Supplementary Materials to "Robust and flexible inference for the covariate-specific ROC curve"

Vanda Inácio, Vanda M. Lourenço, Miguel de Carvalho, Richard A. Parker, and Vincent Gnanapragasam

In this supplementary file we provide additional figures and tables for the Simulation Study and Application sections in the main paper. We further present additional simulation scenarios.

Vanda Inácio, School of Mathematics, University of Edinburgh, Scotland, UK (*vanda.inacio@ed.ac.uk*). Vanda M. Lourenço, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal (*vmml@fct.unl.pt*). Miguel de Carvalho, School of Mathematics, University of Edinburgh, Scotland, UK (*Miguel.deCarvalho@ed.ac.uk*), Richard A. Parker, Edinburgh Clinical Trials Unit, Usher Institute, University of Edinburgh, Scotland, UK (*Richard.Parker@ed.ac.uk*), Vincent Gnanapragasam, Cambridge Urology Translational Research and Clinical Trials Office, Cambridge University Hospitals NHS Foundation Trust, and Academic Urology Group, Department of Surgery, University of Cambridge and Department of Urology, Cambridge University Hospitals Trust, UK (*vjg29@cam.ac.uk*).

Didactic figures

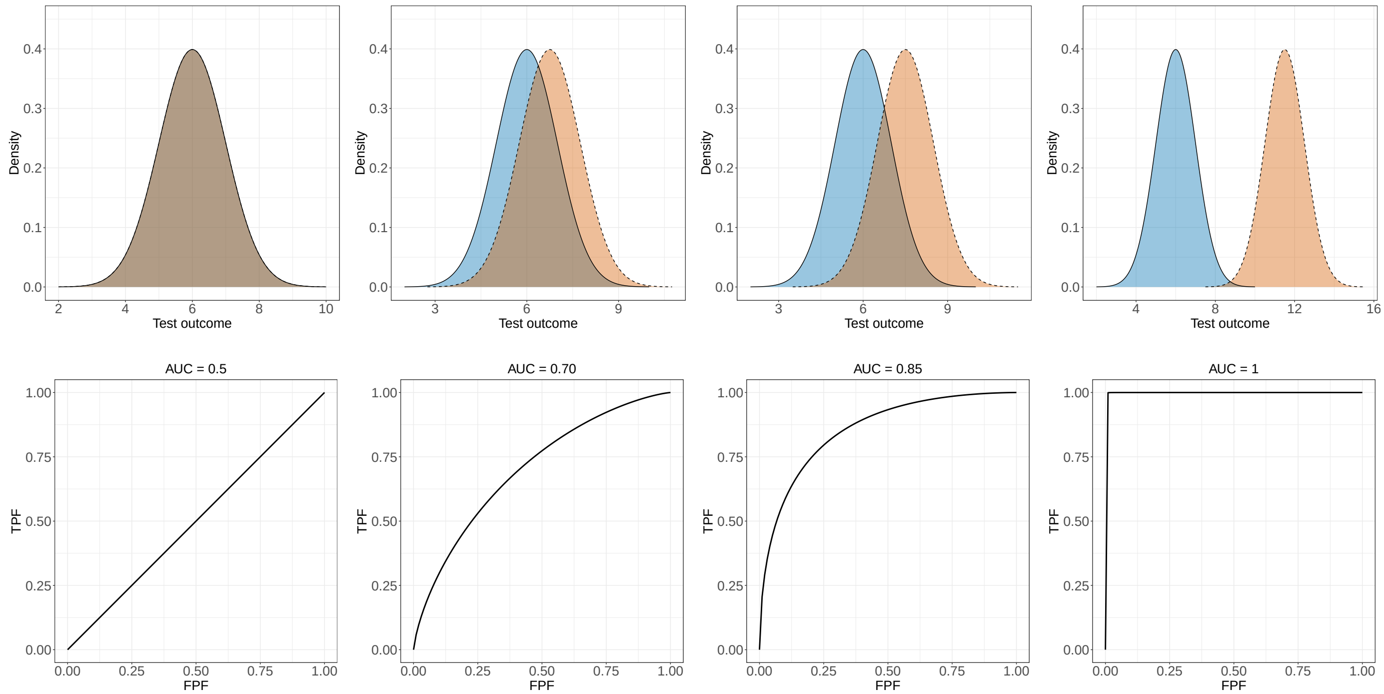


Figure 1: Hypothetical density functions of test outcomes in the diseased (dotted line, orange) and nondiseased (solid line, blue) populations (top) along with the corresponding ROC curves (bottom).

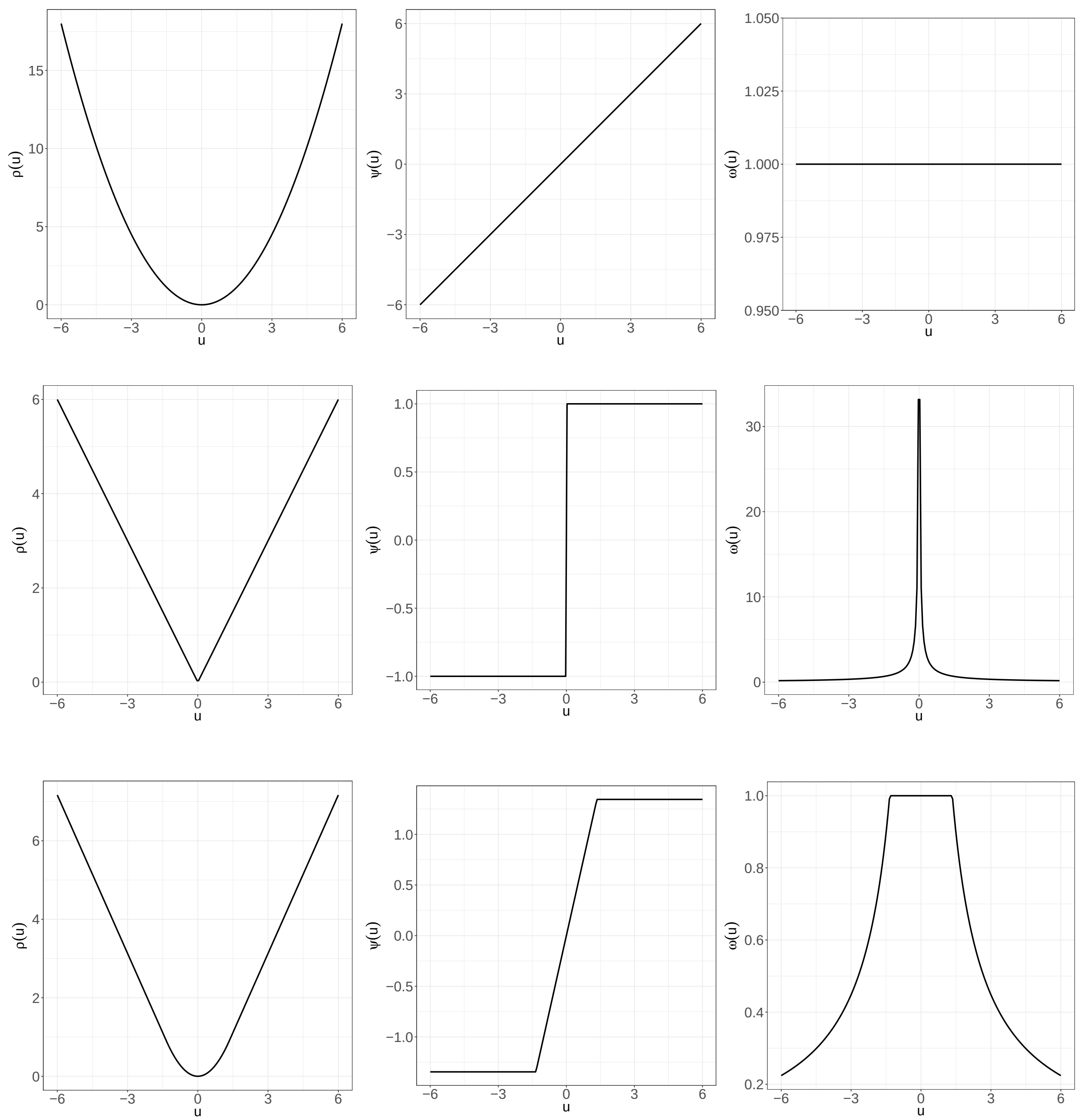


Figure 2: ρ, ψ, and ω functions for the least-squares (first row), least absolute deviations (second row) and Huber (third row) estimators. Huber's tuning constant was set to 1.345.

Additional figures and tables for simulation scenarios I–IV in the main paper

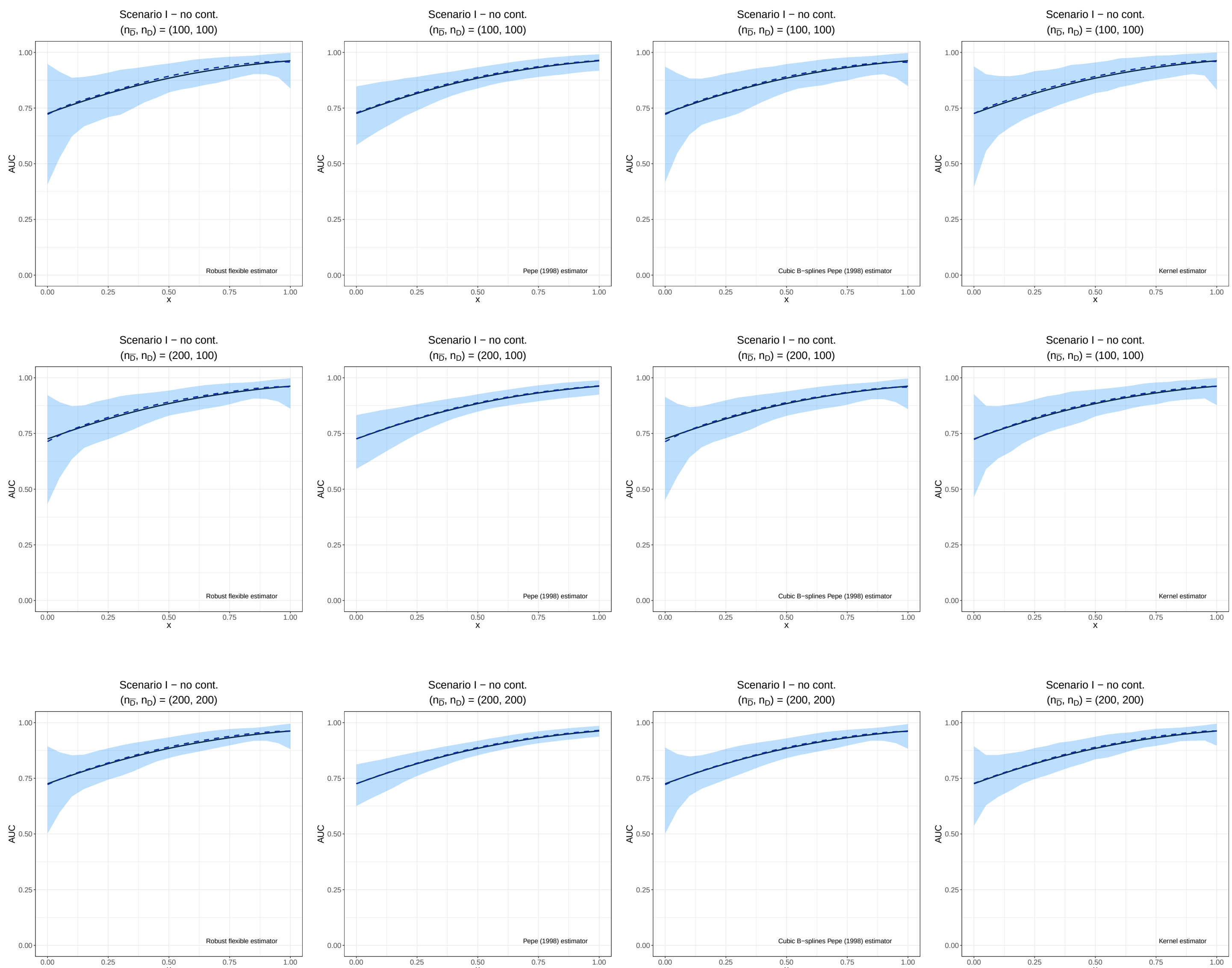


Figure 3: Scenario I. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of no contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

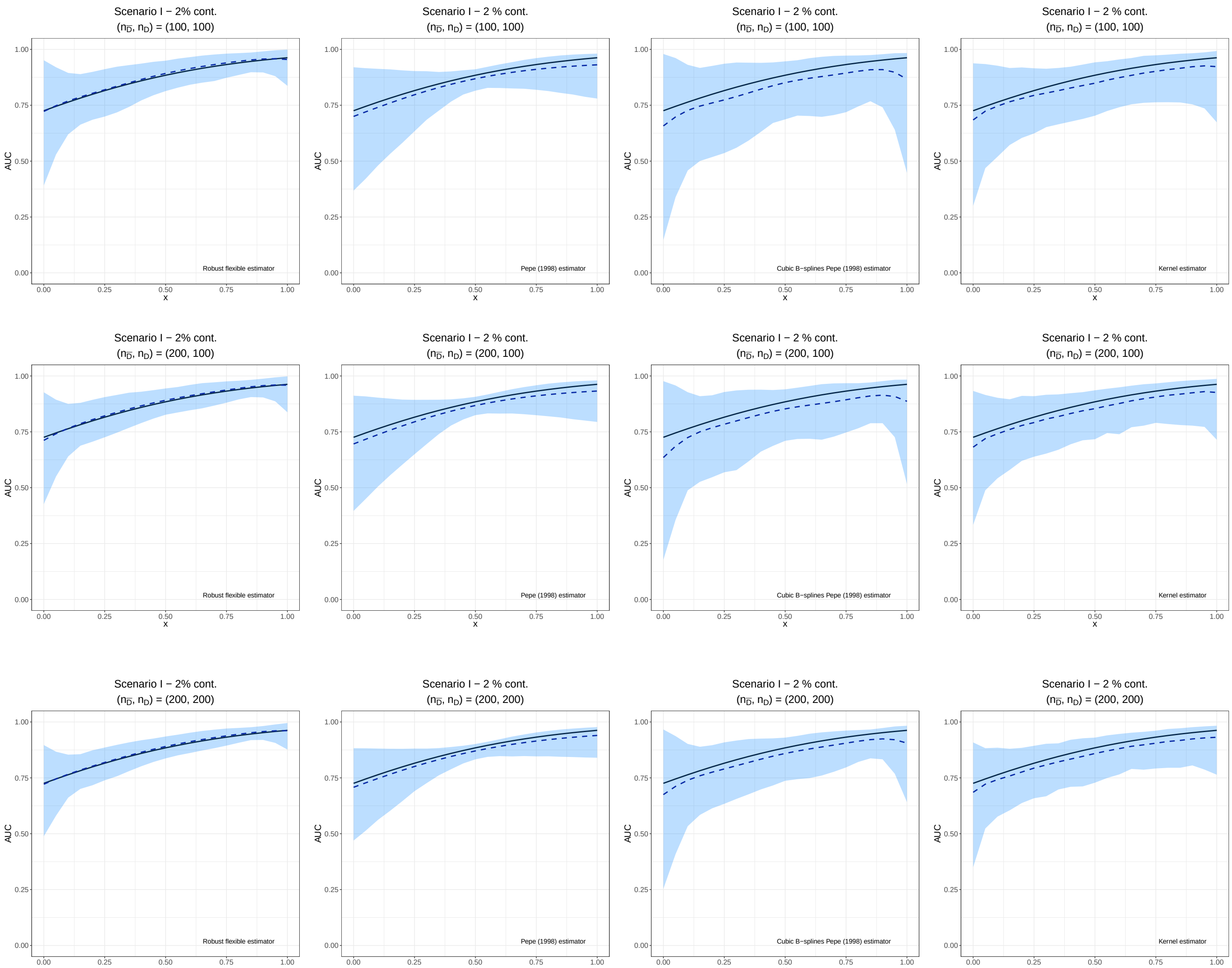


Figure 4: Scenario I. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 2% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

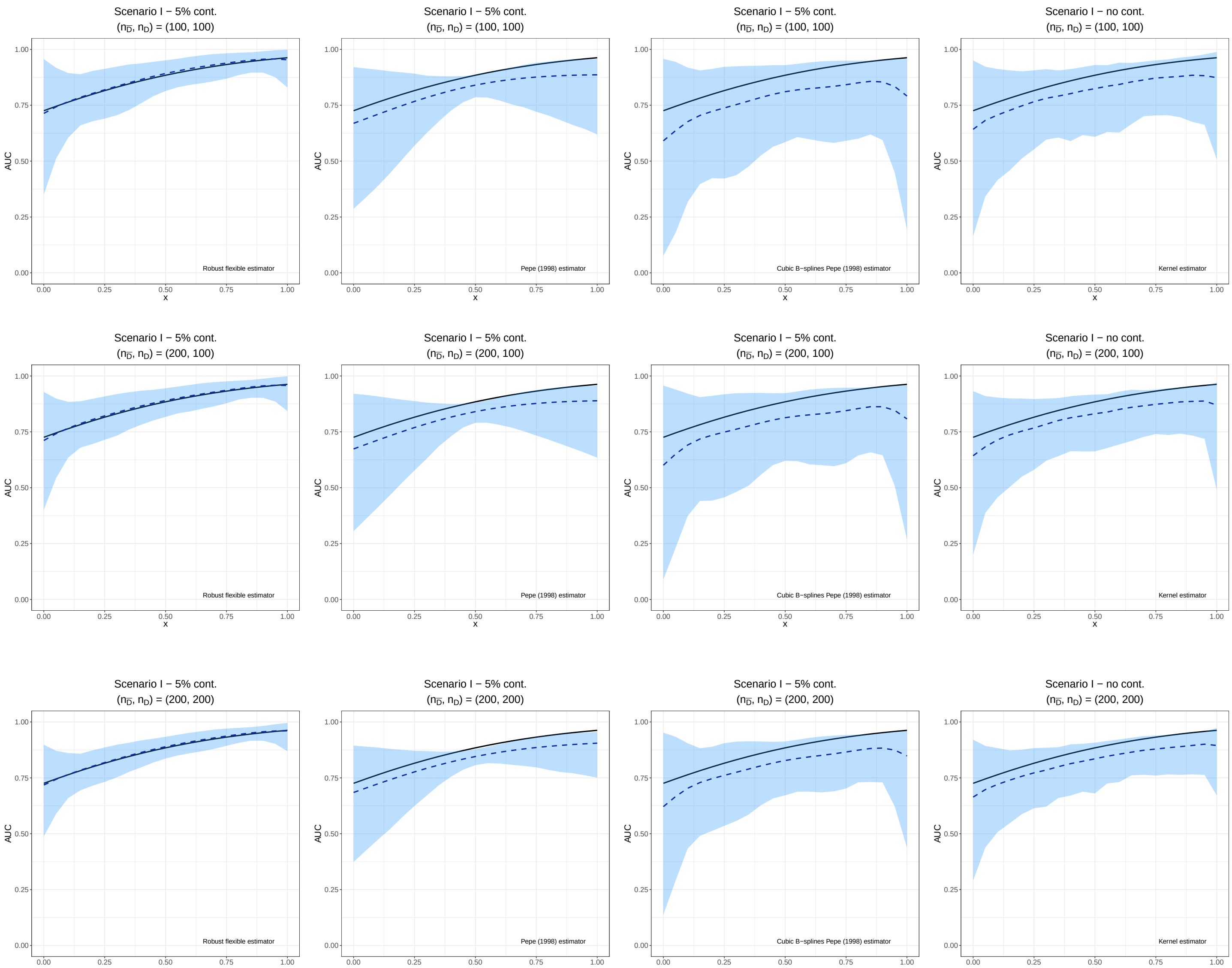


Figure 5: Scenario I. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 5% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

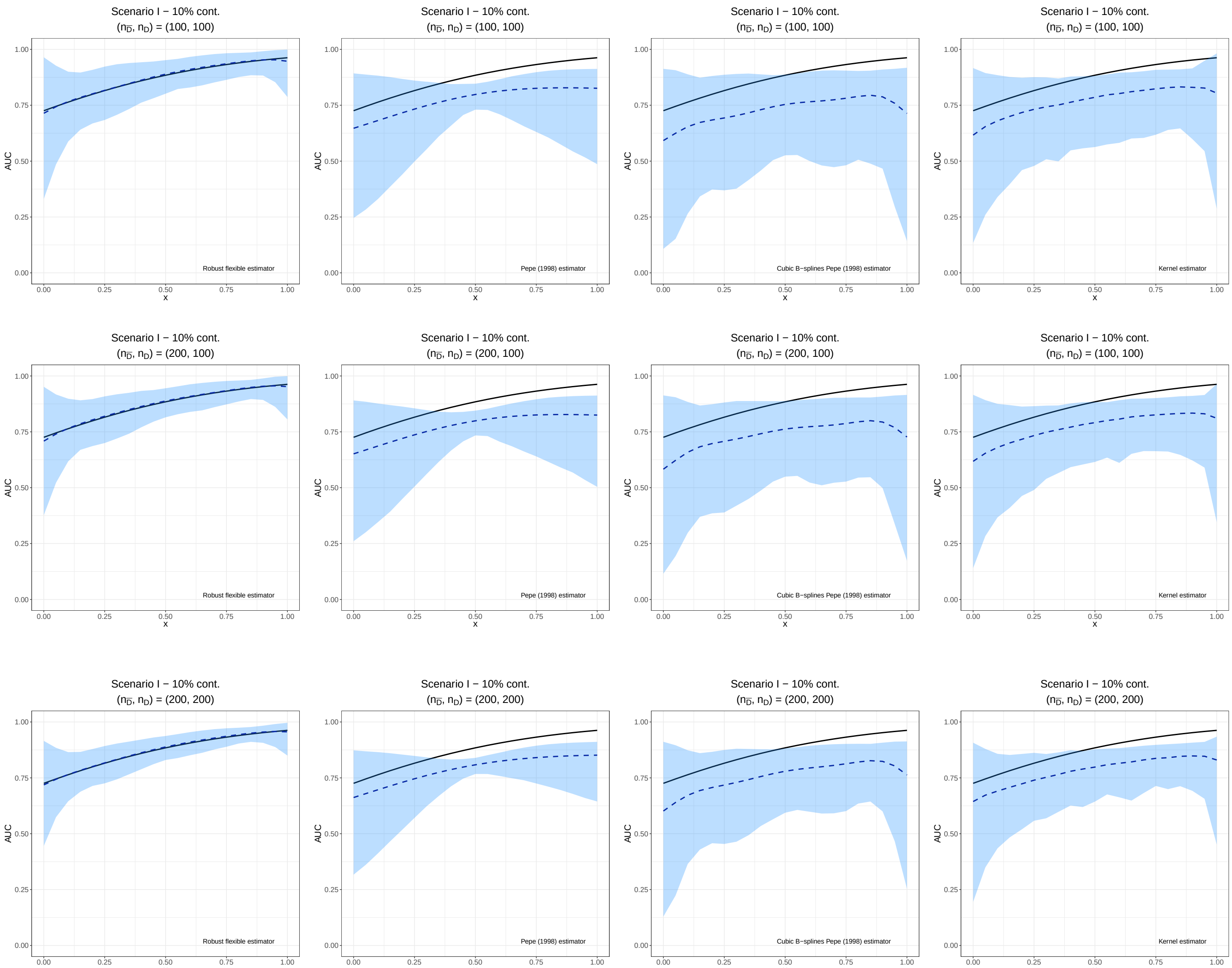


Figure 6: Scenario I. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 10% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

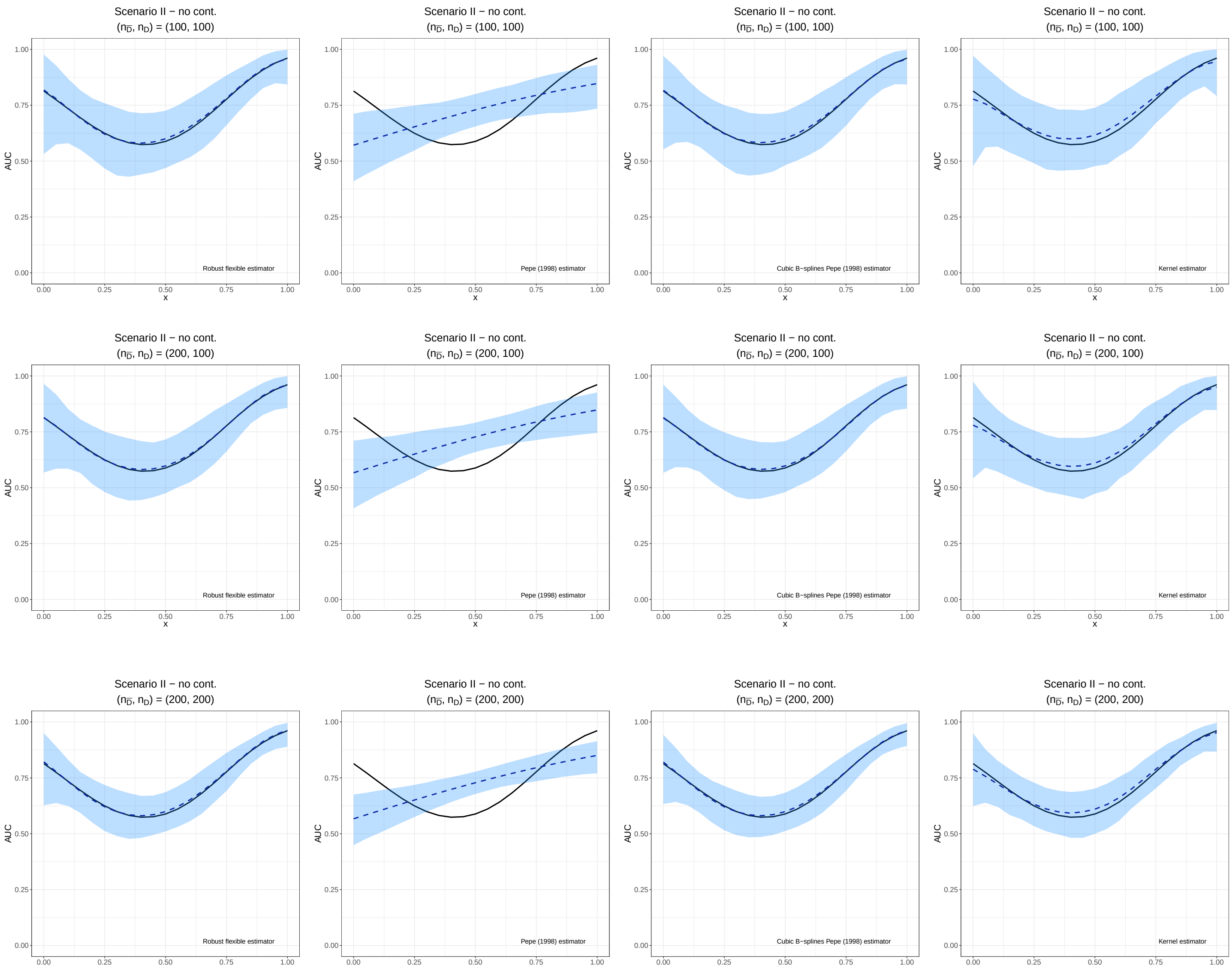


Figure 7: Scenario II. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of no contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

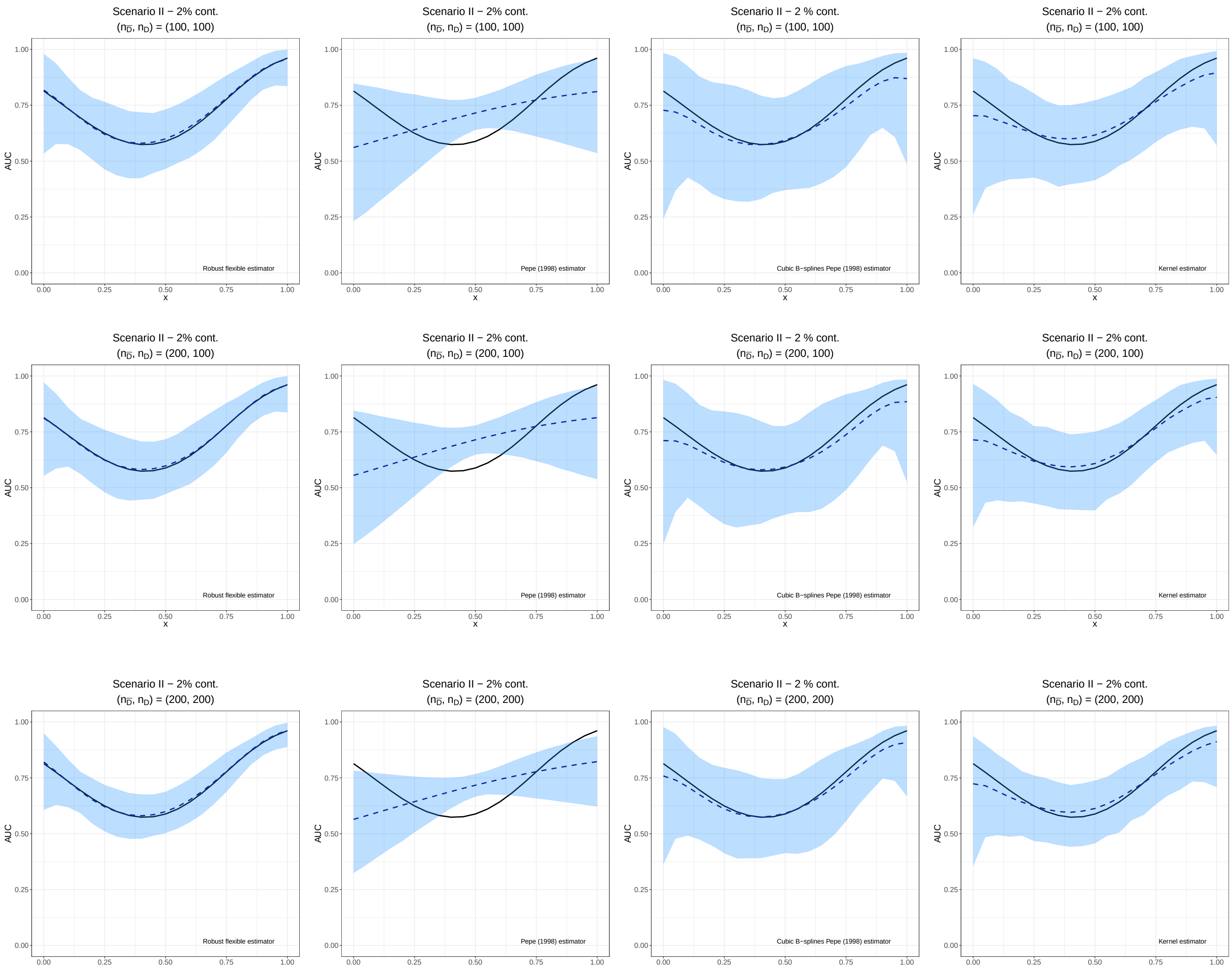


Figure 8: Scenario II. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 2% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

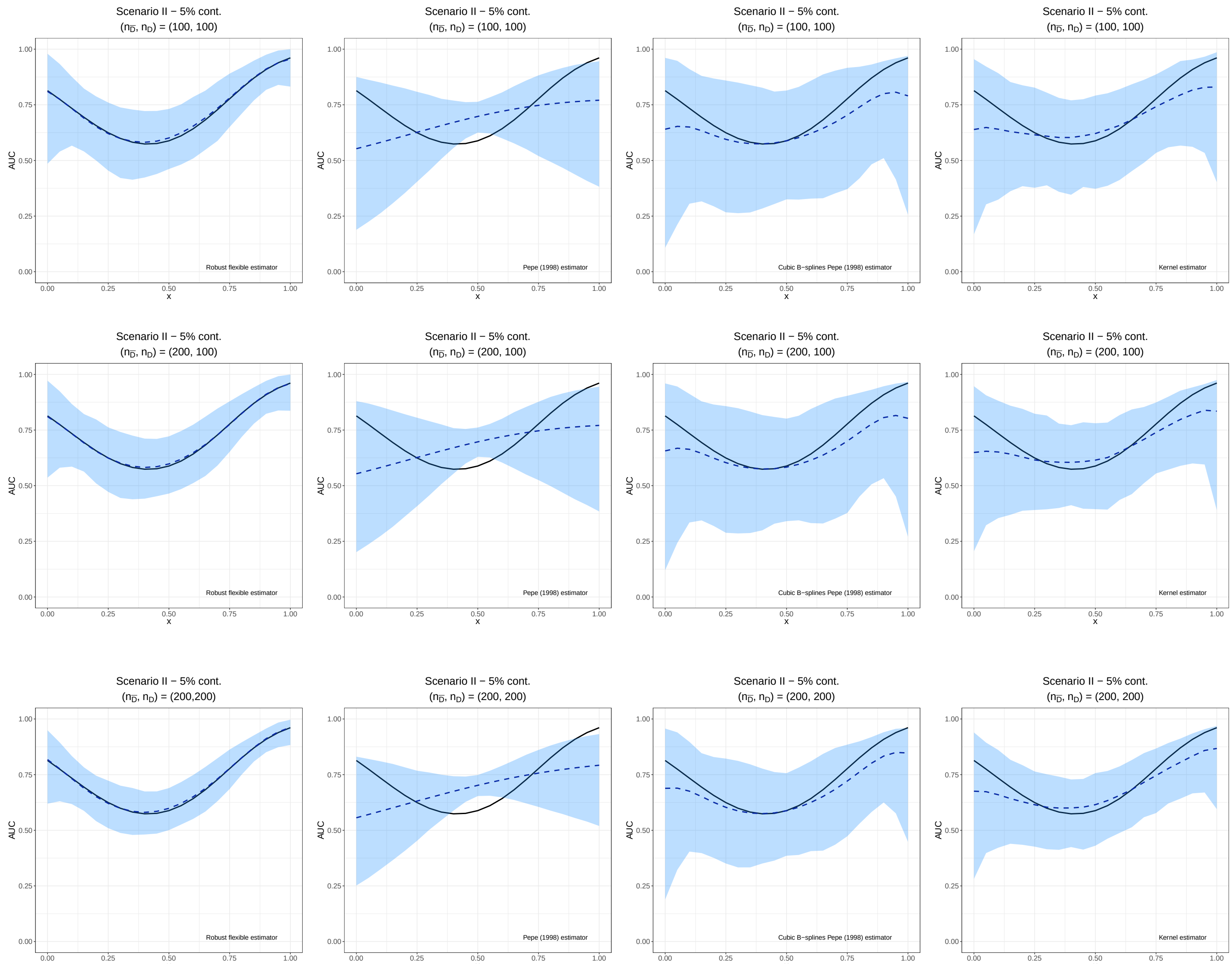


Figure 9: Scenario II. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 5% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

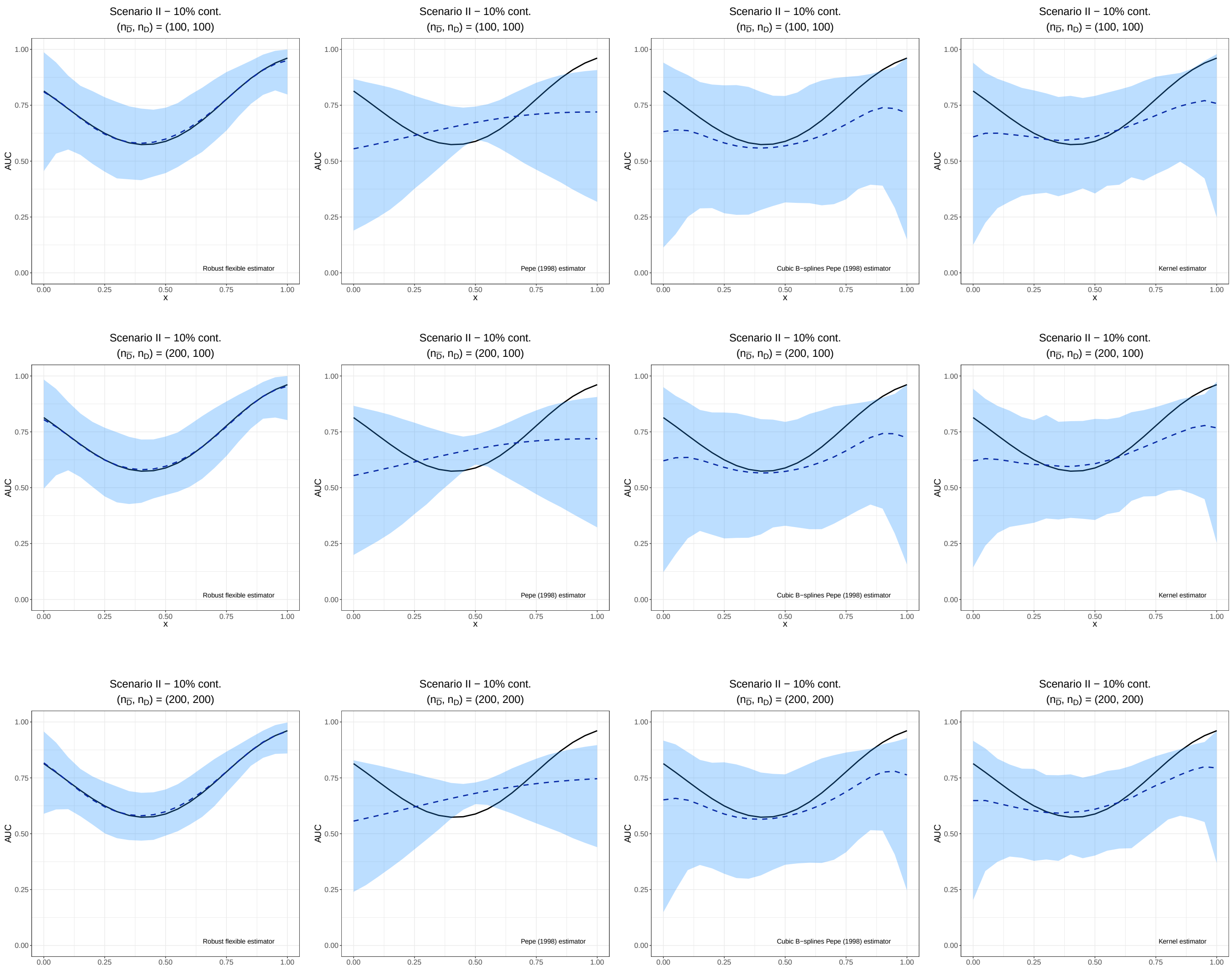


Figure 10: Scenario II. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 10% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

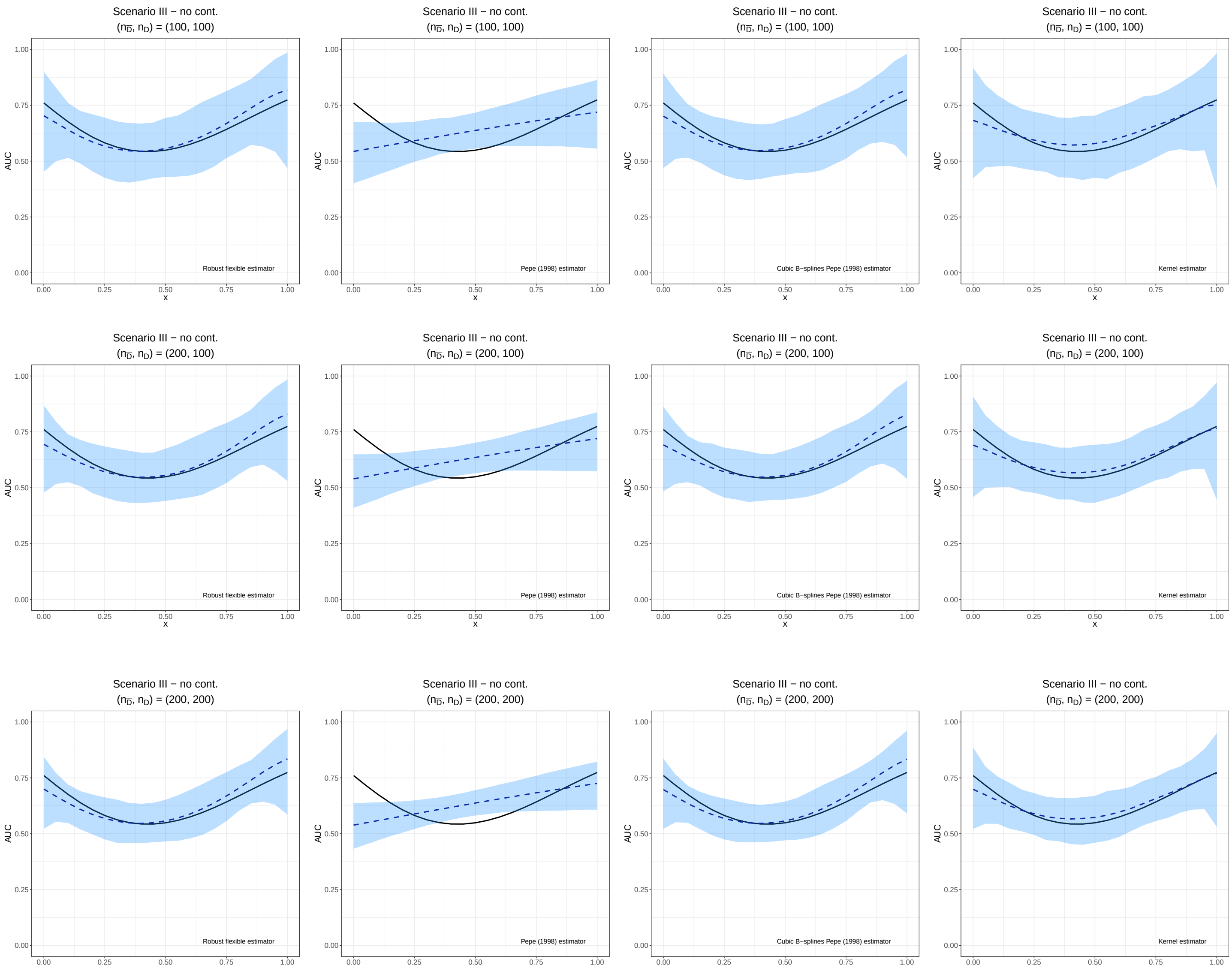


Figure 11: Scenario III. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of no contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

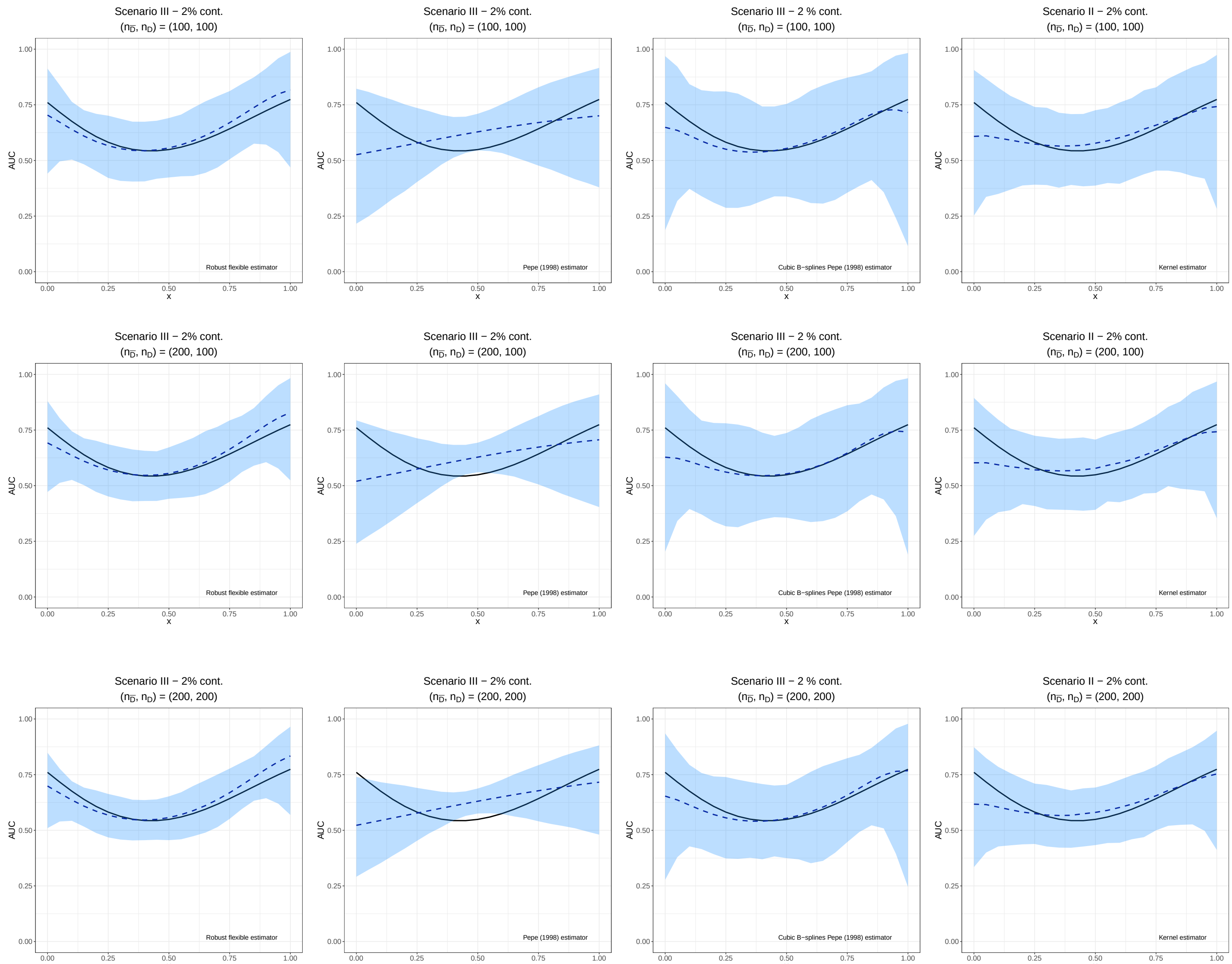


Figure 12: Scenario III. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 2% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

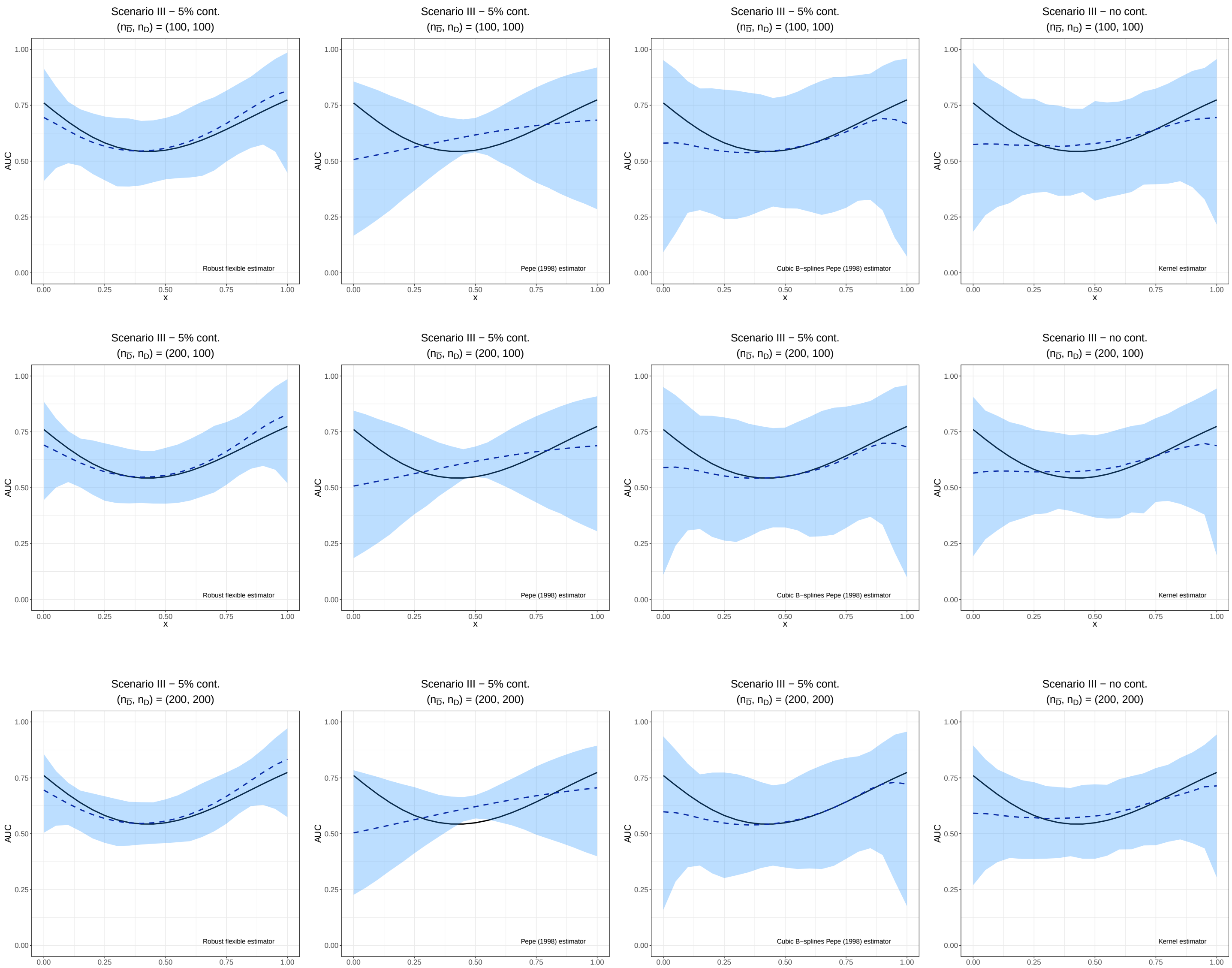


Figure 13: Scenario III. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 5% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

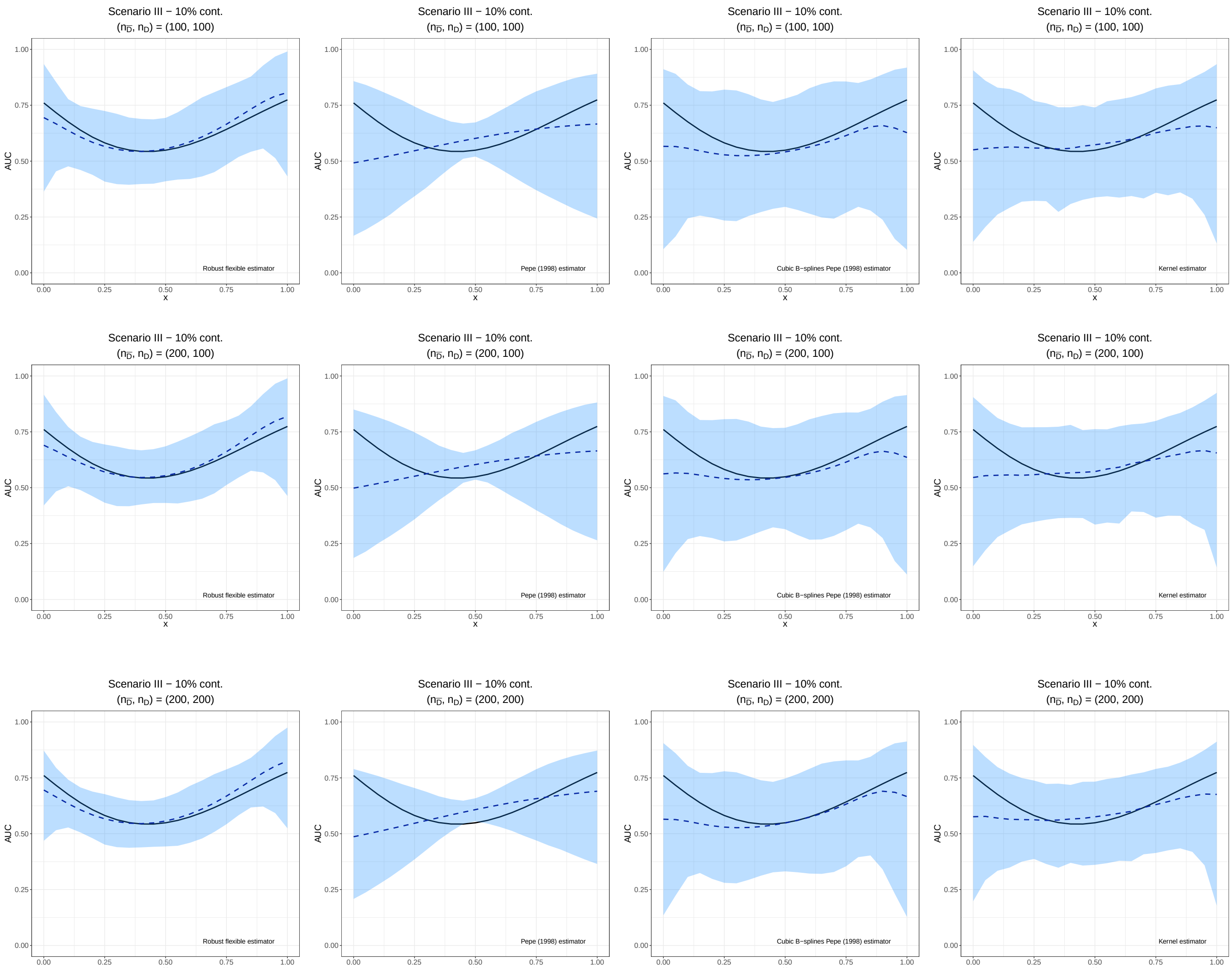


Figure 14: Scenario III. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 10% of contamination. The first row displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, the second row for $(n_{\bar{D}}, n_D) = (200, 100)$, and the third row for $(n_{\bar{D}}, n_D) = (200, 200)$. The first column corresponds to our flexible and robust estimator, the second column to the estimator proposed by Pepe (1998), the third one to the cubic B-splines extension of Pepe (1998), and the fourth column to the kernel estimator.

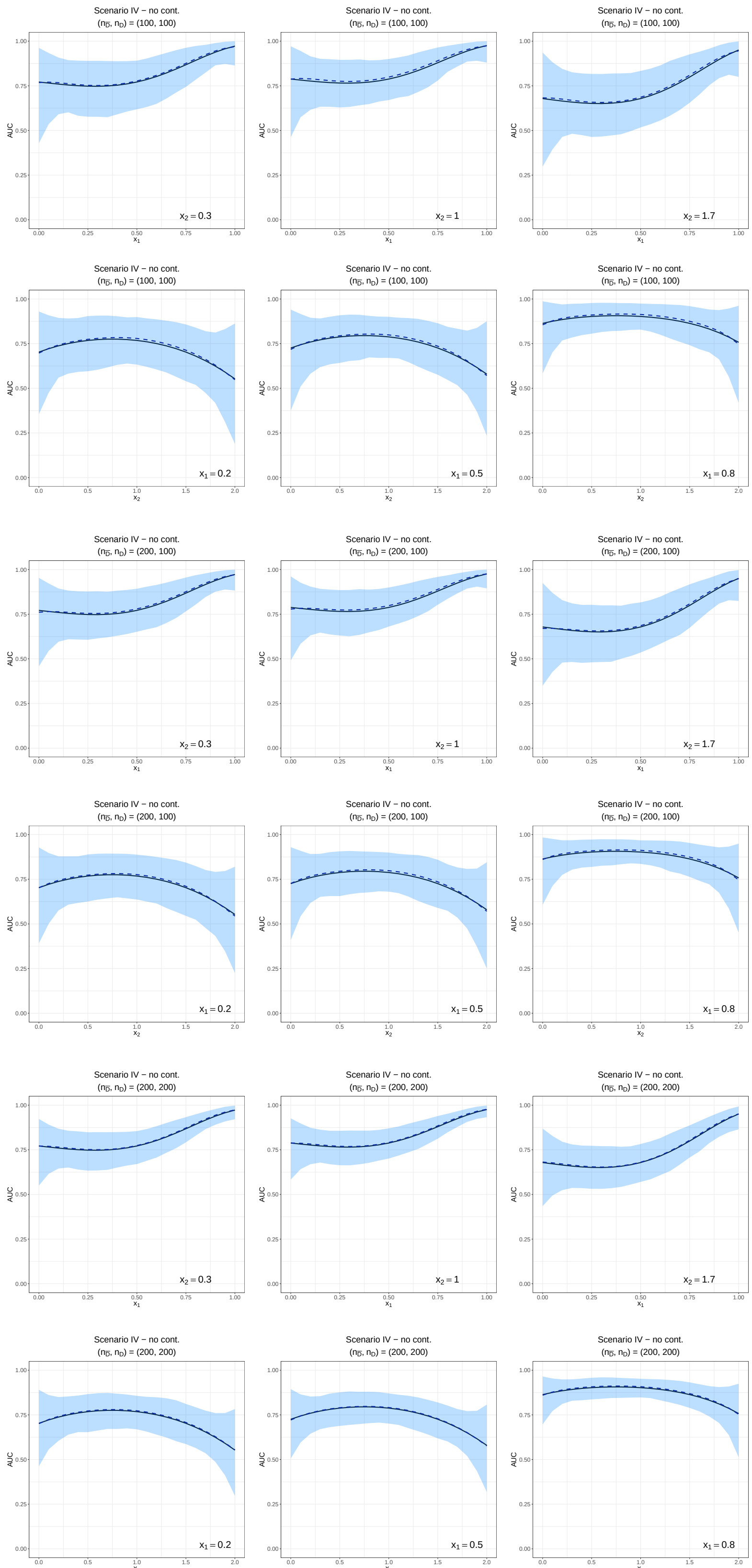


Figure 15: Scenario IV. Multiple profiles of the true covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of no contamination. Rows 1 and 2 displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, rows 3 and 4 for $(n_{\bar{D}}, n_D) = (200, 100)$, and rows 5 and 6 for $(n_{\bar{D}}, n_D) = (200, 200)$.

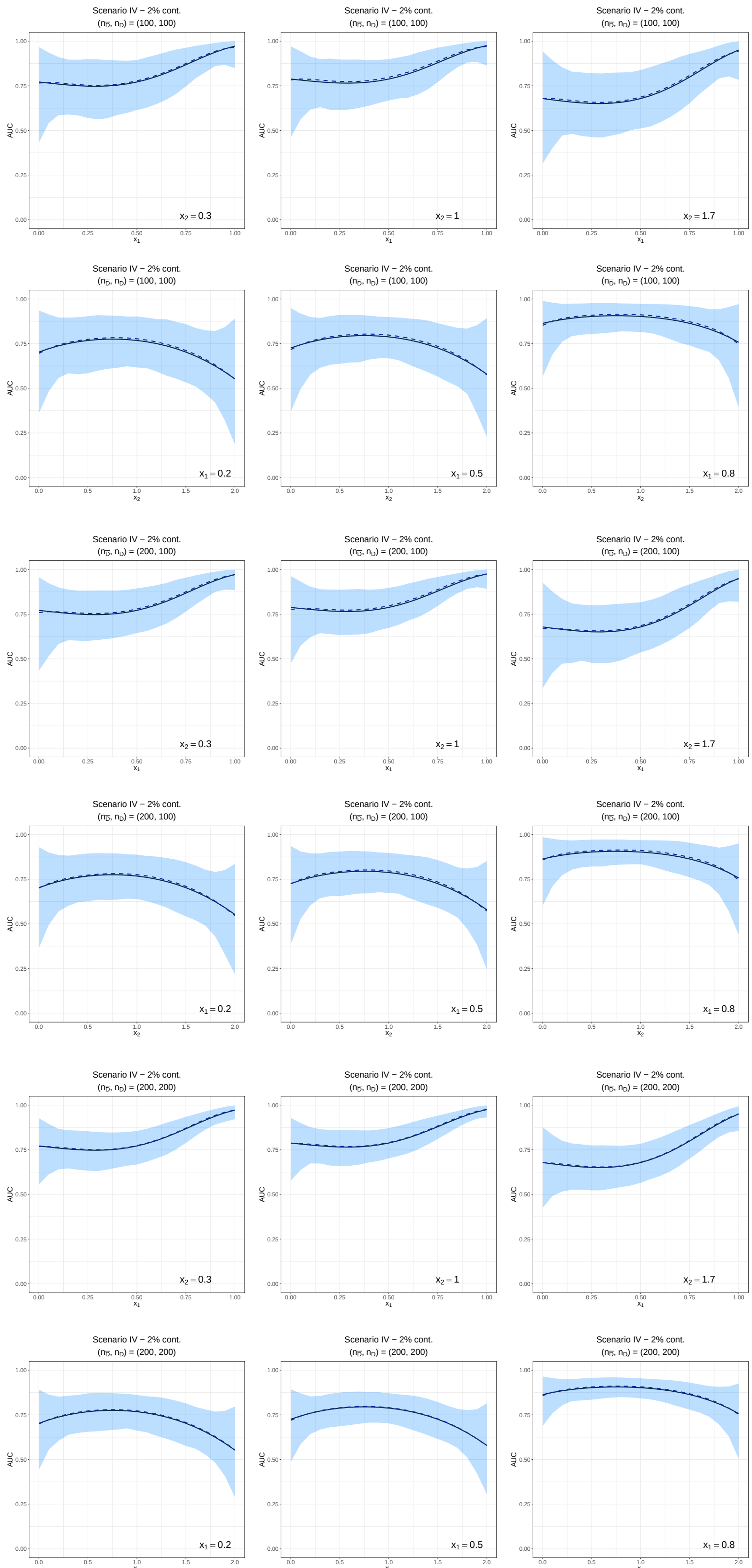


Figure 16: Scenario IV. Multiple profiles of the true covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 2% of contamination. Rows 1 and 2 displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, rows 3 and 4 for $(n_{\bar{D}}, n_D) = (200, 100)$, and rows 5 and 6 for $(n_{\bar{D}}, n_D) = (200, 200)$.

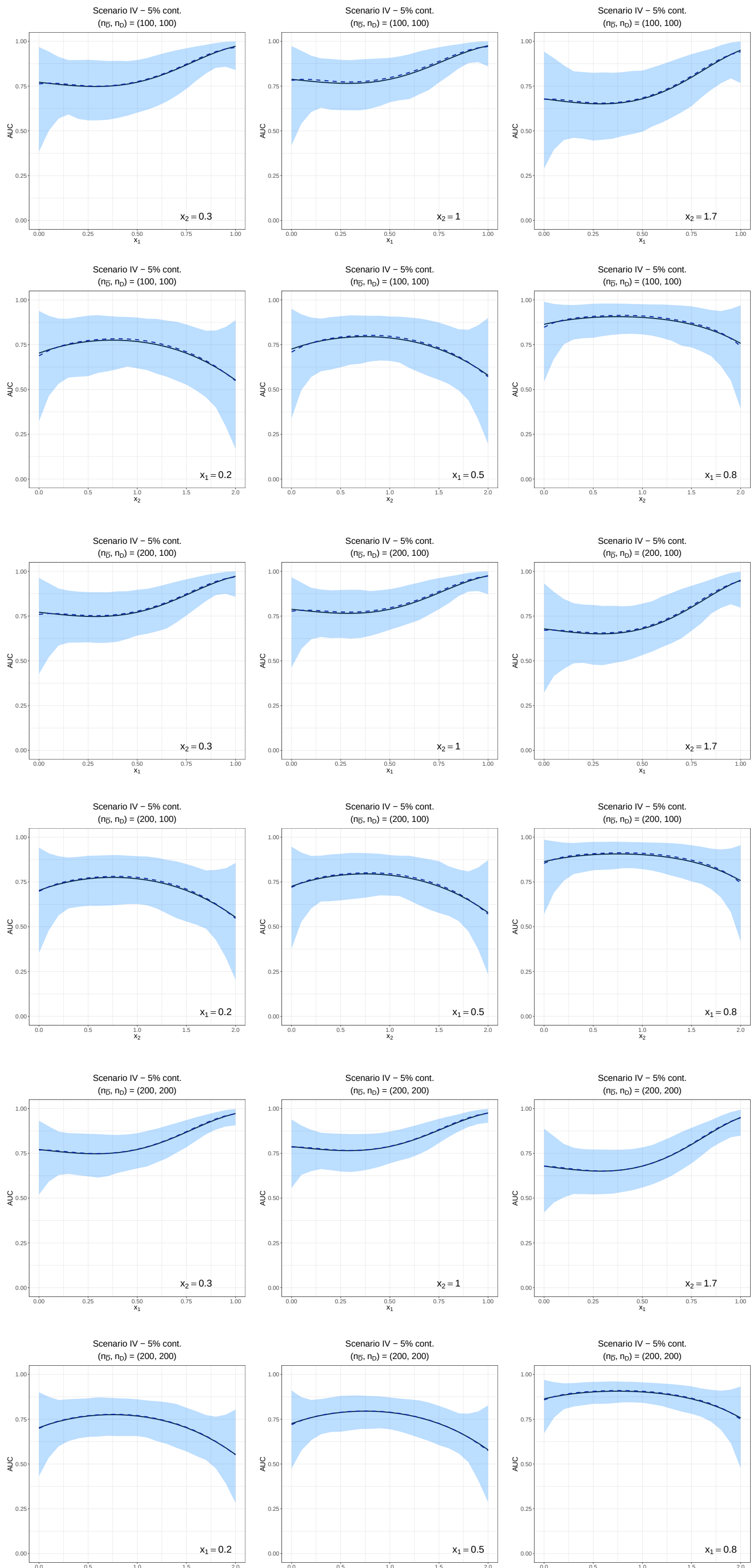


Figure 17: Scenario IV. Multiple profiles of the true covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 5% of contamination. Rows 1 and 2 displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, rows 3 and 4 for $(n_{\bar{D}}, n_D) = (200, 100)$, and rows 5 and 6 for $(n_{\bar{D}}, n_D) = (200, 200)$.

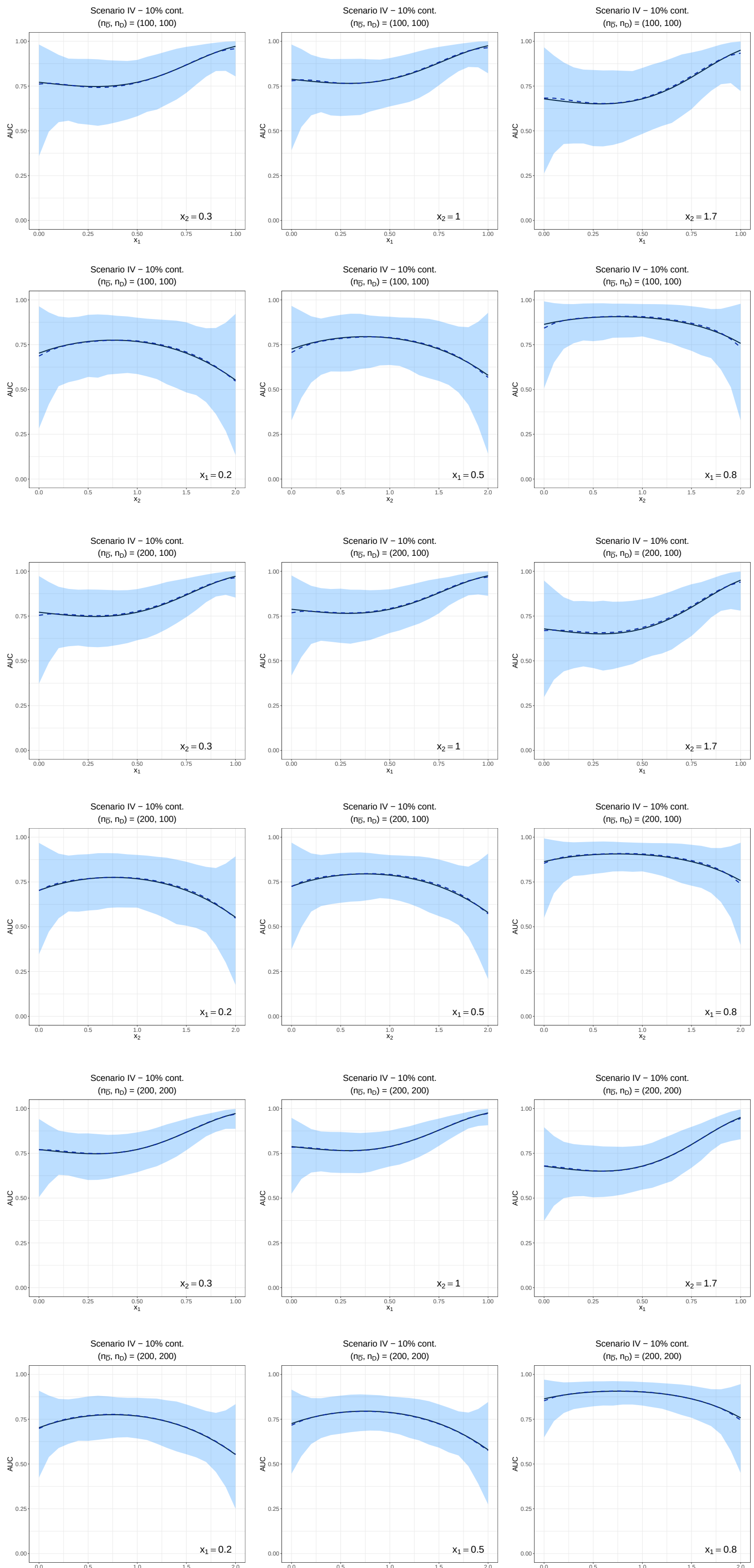


Figure 18: Scenario IV. Multiple profiles of the true covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area) for the case of 10% of contamination. Rows 1 and 2 displays the results for $(n_{\bar{D}}, n_D) = (100, 100)$, rows 3 and 4 for $(n_{\bar{D}}, n_D) = (200, 100)$, and rows 5 and 6 for $(n_{\bar{D}}, n_D) = (200, 200)$.

		Sample size $(n_{\bar{D}}, n_D)$	
Scenario		$(100, 100)$	$(200, 200)$
No contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	70.1	67.4
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	70.6	69.2
2% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	70.7	67.9
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	72.3	70.3
5% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	72.6	69.5
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	75.1	71.4
10% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	80.3	74.1
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	75.8	72.5

Table 1: Scenario I. Percentage over the 1000 simulation runs that the robust AIC favours the robust and flexible model with no interior knots over the same model but with three interior knots.

		Sample size $(n_{\bar{D}}, n_D)$	
Scenario		$(100, 100)$	$(200, 200)$
No contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	69.5	66.4
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	70.6	69.2
2% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	68.9	69.4
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	72.3	70.3
5% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	73.9	70.0
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	75.1	71.4
10% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	78.9	74.4
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	75.8	72.5

Table 2: Scenario II. Percentage over the 1000 simulation runs that the robust AIC favours the robust and flexible model with no interior knots over the same model but with three interior knots.

Scenario		Sample size $(n_{\bar{D}}, n_D)$ (100, 100)	(200, 200)
No contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	75.2	68.5
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	69.2	70.2
2% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	74.1	69.8
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	71.6	68.7
5% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	74.5	70.6
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	75.2	72.8
10% contamination	$\text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 0) < \text{rAIC}_{\bar{D}}(K_{\bar{D}1} = 3)$	79.2	72.7
	$\text{rAIC}_{D}(K_{D1} = 0) < \text{rAIC}_{D}(K_{D1} = 3)$	78.2	73.2

Table 3: Scenario III. Percentage over the 1000 simulation runs that the robust AIC favours the robust and flexible model with no interior knots over the same model but with three interior knots.

Scenario		Sample size $(n_{\bar{D}}, n_D)$ (100, 100)	(200, 200)
No contamination	$\text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (0,0)) < \text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (3,3))$	82.1	77.8
	$\text{rAIC}_{D}(K_{D1}, K_{D2}) = (0,0)) < \text{rAIC}_{D}((K_{D1}, K_{D2}) = (3,3))$	83.3	78.2
2% contamination	$\text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (0,0)) < \text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (3,3))$	84.4	79.2
	$\text{rAIC}_{D}(K_{D1}, K_{D2}) = (0,0)) < \text{rAIC}_{D}((K_{D1}, K_{D2}) = (3,3))$	86.0	78.7
5% contamination	$\text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (0,0)) < \text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (3,3))$	85.8	82.0
	$\text{rAIC}_{D}(K_{D1}, K_{D2}) = (0,0)) < \text{rAIC}_{D}((K_{D1}, K_{D2}) = (3,3))$	87.9	80.8
10% contamination	$\text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (0,0)) < \text{rAIC}_{\bar{D}}((K_{\bar{D}1}, K_{\bar{D}2}) = (3,3))$	91.3	84.7
	$\text{rAIC}_{D}(K_{D1}, K_{D2}) = (0,0)) < \text{rAIC}_{D}((K_{D1}, K_{D2}) = (3,3))$	89.4	85.7

Table 4: Scenario IV. Percentage over the 1000 simulation runs that the robust AIC favours the robust and flexible model with no interior knots over the same model but with three interior knots.

Further simulation scenarios

Scenarios description

We replicate simulation Scenarios I and II from the main paper but now under different distributional error assumptions. Specifically, we consider

- **Scenario V**: the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario I but with $\varepsilon_D \sim 0.5\mathrm{N}(1, 1^2) + 0.5\mathrm{N}(7, 1^2)$ and $\varepsilon_{\bar{D}} \sim 0.5\mathrm{N}(-3, 1^2) + 0.5\mathrm{N}(3, 1^2)$.

- **Scenario VI**:the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario II but with the same error distribution as in Scenario V.

- **Scenario VII**: the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario I but now the error term in the two populations follow a Weibull distribution with scale and shape parameters both equal to 2. This leads to a moderate skewness of around 0.6.

- **Scenario VIII**: the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario II but with the same error distribution as in Scenario VII.

- **Scenario IX**: the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario I but now the error term in the two populations follow a Weibull distribution with scale parameter equal to 1.5 and shape parameter equal to 1. This error configuration leads to a skewness of around 2.

- **Scenario X**: the same mean functions and scale parameters in the diseased and nondiseased populations as in Scenario II but with the same error distribution as in Scenario IX.

To meet the model assumptions the errors were standardised to have mean 0 and variance 1. Finally, in Scenario XI the regression models in each population are as follows:

$$
\begin{aligned}
&y_{\bar{D}i} = \sin(2.75\pi x_{\bar{D}i}) + \varepsilon_{\bar{D}i}, \quad y_{Dj} = 1.25 + \sin(1.25\pi(x_{Dj} + 2)) + 1.5\varepsilon_{Dj},\\
&x_{\bar{D}i} \overset{\text{iid}}{\sim} \mathrm{U}(-1, 1), \quad x_{Dj} \overset{\text{iid}}{\sim} \mathrm{U}(-1, 1), \quad \varepsilon_{\bar{D}i} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1), \quad \varepsilon_{Dj} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1),
\end{aligned}
$$

for $i = 1, \ldots, n_{\bar{D}}$ and $j = 1, \ldots, n_D$. The same sample sizes of the previous scenarios were considered.

Results

As can be observed from Figures 19 and 20, in Scenarios V and VI where the error term follows a two-component (symmetric) mixture of normal distributions, our estimator is able to recover the true functional form of the covariate-specific AUC for all percentages of contamination and sample sizes considered.

In Figures 21 and 22 we show the results for Scenarios VII and VIII where a Weibull distribution with scale and shape parameters both equal to 2 was considered for the error term in the two populations. In both cases we used $v_{\bar{D}} = v_D = 3$ as in all previous scenarios. In this case, and especially in the linear scenario, we can observe some bias in the estimates of the covariate-specific AUC. For larger sample sizes than those shown here, the true true conditional AUC curve may fall slightly outside the band formed by the 2.5% and 97.5% simulation quantiles. We shall note that, and again especially in Scenario VII, it may seem that the bias decreases and the contamination percentage increases. However, we believe this is only due to the way outlying test outcomes were generated: a given percentage of 'clean' test outcomes, from each population, is randomly selected and the corresponding mean shift is applied to such subset of test results and therefore, as the contamination percentage increases, the skewness tends to be less marked. This should however be no problem when working with real data.

In the case of Scenarios IX and X, where a Weibull distribution with scale parameter equal to 1.5 and shape parameter equal to 1 was considered, therefore leading to a highly skewed error distribution, $v_{\bar{D}} = v_D = 3$ would lead to considerable biased estimates of the covariate-specific AUC (results not shown). By inspecting the histogram of the standardised residuals. the choice of $v_{\bar{D}} = v_D = 4$ was deemed as reasonable and this is indeed supported by the results presented in Figures 23 and 24. We note that although we have considered $v_{\bar{D}}$ and v_D to be the same across all contamination percentages, in order to run the simulation study in a little more automatic way, this might not always be the optimal choice for the reasons explained above (skewness getting milder with increased contamination percentages).

With regard to the performance of the robust AIC when the error distribution is skewed, we found its performance to still be good for the case of moderately skewed errors. In the case of highly skewed error distributions, we found the performance of the robust AIC to slightly deteriorate.

Finally, in Scenario XI, whose purpose is to check whether a few number of knots suffice to accurately recover the true functional form of the AUC when it shows a marked nonlinearity, we have fitted the model in each population considering a number of interior knots ranging from 0 to 4. Tables 5 and 6 show, for the cases of $(n_{\bar{D}}, n_D) = (100, 100)$ and $(n_{\bar{D}}, n_D) = (200, 200)$, the percentage over the 1000 simulation of runs each interior knots number was preferred. As can be observed from these tables, in the nondiseased population, $K_{\bar{D}} = 4$ was clearly preferred. In turn, in the diseased population and for $n_D = 100$, $K_D = 0$ and $K_D = 1$ are more or less equally likely. For the case of $n_D = 200$, results show a stronger preference for $K_D = 1$. Results under both choices of K_D are presented in Figures 25 and 26 and as can be noticed, $K_D = 1$ leads to less biased results of the covariate-specific AUC than $K_D = 0$, although this latter value still produced very decent estimates of the conditional AUC.

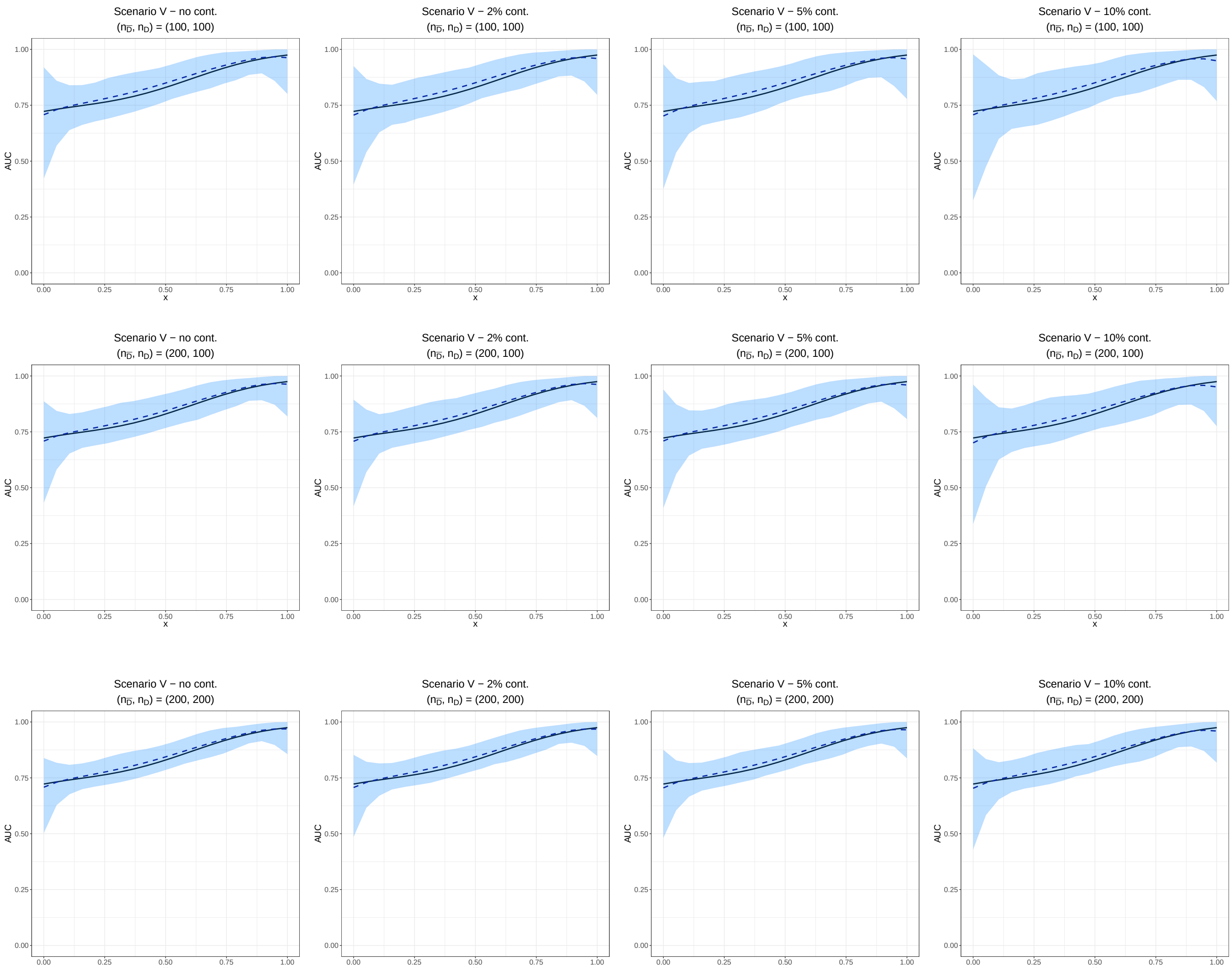


Figure 19: Scenario V. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

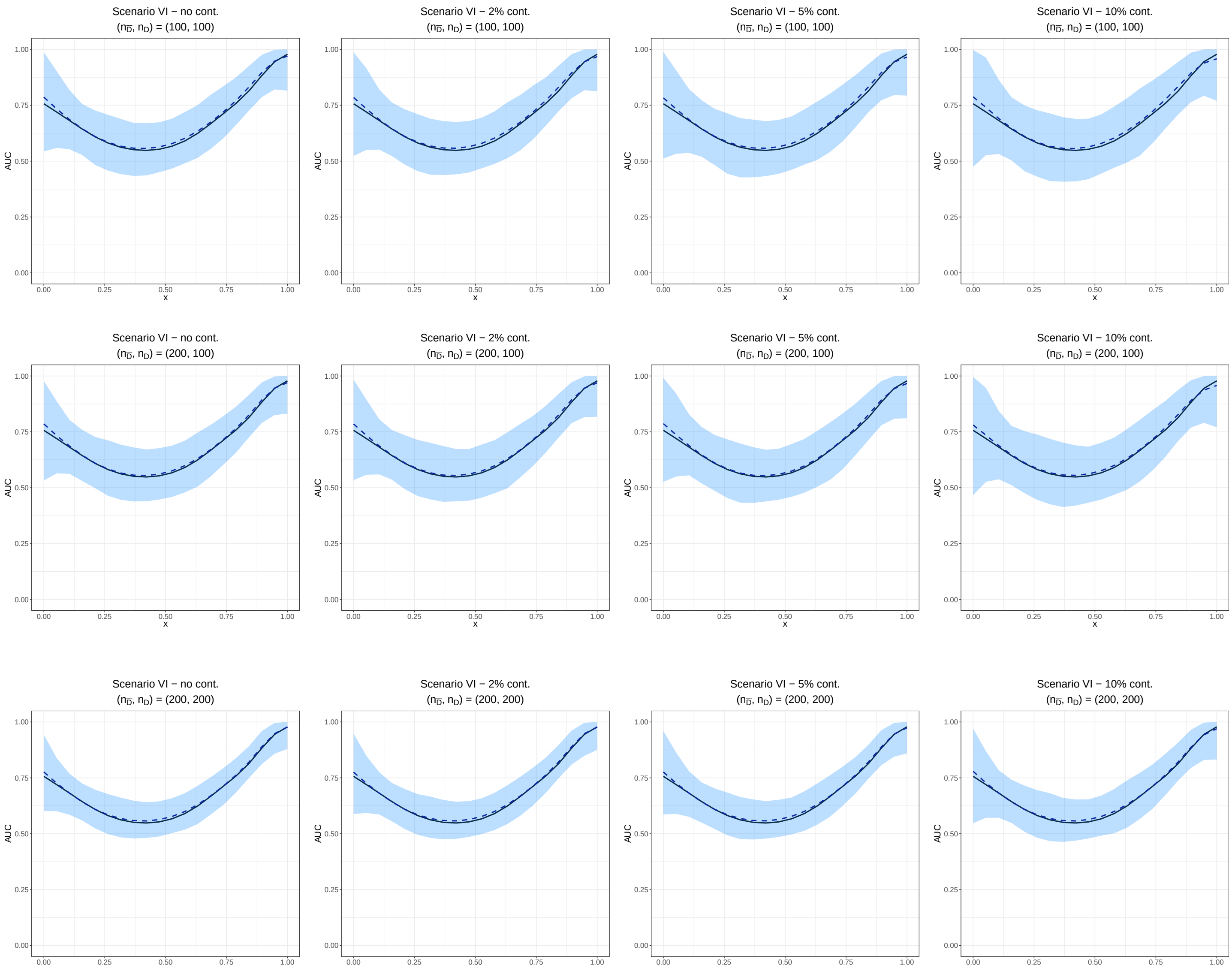


Figure 20: Scenario VI.True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

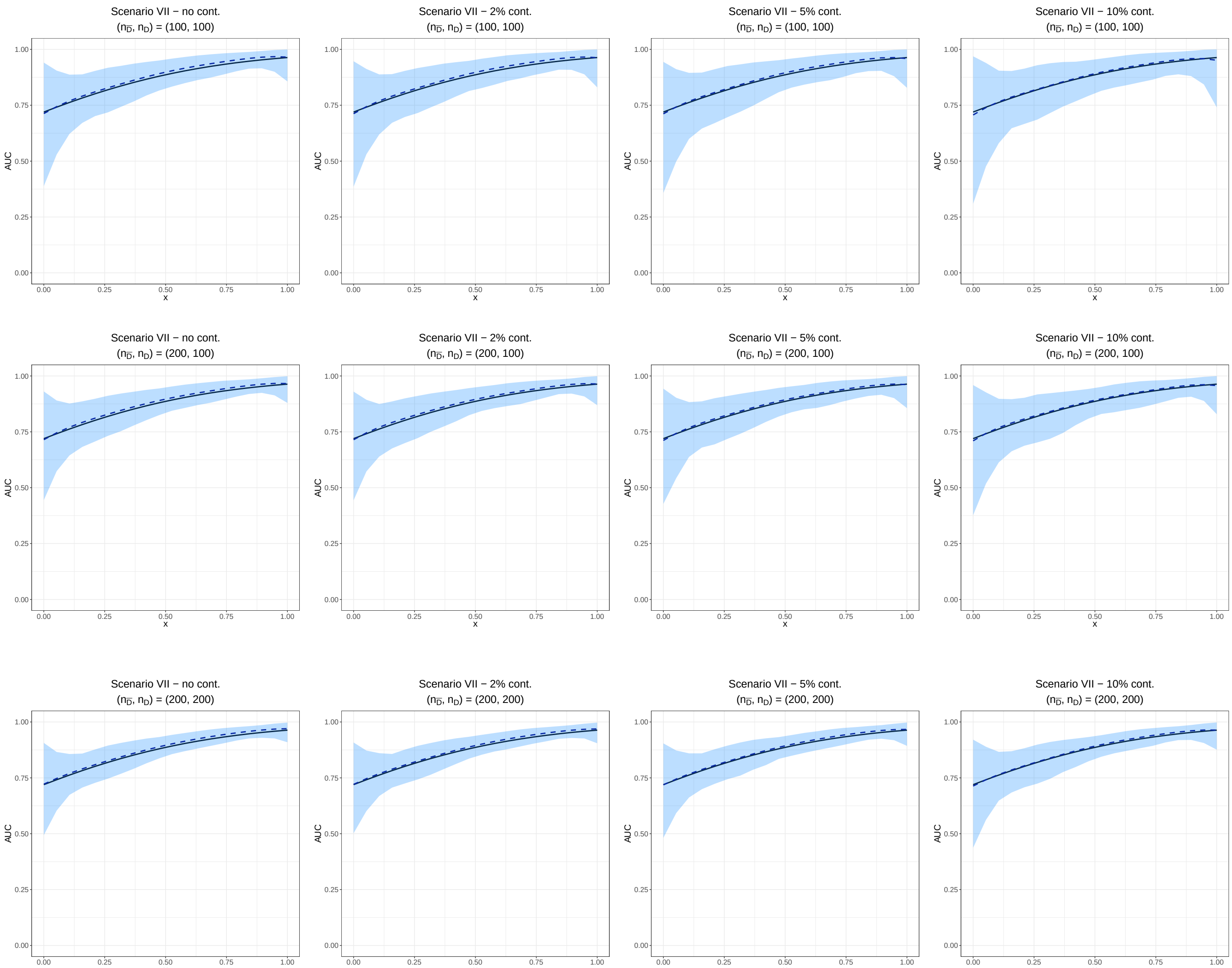


Figure 21: Scenario VII. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

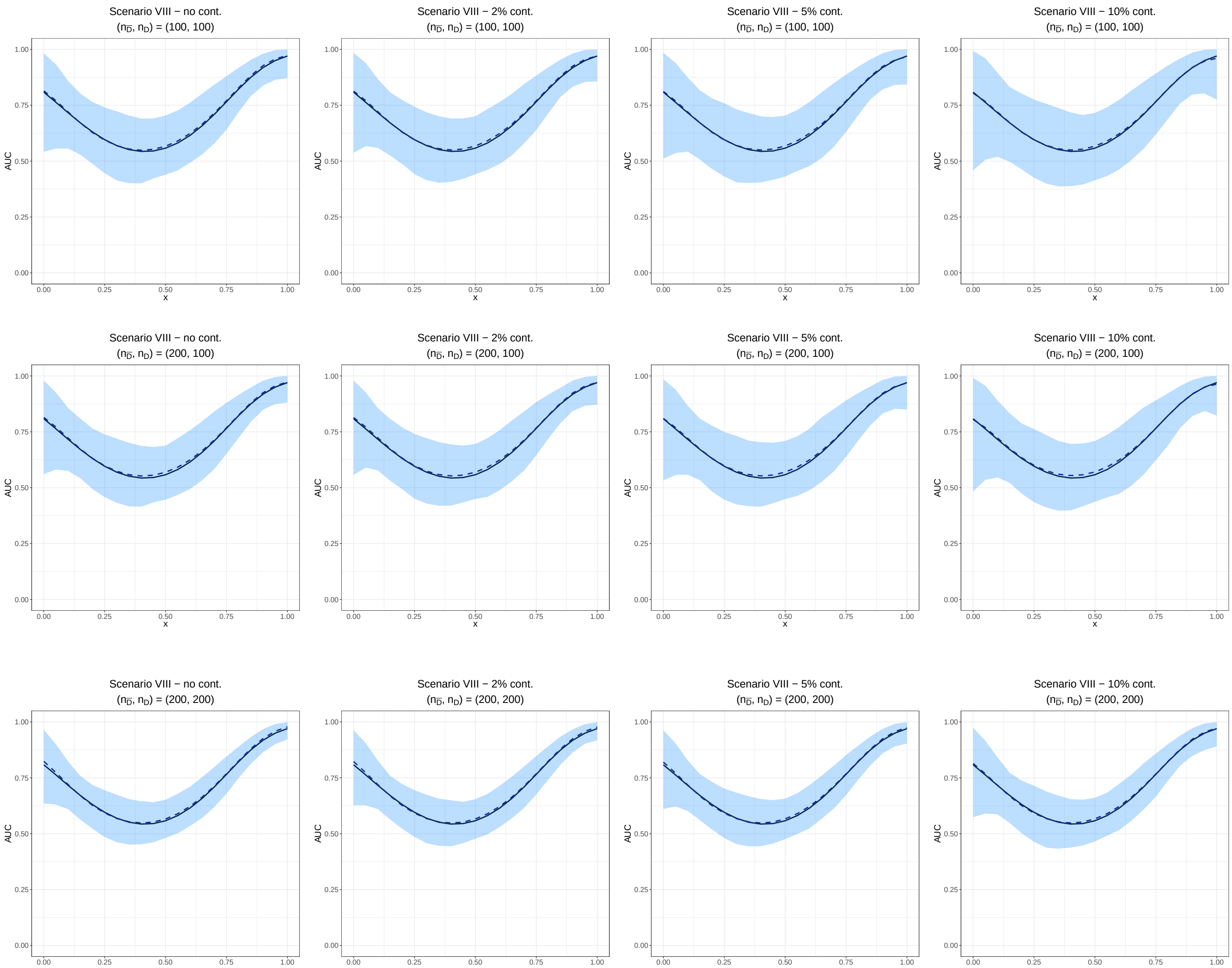


Figure 22: Scenario VIII. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

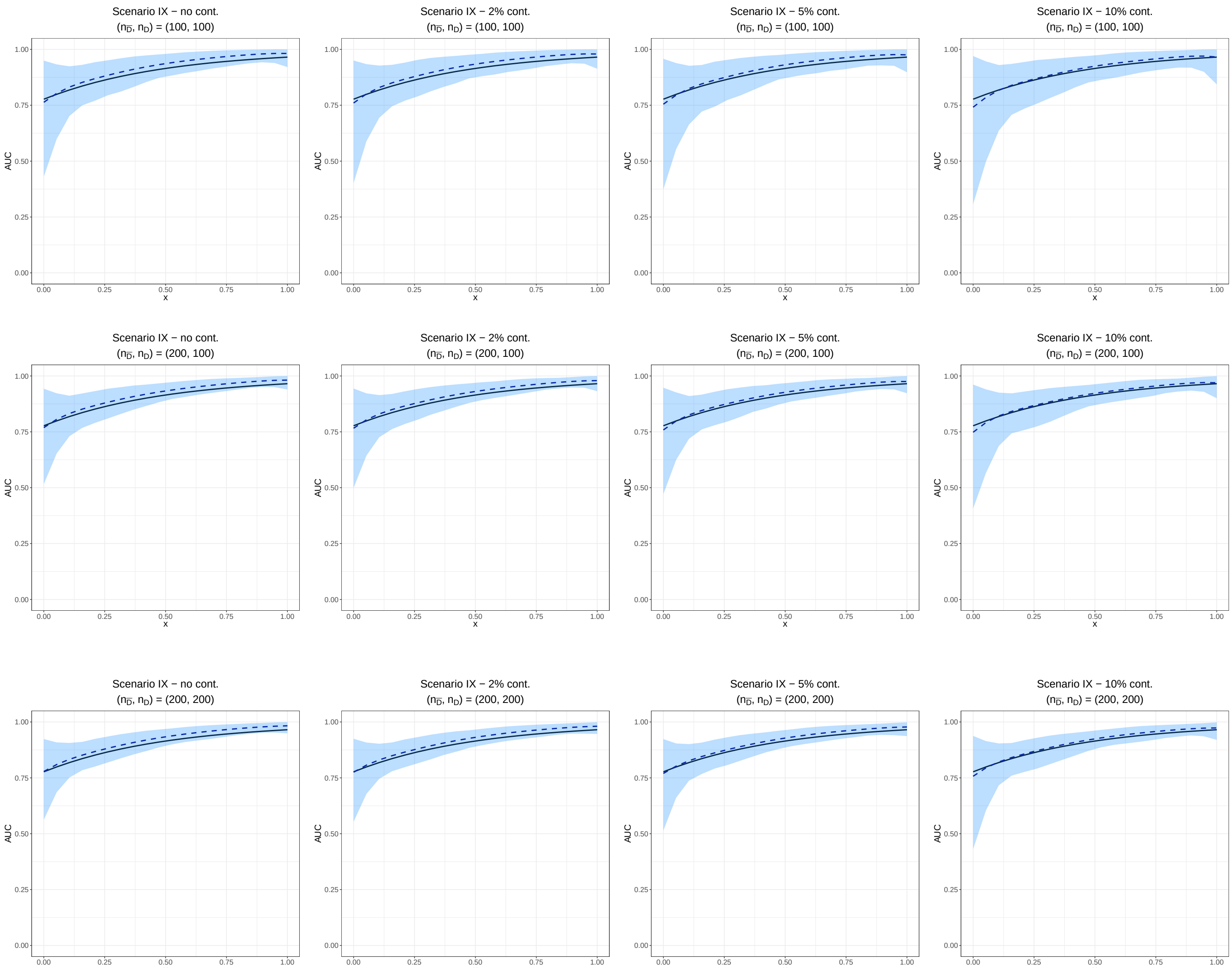


Figure 23: Scenario IX. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

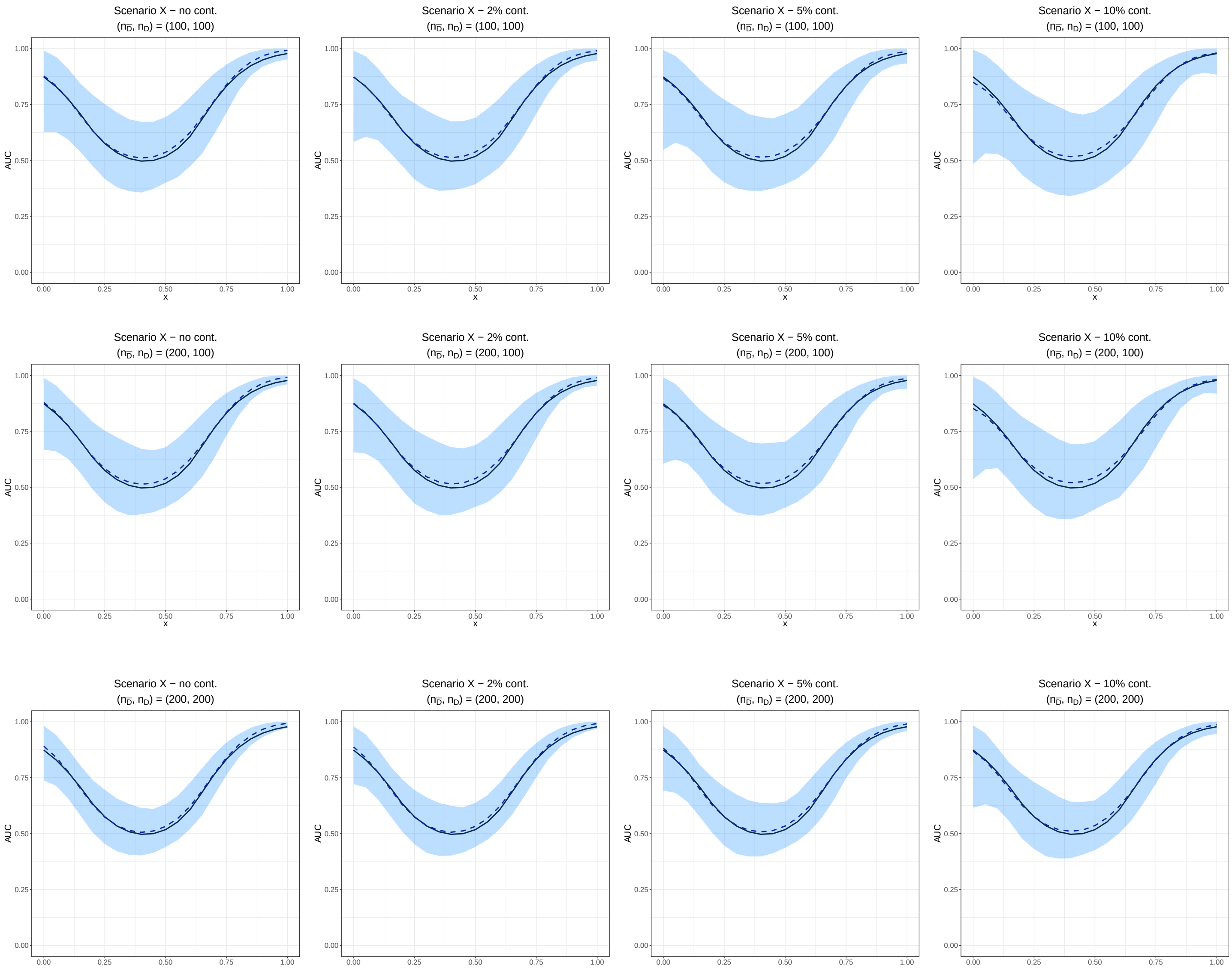


Figure 24: Scenario X. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

rAIC					
	$K_{\bar{D}}$				
	0	1	2	3	4
No contamination	12.2	5.2	9.7	8.4	64.5
2% contamination	10.9	5.9	11.3	8.9	63.0
5% contamination	11.5	5.7	11.2	8.2	63.4
10% contamination	16.6	6.0	11.3	7.1	59.0
	K_D				
	0	1	2	3	4
No contamination	29.8	27.8	19.1	12.0	11.3
2% contamination	30.5	30.1	18.2	11.3	9.9
5% contamination	32.8	32.0	15.4	10.4	9.4
10% contamination	37.7	30.3	13.8	11.0	7.2

Table 5: Scenario XI. Percentage over the 1000 simulation runs that the model with each of the referred number of knots was preferred for the sample sizes $(n_{\bar{D}}, n_D) = (100, 100)$.

rAIC					
	$K_{\bar{D}}$				
	0	1	2	3	4
No contamination	3.0	1.3	4.5	2.6	88.6
2% contamination	3.2	1.1	4.5	3.2	88.0
5% contamination	3.4	1.8	3.8	4.0	87.0
10% contamination	4.2	1.9	5.3	2.5	86.1
	K_D				
	0	1	2	3	4
No contamination	22.3	30.7	19.8	13.2	14.0
2% contamination	24.4	31.6	20.4	10.9	12.7
5% contamination	24.8	30.0	19.5	13.3	12.4
10% contamination	23.4	35.5	18.8	12.3	10.0

Table 6: Scenario XI. Percentage over the 1000 simulation runs that the model with each of the referred number of knots was preferred for the sample sizes $(n_{\bar{D}}, n_D) = (200, 200)$.

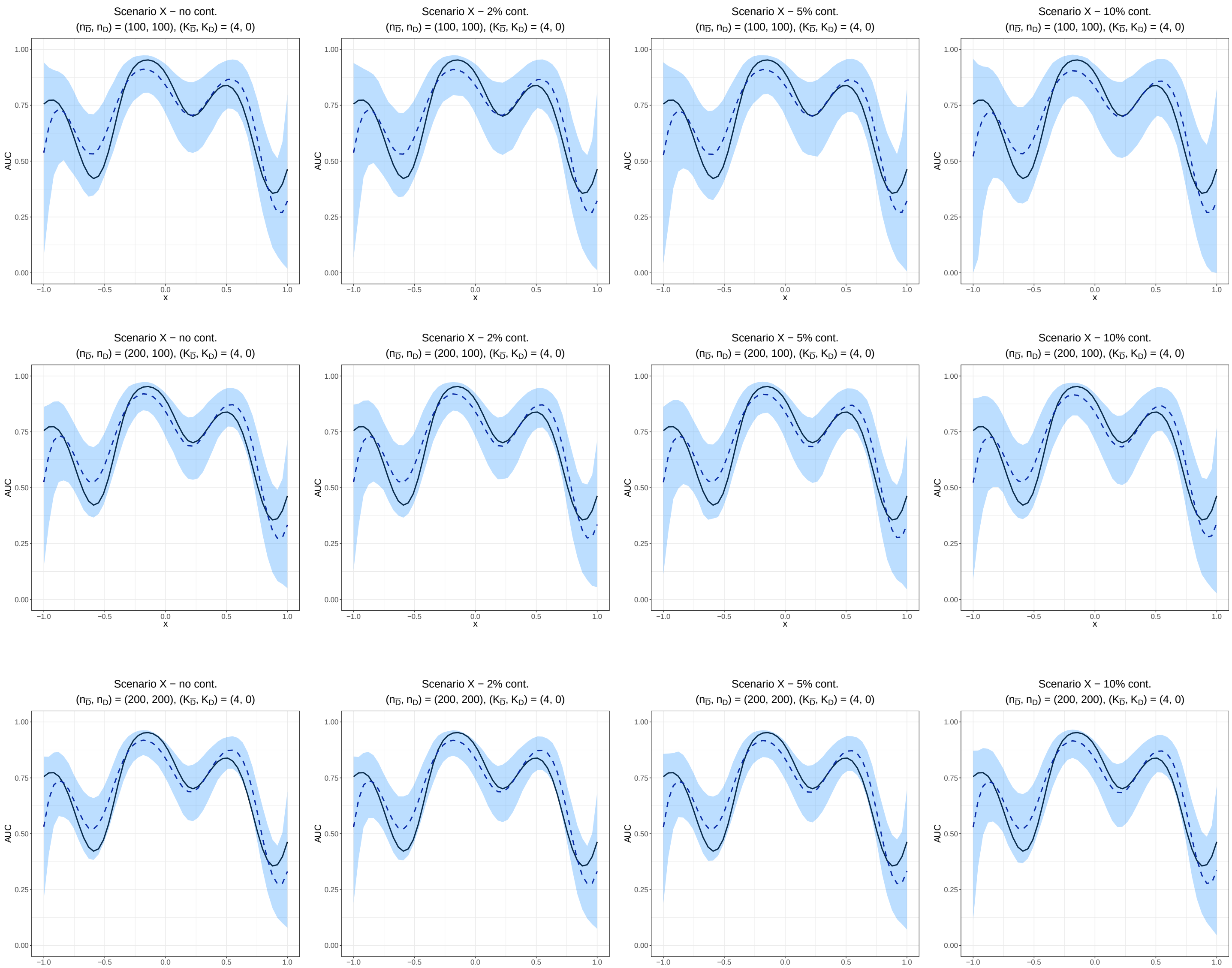


Figure 25: Scenario XI and $(K_{\bar{D}}, K_D) = (4, 0)$. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

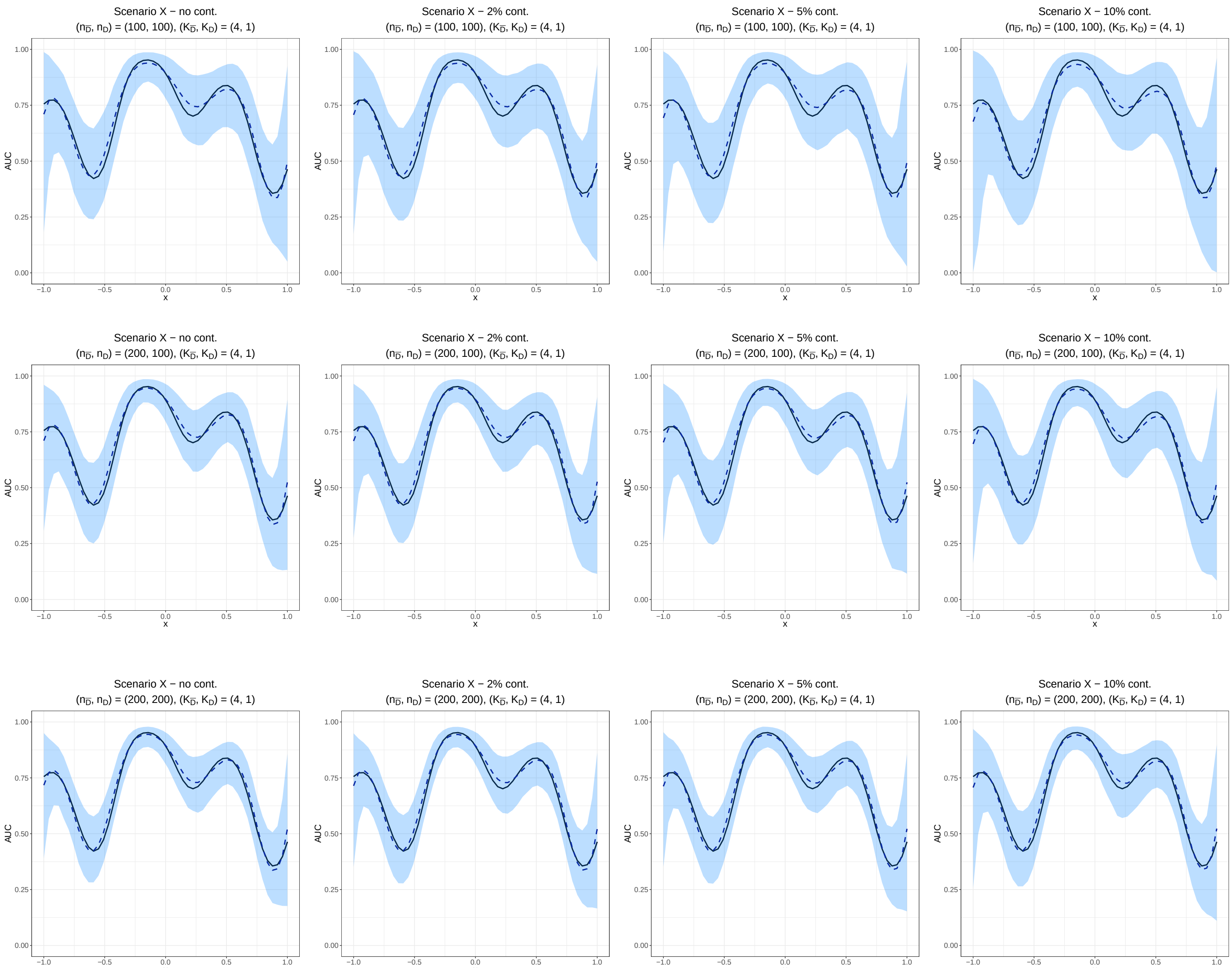


Figure 26: Scenario XI and $(K_{\bar{D}}, K_D) = (4, 1)$. True covariate-specific AUC (solid line) versus the mean of the Monte Carlo estimates (dashed line) along with the 2.5% and 97.5% simulation quantiles (shaded area). First row: $(n_{\bar{D}}, n_D) = (100, 100)$. Second row: $(n_{\bar{D}}, n_D) = (200, 100)$. Third row: $(n_{\bar{D}}, n_D) = (200, 200)$. First column: no contamination. Second column: 2% contamination. Third column: 5% contamination. Fourth column: 10% contamination.

Additional figures for the Application

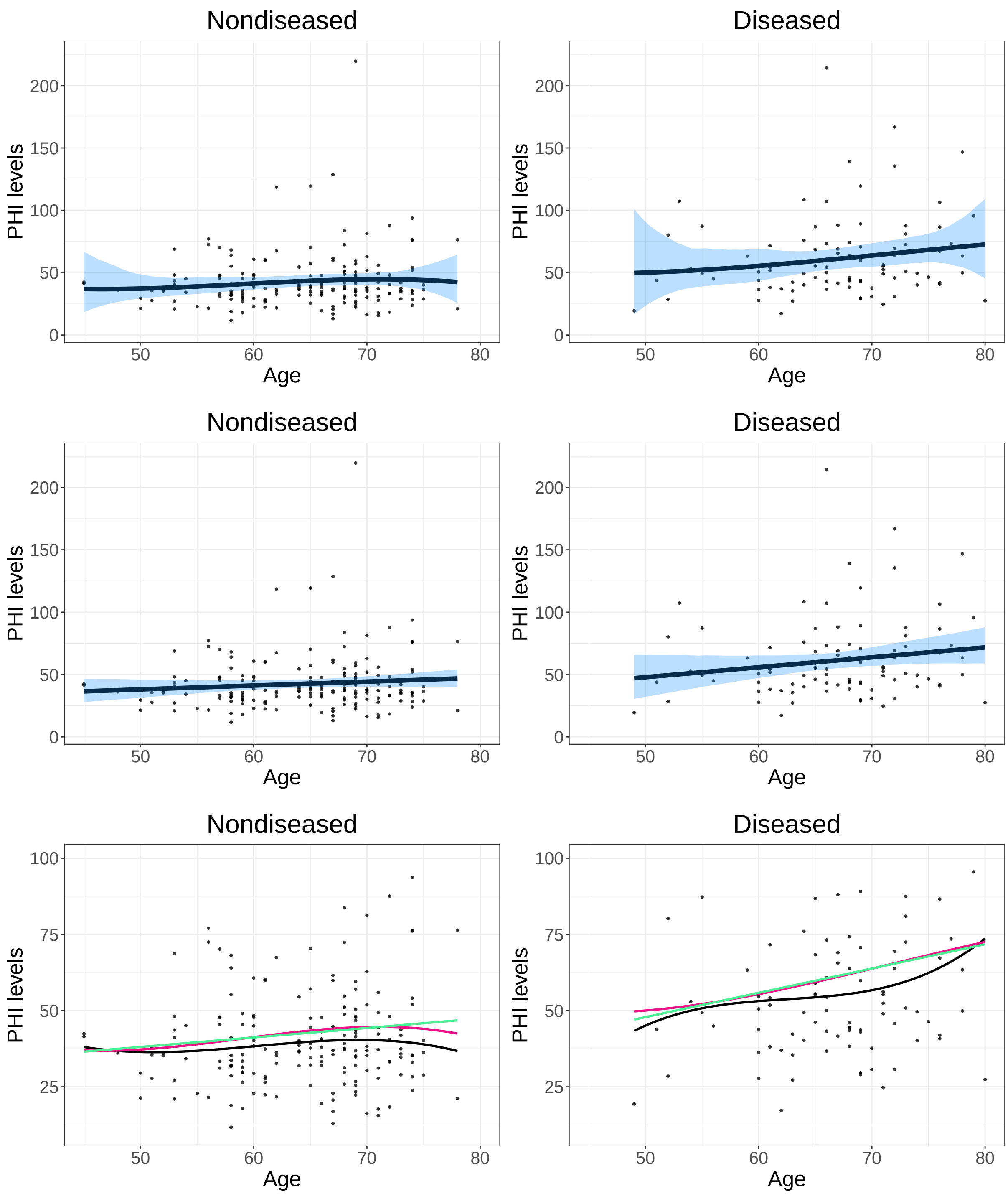


Figure 27: First row: Regression functions resulting from fitting a linear model with a cubic B-splines basis expansion (no interior knots) for the mean function. Second row: Regression functions resulting from fitting a linear model. Third row: Point estimates from the three different fits, where the black line is the point estimate from the robust flexible model, the pink line is the point estimate corresponding to the B-splines linear model and the light green line is the estimate from the linear model. Note that for a better visualization the y axis has been restricted to the range $(10, 100)$. The shaded areas represent the 95% pointwise bootstrap confidence bands (based on 1000 resamples).

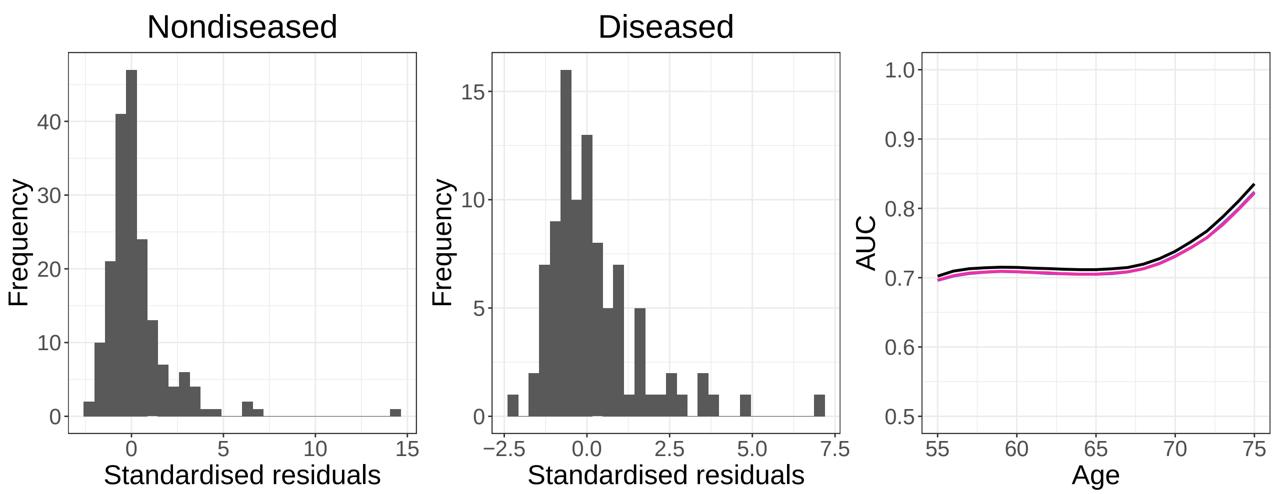
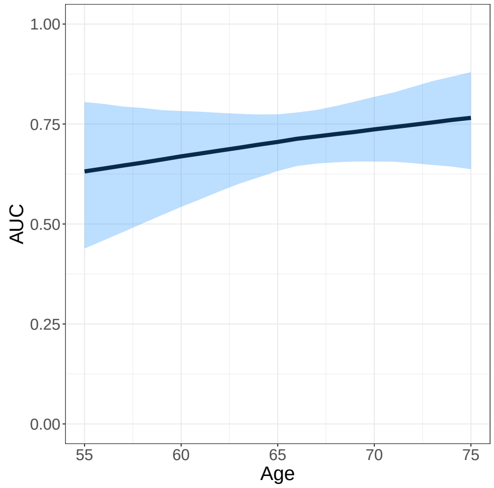
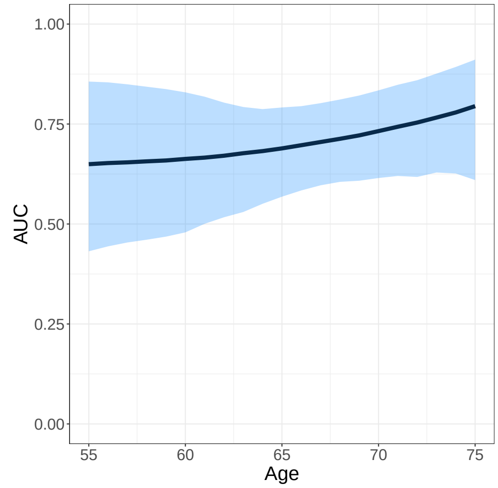


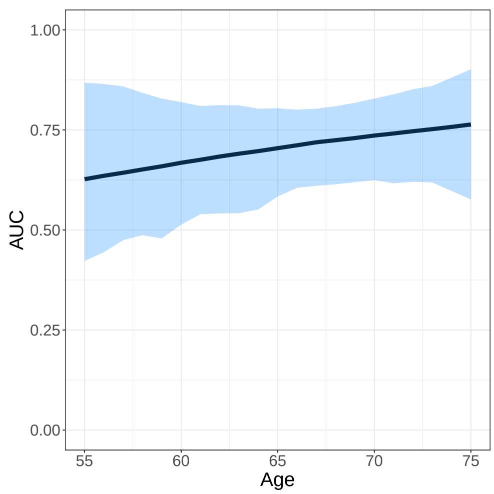
Figure 28: Histogram of the standardised residuals in the nondiseased (left) and diseased populations (right). Age-specific AUC for the case of $v_{\bar{D}} = v_D = 3$ (black line), $v_{\bar{D}} = v_D = 4$ (blue line) and $v_{\bar{D}} = v_D = 5$ (maroon line). The latter two curves are indistinguishable. The y limit ranging from 0.5 to 1 is used to facilitate comparison between the three age-specific AUCs.



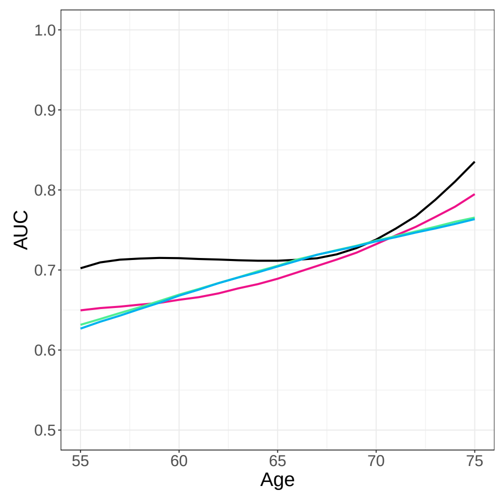
(a)



(b)



(c)



(d)

Figure 29: Age-specific AUCs. (a) Estimate resulting from fitting a linear model in each group. (b) Estimate resulting from fitting a linear model in each group with a cubic B-splines basis expansion for the mean function. (c) Estimate resulting from fitting the kernel approach in each group. The shaded areas represent the 95% pointwise bootstrap confidence bands (based on 1000 resamples). (d) Comparison of point estimates from the different approaches. Black line: our approach. Light green: linear model. Pink Line: linear model with B-splines basis expansion for the mean function. Blue line: kernel method. The green and blue line are indistinguishable.