine Learning*, 305–330, DOI: [10.7551/mitpress/8996.003.0013](https://doi.org/10.7551/mitpress/8996.003.0013) (The MIT Press, Cambridge, MA, USA, 2011).
8. Boyd, S. P. & Vandenberghe, L. *Convex optimization* (Cambridge university press, Cambridge, United Kingdom, 2004).
9. Aggarwal, C. C. & Yu, P. S. *Privacy-Preserving Data Mining: Models and Algorithms* (Springer, New York, NY, USA, 2008).
10. Givens, W. Computation of Plane Unitary Rotations Transforming a General Matrix to Triangular Form. *J. Soc. for Ind. Appl. Math.* **6**, 26–50 (1958).