SUPPLEMENTARY MATERIALS

A.1 Data Generation

The simulations reported in Table 1 are based on data generated as

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X D \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

where case status, D, is a random binary variable with specified $P(D = 1)$. We chose $\mu_X = \sqrt{2}\, \Phi^{-1}(\mathrm{AUC}_X)$ so that the AUC for the baseline model was fixed by design.

Data for Table 2 were simulated as

$$\begin{pmatrix} X \\ Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X D \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right).$$

Data for Table 3 were generated from the distribution

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X D \\ \mu_Y D \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

where $\mu_X = \sqrt{2}\Phi^{-1}(\mathrm{AUC}_X)$ and $\mu_Y = \sqrt{2}\Phi^{-1}(\mathrm{AUC}_Y)$.

A.2 Net Benefit is a Proper Scoring Statistic

Let X denote the available predictors and let $r^*(X)$ and $r(X)$ be two functions of X. Assume that the true model is $P(D = 1|X) = r(X)$ and the aim is to assess the decision rule based on the model $r^*(X)$ defined by $r^*(X) \geq t$. The standardized net benefit statistic at threshold t associated with the function $r^*(X)$ is

$$S\mathrm{NB}(t, r^*(X)) = P(r^*(X) \geq t|D = 1) - \frac{1-\rho}{\rho}\frac{t}{1-t}P(r^*(X) \geq t|D = 0)$$

where $\rho = P(D = 1)$. In terms of expectation over the marginal distribution of X, we write

$$\begin{aligned} \rho S\mathrm{NB}(t, r^*(X)) &= E\left\{I(r^*(X) \geq t, D = 1) - \frac{t}{1-t}I(r^*(X) \geq t, D = 0)\right\} \\ &= E\left\{P(D = 1|X)I(r^*(X) \geq t) - \frac{t}{1-t}[1 - P(D = 1|X)]I(r^*(X) \geq t)\right\} \\ &= E\left\{I(r^*(X) \geq t)[r(X) - \frac{t}{1-t}(1 - r(X))]\right\} \\ &= E\left\{I(r^*(X) \geq t)[\frac{r(X) - t}{1-t}]\right\}. \end{aligned}$$

Now consider the difference between $\rho S\mathrm{NB}(t, r(X))$ and $\rho S\mathrm{NB}(t, r^*(X))$: $\rho(S\mathrm{NB}(t, r(X)) - S\mathrm{NB}(t, r^*(X))) = \frac{1}{1-t} E\{(r(X)-t)(I(r(X) \geq t) - I(r^*(X)) \geq t))\}$

The entity inside the expectation, namely

$$(r(X) - t)\{I(r(X) \geq t) - I(r^*(X) \geq t)\} \tag{A.1}$$

is non-negative with probability 1. To see this, consider the various cases possible: (i) $r(X) = t$; (ii) $r(X) > t, r^*(X) \geq t$; (iii) $r(X) > t, r^*(X) < t$ (iv) $r(X) < t, r^*(X) < t$ and (v) $r(X) < t, r^*(X) \geq t$ and observe that A.1 is ≥ 0 in each case. Therefore the expectation is nonnegative. In other words,

$$S\mathrm{NB}(t, r(X))) \geq S\mathrm{NB}(t, r^*(X))$$

for every function $r^*(X)$. That is, $S\mathrm{NB}(t, r^*(X))$ is maximized by $r^*(X) = r(X)$, the true risk function.

A.3 Supplementary Figures and Tables

Table A.1 This table adds standard errors (in parentheses) to the information shown in Table 1.

One Uninformative Marker

Simulation Scenario				Performance Increment ×100 Average (standard error)				
$\rho = P(D = 1)$	AUC_X	N-training	N-test	NRI	ΔROC(0.2)	ΔAUC	ΔBrier	ΔSNB(ρ)
0.1	0.6	250	25,000	0.27(0.09)	−1.70(0.09)	−1.28(0.08)	−0.044(0.002)	−1.85(0.12)
0.1	0.7	250	25,000	1.38(0.16)	−1.37(0.07)	−0.86(0.05)	−0.049(0.002)	−1.31(0.07)
0.1	0.8	250	25,000	3.22(0.28)	−0.90(0.05)	−0.48(0.02)	−0.058(0.003)	−0.80(0.04)
0.1	0.9	250	25,000	7.72(0.52)	−0.52(0.03)	−0.25(0.01)	−0.066(0.003)	−0.57(0.03)
0.5	0.6	50	5,000	0.57(0.15)	−1.67(0.12)	−1.19(0.11)	−0.479(0.023)	−1.69(0.15)
0.5	0.7	50	5,000	2.78(0.28)	−2.59(0.12)	−1.69(0.08)	−0.540(0.024)	−2.49(0.13)
0.5	0.8	50	5,000	6.56(0.47)	−1.83(0.09)	−1.00(0.05)	−0.492(0.022)	−1.62(0.08)
0.5	0.9	50	5,000	17.09(0.91)	−1.11(0.05)	−0.56(0.03)	−0.433(0.021)	−1.17(0.06)

Table A.2 Average values of the NRI (%) in a test set with 5,000 observations (ρ = 0.5) when risk models are fit in training sets of various sizes and with various numbers of covariates. Settings where the ratio of the training set sample size to the number of predictors is fixed at 25 are highlighted. The markers $(Y_1, \ldots, Y_5)$ are not predictive so the NRI= 0 for the true risk models. Results are based on 1,000 simulations per scenario.

AUC_X	Size for Training Data	(X, Y_1)	(X, Y_1, Y_2, Y_3)	$(X, Y_1, Y_2, Y_3, Y_4, Y_5)$
0.70	50	2.3	4.8	6.5
0.70	100	1.4	3.4	4.6
0.70	150	1.3	2.8	3.7
0.80	50	7.0	13.3	17.2
0.80	100	4.4	8.8	11.9
0.80	150	3.5	7.5	10.1

Table A.3 Estimated performance in the same simulation scenarios considered in Table 1. Here we contrast average estimates calculated in the training data with those calculated in the independent validation data.

Simulation Scenario	Average Performance Increment (%)									
Row in Table 1	NRI		ΔROC(0.2)		ΔAUC		ΔBrier		ΔSNB(ρ)	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
1	13.61	0.27	2.26	−1.70	1.81	−1.28	0.041	−0.044	2.55	−1.85
2	14.30	1.38	1.24	−1.37	0.77	−0.86	0.046	−0.049	1.19	−1.31
3	16.16	3.22	1.19	−0.90	0.46	−0.48	0.056	−0.058	0.94	−0.80
4	18.90	7.72	0.54	−0.52	0.26	−0.25	0.059	−0.066	0.62	−0.57
5	17.45	0.57	4.19	−1.67	2.77	−1.19	0.437	−0.479	3.99	−1.69
6	19.80	2.78	3.13	−2.59	1.69	−1.69	0.486	−0.540	2.62	−2.49
7	23.14	6.56	1.78	−1.83	0.94	−1.00	0.426	−0.492	1.18	−1.62
8	31.36	17.09	0.94	−1.11	0.56	−0.56	0.391	−0.433	1.12	−1.17

Fig. A.1 Scatterplots showing the relationship between the ΔROC(0.2) statistic ($\times$100) and $\hat{\beta}_1 - \hat{\alpha}_1$ in 1000 simulated datasets generated according to the scenario shown in the second to last row of Table 1. The coefficients are calculated by fitting the models logit$P(D = 1|X) = \alpha_0 + \alpha_1 X$ and logit$P(D = 1|X, Y) = \beta_0 + \beta_1 X + \beta_2 Y$ to the training data. The ΔROC(0.2) is calculated using the test dataset.

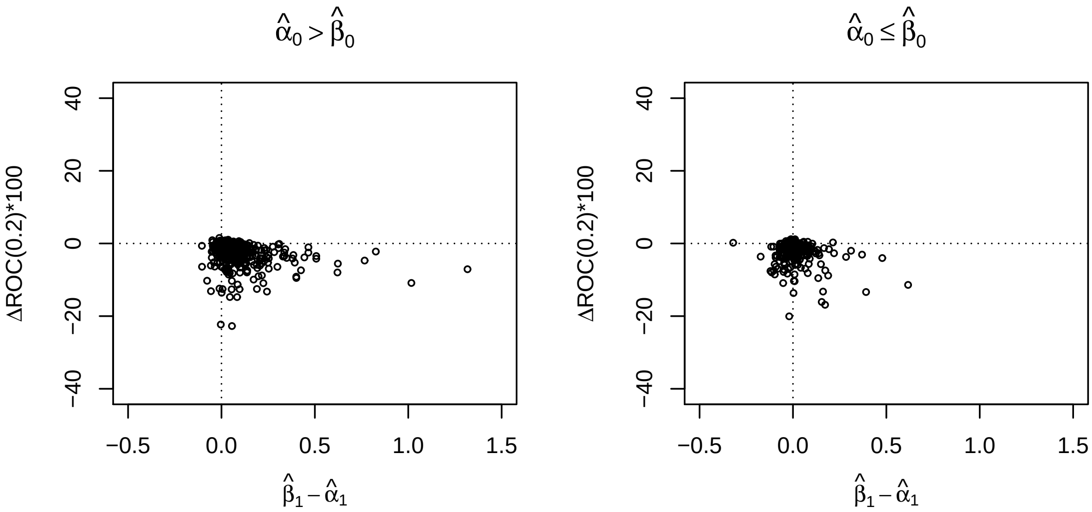

Fig. A.2 Scatterplots showing the relationship between the $\Delta S\mathrm{NB}(\rho)$ statistic ($\times 100$) and $\hat{\beta}_1 - \hat{\alpha}_1$ in 1000 simulated datasets generated according to the scenario shown in the second to last row of Table 1. The coefficients are calculated by fitting the models $\mathrm{logit}P(D = 1|X) = \alpha_0 + \alpha_1 X$ and $\mathrm{logit}P(D = 1|X, Y) = \beta_0 + \beta_1 X + \beta_2 Y$ to the training data. The $\Delta S\mathrm{NB}(\rho)$ is calculated using the test dataset.

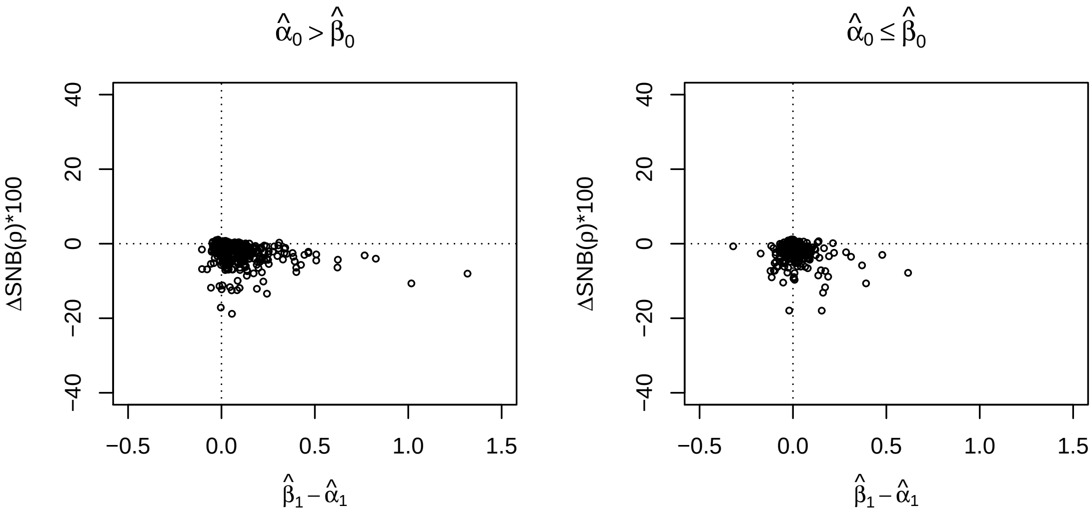

Fig. A.3 Scatterplots showing the relationship between the ΔAUC statistic ($\times 100$) and $\hat{\beta}_1 - \hat{\alpha}_1$ in 1000 simulated datasets generated according to the scenario shown in the second to last row of Table 1. The coefficients are calculated by fitting the models $\text{logit}P(D = 1|X) = \alpha_0 + \alpha_1 X$ and $\text{logit}P(D = 1|X, Y) = \beta_0 + \beta_1 X + \beta_2 Y$ to the training data. The ΔAUC is calculated using the test dataset.

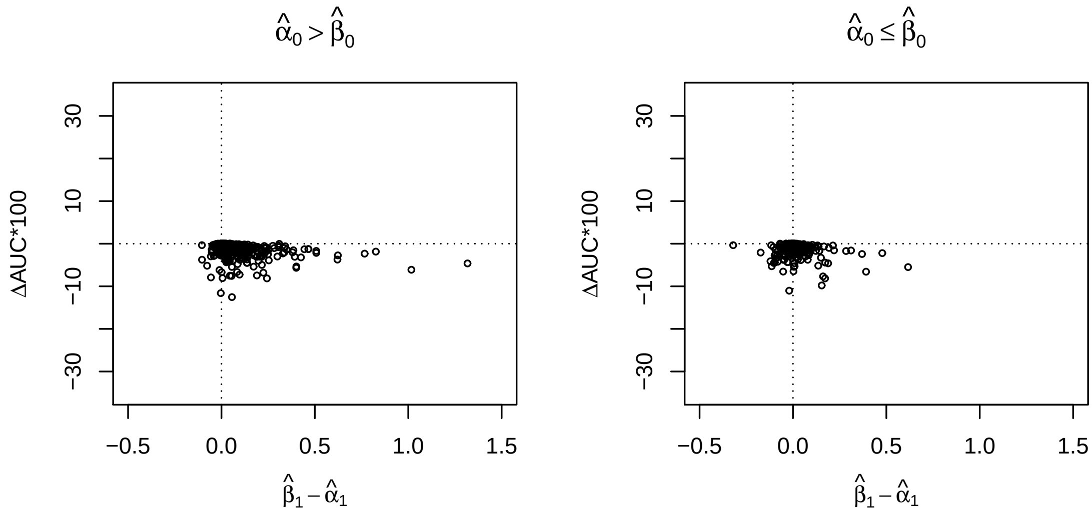

Fig. A.4 Scatterplots showing the relationship between the ΔBrier statistic ($\times 100$) and $\hat{\beta}_1 - \hat{\alpha}_1$ in 1000 simulated datasets generated according to the scenario shown in the second to last row of Table 1. The coefficients are calculated by fitting the models $\text{logit}P(D = 1|X) = \alpha_0 + \alpha_1 X$ and $\text{logit}P(D = 1|X,Y) = \beta_0 + \beta_1 X + \beta_2 Y$ to the training data. The ΔBrier is calculated using the test dataset.