Informing a Risk Prediction Model for Binary Outcomes with External Coefficient Information: Supplementary Materials

Wenting Cheng, Jeremy M. G. Taylor, Tian Gu, Scott A. Tomlins and Bhramar Mukherjee

University of Michigan, Ann Arbor, Michigan, USA

†

1. Web Appendix A

Further Details on the Statistical Methods for Gaussian B

Standard Bayes

Analogous to direct linear regression, we can perform standard Bayesian linear regression with flat conjugate priors for parameters in model (3) (in main text). For model (2) (in main text), we can perform standard Bayesian logistic regression. Posterior distributions for Bayesian analysis of a logistic regression model are not available as closed-form expressions based on a conjugate prior and instead standard Bayes can be implemented by a Metropolis-Hasting sampling technique with either flat priors or Jeffrey's priors. Gelman et al. (2008) suggested weakly informative Cauchy distributions as priors for the regression coefficients in logistic regression to reduce the separability issue. With an approximate EM algorithm, this non-informative Bayes method can be implemented in a fast and easy way to obtain posterior draws. We will use this method with weakly

†*Address for correspondence:* Wenting Cheng, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.

Email: chengwt@umich.edu

informative Cauchy priors throughout this paper. The implementation of standard Bayes was done in R using the package 'arm' with the function 'bayesglm'.

Constrained Maximum Likelihood

The constrained maximum likelihood (constrained ML) estimation optimizes the joint log-likelihood under the set of constraints generated based on the approximate relationship equations in (6) (in main text). As we have the point estimates and the standard errors of $\boldsymbol{\beta}$ from the established model, we require the parameter estimates for $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ to result in the derived estimated $\boldsymbol{\beta}$ to be within d standard errors of the old point estimates:

$$\begin{aligned}\min_{\boldsymbol{\gamma},\boldsymbol{\theta}}\Big\{ &\sum_{i=1}^{n}[-\mathrm{Y_i}(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})+\log(1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i}))] \\ &+\sum_{\mathrm{i=1}}^{\mathrm{n}}\frac{(\mathrm{B_i}-\sum_{\mathrm{j=0}}^{\mathrm{p}}\theta_{\mathrm{j}}\mathrm{X_{ij}})^2}{2\hat{\sigma}_2^2}\Big\} \\ \text{s.t.}&\frac{\gamma_j+\gamma_{p+1}\theta_j}{(1+\gamma_{p+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}}\in[\bar{\beta}_j-d\bar{S}_j,\bar{\beta}_j+d\bar{S}_j], j=0,\ldots,p\end{aligned} \tag{1}$$

In this optimization problem, $\hat{\sigma}_2^2$ is a plug-in estimator and is the OLS residual variance from fitting $\mathrm{E(B|\mathbf{X})}$ and d is a scale parameter representing the strength of external information. A value of d approaching zero would correspond to assuming that the β's are known precisely. Intuitively a value of d in the range of 1 to 2 would capture much of the uncertainty regarding the β's, and a large value of d would correspond to ignoring the established model. From simulations, we find that fixing d as $d = 1$ leads to decent properties of the estimates of $\boldsymbol{\gamma}$. To solve this optimization problem, we use function **solnp** in R package **Rsolnp**, a function that efficiently solves a general nonlinear optimization problem using Lagrange multipliers. For computational convenience, we further specify wide lower and upper bounds for each of these parameters: $\gamma_{\mathrm{j}} \in [\hat{\gamma}_{\mathrm{j}} - 5\hat{\mathrm{SE}}(\gamma_{\mathrm{j}}), \hat{\gamma}_{\mathrm{j}} + 5\hat{\mathrm{SE}}(\gamma_{\mathrm{j}})], \mathrm{j} = 0, \cdots, \mathrm{p}+1, \theta_{\mathrm{j}} \in [\hat{\theta}_{\mathrm{j}} - 5\hat{\mathrm{SE}}(\theta_{\mathrm{j}}), \hat{\theta}_{\mathrm{j}} + 5\hat{\mathrm{SE}}(\theta_{\mathrm{j}})], \mathrm{j} = 0, \cdots, \mathrm{p}$ where $\hat{\gamma}_j, j = 0, \cdots, p+1$ and $\hat{\theta}_j, j = 0, \cdots, p$ are the MLEs and $\hat{\mathrm{SE}}(\gamma_{\mathrm{j}}), \mathrm{j} = 0, \cdots, \mathrm{p}+1$ and $\hat{\mathrm{SE}}(\theta_{\mathrm{j}}), \mathrm{j} = 0, \cdots, \mathrm{p}$ are the corresponding standard errors.

We also consider a modification to the constrained ML solution above by adding a

Firth penalty term to the objective function:

$$\begin{aligned}\min_{\boldsymbol{\gamma},\boldsymbol{\theta}}\Big\{ &\sum_{i=1}^{n}[-\mathrm{Y_i}(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})+\log(1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i}))] \\ &+\sum_{\mathrm{i=1}}^{\mathrm{n}}\frac{(\mathrm{B_i}-\sum_{\mathrm{j=0}}^{\mathrm{p}}\theta_{\mathrm{j}}\mathrm{X_{ij}})^2}{2\hat{\sigma}_2^2}-0.5\log|\mathbf{I}(\boldsymbol{\gamma})|\Big\} \\ \text{s.t.}&\frac{\gamma_j+\gamma_{p+1}\theta_j}{(1+\gamma_{p+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}}\in[\bar{\beta}_j-d\bar{S}_j,\bar{\beta}_j+d\bar{S}_j],j=0,\ldots,p\end{aligned} \tag{2}$$

where $|\mathbf{I}(\boldsymbol{\gamma})|$ is the determinant of the Fisher information matrix of the likelihood function $\mathrm{L(Y|\mathbf{X},B)}$.

Informative Full Bayes

In informative full Bayes, starting with the joint likelihood $\mathrm{L(Y|\mathbf{X},B)L(B|\mathbf{X})}$ we translate the constraints in (6) (in main text) to prior information. The first step is to write down the joint likelihood function with priors on $\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2$:

$$\begin{aligned}&p(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2|\text{data})\propto \mathrm{L(Y|\mathbf{X},B},\boldsymbol{\gamma})\cdot\mathrm{L(B|\mathbf{X}},\boldsymbol{\theta},\sigma_2^2)\cdot\pi(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2) \\ &=\Big\{\prod_{\mathrm{i=1}}^{\mathrm{n}}\frac{\exp((\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})\mathrm{Y_i})}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})}\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\big(-\frac{1}{2\sigma_2^2}(\mathrm{B_i}-\sum_{\mathrm{j=0}}^{\mathrm{p}}\theta_{\mathrm{j}}\mathrm{X_{ij}})^2\big)\Big\}\cdot\pi(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2)\end{aligned} \tag{3}$$

The logistic regression approximation result (6) (in the main manuscript) suggests that $\theta_j=\frac{1}{\gamma_{p+1}}(\beta_j\sqrt{1+\frac{\gamma_{p+1}^2\sigma_2^2}{1.7^2}}-\gamma_j),j=0,\ldots,p$. We can re-parametrize (3) in terms of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and σ_2^2, and include a Jacobian transformation matrix denoted by $\mathbf{J}$, where $|\mathbf{J}|=\frac{1}{|\gamma_{p+1}^{p+1}|}(1+\frac{\gamma_{p+1}^2\sigma_2^2}{1.7^2})^{\frac{p+1}{2}}$. Now the likelihood is represented in terms of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and σ_2^2.

We specify a non-informative prior inverse-gamma$(0.01,0.01)$ for σ_2^2 and independent weakly informative Cauchy priors for $\boldsymbol{\gamma}$ (Gelman et al., 2008). For γ_0 we specify a Cauchy prior with location parameter 0, scale parameter 10. For $\gamma_j,j=1,\ldots,p+1$ we specify a Cauchy prior with location parameter 0, scale parameter 2.5. This is achieved through the hierarchical representation:

$$\begin{aligned}&\gamma_0\sim\mathrm{N}(0,\mathrm{k_0^2}),\gamma_1\sim\mathrm{N}(0,\mathrm{k_1^2}),\ldots,\gamma_{\mathrm{p+1}}\sim\mathrm{N}(0,\mathrm{k_{p+1}^2}) \\ &k_0^2\sim\mathrm{Inv}-\chi^2(1,10^2),\mathrm{k_1^2}\sim\mathrm{Inv}-\chi^2(1,2.5^2),\ldots,\mathrm{k_{p+1}^2}\sim\mathrm{Inv}-\chi^2(1,2.5^2)\end{aligned} \tag{4}$$

As a result, the prior distribution for the coefficient $\gamma_j, j = 0, \ldots, p+1$ is a mixture of normals with unknown scale parameter $\mathrm{k_j}$ that follow an inverse chi-square distribution.

For parameters $\boldsymbol{\beta}$, we use the constraints directly as priors:

$$\beta_{\mathrm{j}} = \frac{\gamma_{\mathrm{j}} + \gamma_{\mathrm{p+1}}\theta_{\mathrm{j}}}{(1 + \gamma_{\mathrm{p+1}}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}} \sim \mathrm{N}(\bar{\beta}_{\mathrm{j}}, \bar{\mathrm{S}}_{\mathrm{j}}^2), \mathrm{j} = 0, \ldots, \mathrm{p} \tag{5}$$

Then we can rewrite the joint distribution in terms of $\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_2^2, \mathbf{k}^2$ as $p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_2^2, \mathbf{k}^2|\mathbf{Y}, \mathbf{X}, \mathbf{B}) \propto$
$\{\prod_{i=1}^{n} \frac{\exp((\sum_{j=0}^{p}\gamma_j X_{ij} + \gamma_{p+1}B_i)Y_i)}{1+\exp(\sum_{j=0}^{p}\gamma_j X_{ij} + \gamma_{p+1}B_i)}\} \cdot \{\prod_{\mathrm{i=1}}^{\mathrm{n}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp(-\frac{1}{2\sigma_2^2}(\mathrm{B_i} - \frac{\beta_0\sqrt{1+\frac{\gamma_{p+1}^2\sigma_2^2}{1.7^2}} - \gamma_0}{\gamma_{\mathrm{p+1}}} - \sum_{\mathrm{j=1}}^{\mathrm{p}} \frac{\beta_{\mathrm{j}}\sqrt{1+\frac{\gamma_{p+1}^2\sigma_2^2}{1.7^2}} - \gamma_{\mathrm{j}}}{\gamma_{\mathrm{p+1}}}\mathrm{X_{ij}})^2)$
$\pi(\boldsymbol{\beta}) \cdot \{\prod_{j=0}^{p+1} \frac{1}{\sqrt{2\pi k_j^2}} \exp(-\frac{\gamma_j^2}{2k_j^2})\} \cdot \{\prod_{j=0}^{p+1} \pi(k_j^2)\} \cdot \pi(\sigma_2^2) \cdot |\mathbf{J}|$

After some algebraic calculations, the full conditional distributions of $\beta_0, \ldots, \beta_p$ turn out to be normal, each with distribution function $\mathrm{N}(\mu_{\beta_{\mathrm{j}},\mathrm{n}}, \sigma^2_{\beta_{\mathrm{j}},\mathrm{n}}), \mathrm{j} = 0, \ldots, \mathrm{p}$. The full conditional distributions of $k_0^2, k_1^2, \ldots, k_{p+1}^2$ are inverse chi-square, each with distribution function $\mathrm{Inv} - \chi^2(2, \frac{1}{2}(\mathrm{s_j^2} + \gamma_{\mathrm{j}}^2)), \mathrm{s_0} = 10, \mathrm{s_1} = \cdots = \mathrm{s_{p+1}} = 2.5, \mathrm{j} = 0, \ldots, \mathrm{p}+1$. The full conditional distributions of $\gamma_0, \ldots, \gamma_{p+1}$ and the full conditional distribution of σ_2^2 do not have closed form expressions. A Metropolis-Hastings sampling algorithm is needed.

2. Web Appendix B

Logistic Regression Approximation in Main Text, Equations (6) and (13)

The logistic-normal integral of the form $\mathrm{G}(\eta, \tau) = \int_{-\infty}^{+\infty} \mathrm{H(z)}\tau^{-1}\phi(\frac{\mathrm{z}-\eta}{\tau})\mathrm{dz}$ often appears in the studies of logistic regression model calibration where a subset of predicting variables are measured with errors. Monahan and Stefanski (1992) demonstrated a normal scale mixture representation of the logistic cumulative distribution function H(z), showing that H(z) can be approximated by a finite location-scale mixture of normal distribution functions: $\mathrm{H(z)} \approx \mathrm{H_k(z)} = \sum_{\mathrm{i=1}}^{\mathrm{k}} \mathrm{p_{k,i}}\Phi(\mathrm{z} \times \mathrm{s_{k,i}}), \mathrm{k} = 1, 2, \ldots,$ where $p_{k,i}$ is a fixed value and can be considered the weight of each normal CDF. $s_{k,i}$ is also a fixed value and can be considered as the corresponding scale parameter. All values of $p_{k,i}$ and $s_{k,i}$ can be found in their Least Maximum Approximants Table. Numerically studies show that this approximation is remarkably good for k as small as 3. In logistic regression calibration it is generally acceptable to take $k = 1$. Based on the Least Maximum Approximants Table, the corresponding values of $p_{k,i}, s_{k,i}$ are $p_{1,1} = 1$ and $s_{1,1} \approx 0.59$. Then we have

the following conclusion:

$$\mathrm{H(z)} \approx \Phi(\mathrm{z} \times \mathrm{s_{1,1}}) \approx \Phi(0.59\mathrm{z}) \approx \Phi(\mathrm{z}/1.7) \tag{6}$$

Sketch of Proof for Logistic Regression Approximation Equation (6)

Assume that $\mathbf{B}|\mathbf{X}$ is univariate normal with mean $\mathrm{m_B} = \mathbf{X}\boldsymbol{\theta}$ and variance σ_2^2.

$\Pr(\mathbf{Y} = 1|\mathbf{X}) = \int \mathrm{H}(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + \mathbf{B}\gamma_\mathrm{B}) \frac{1}{\sqrt{2\pi\sigma_2^2}} \mathrm{e}^{-\frac{(\mathbf{B}-\mathbf{X}\boldsymbol{\theta})^2}{2\sigma_2^2}} \mathrm{d}\mathbf{B}$

$= \int \mathrm{H(z)} \frac{1}{\sqrt{2\pi\sigma_2^2\gamma_B^2}} \mathrm{e}^{-\frac{(\frac{\mathrm{z}-\mathbf{X}\boldsymbol{\gamma}_\mathbf{x}}{\gamma_\mathrm{B}} - \mathbf{X}\boldsymbol{\theta})^2}{2\sigma_2^2}} \mathrm{dz}$ by changing $\mathbf{X}\boldsymbol{\gamma}_x + \mathbf{B}\gamma_B$ to z

$\approx \int \Phi(\mathrm{z} \times \mathrm{s_{1,1}}) \frac{1}{\sqrt{2\pi\sigma_2^2\gamma_B^2}} \mathrm{e}^{-\frac{(\frac{\mathrm{z}-\mathbf{X}\boldsymbol{\gamma}_\mathrm{X}}{\gamma_\mathrm{B}} - \mathbf{X}\boldsymbol{\theta})^2}{2\sigma_2^2}} \mathrm{dz}$

$= \int \Phi[(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{X}\boldsymbol{\theta})\gamma_\mathrm{B} + \mathrm{C}\gamma_\mathrm{B}\sigma_2)\mathrm{s_{1,1}}] \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\mathrm{C}^2}{2}} \mathrm{dC}$

$= \Phi\big(\frac{(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{X}\boldsymbol{\theta})\gamma_\mathrm{B})\mathrm{s_{1,1}}}{\sqrt{1+\gamma_B^2\sigma_2^2 s_{1,1}^2}}\big)$

$\approx \mathrm{H}\big(\frac{(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{X}\boldsymbol{\theta})\gamma_\mathrm{B})}{\sqrt{1+\gamma_B^2\sigma_2^2 s_{1,1}^2}}\big)$

$\approx \mathrm{H}\big(\frac{(\mathbf{X}\gamma_\mathrm{x} + (\mathbf{X}\theta)\gamma_\mathrm{B})}{\sqrt{1+\gamma_B^2\sigma_2^2/1.7^2}}\big)$

Note that the above derivation is based on calculating the integral of product of a normal CDF and a standard normal PDF. In most cases the above is a good approximation, unless $\boldsymbol{\gamma}_{p+1}^2\sigma_2^2$ is too large (Carroll et al., 2006).

Sketch of Proof for Logistic Regression Approximation Equation (13)

For the case that $\mathbf{B}|\mathbf{X}$ is multivariate normal with L dimensions with mean $\mathbf{X}\boldsymbol{\theta}$ and covariance matrix $\mathbf{V}_{L\times L}$, the derivation will be slightly different:

$\Pr(\mathbf{Y} = 1|\mathbf{X} = \mathbf{x}) = \int \Pr(\mathbf{Y} = 1|\mathbf{X} = \mathbf{x}, \mathbf{B} = \mathbf{b}) \frac{1}{(2\pi)^{\mathrm{L}/2}|\mathbf{V}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})^\mathrm{T}\mathbf{V}^{-1}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})} \mathrm{d}\mathbf{b}$

$= \int \mathrm{H}(\mathbf{x}\boldsymbol{\gamma}_\mathrm{x} + \mathbf{b}\gamma_\mathrm{B}) \frac{1}{(2\pi)^{\mathrm{L}/2}|\mathbf{V}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})^\mathrm{T}\mathbf{V}^{-1}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})} \mathrm{d}\mathbf{b}$

$\approx \int \Phi(\mathrm{s_{1,1}}(\mathbf{x}\boldsymbol{\gamma}_\mathrm{x} + \mathbf{b}\gamma_\mathrm{B})) \frac{1}{(2\pi)^{\mathrm{L}/2}|\mathbf{V}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})^\mathrm{T}\mathbf{V}^{-1}(\mathbf{b}-\mathbf{x}\boldsymbol{\theta})} \mathrm{d}\mathbf{b}$

$= \int \Phi(\mathrm{s_{1,1}}(\mathbf{x}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{x}\boldsymbol{\theta} + \boldsymbol{\Sigma}\mathbf{c})\gamma_\mathrm{B}))\phi(\mathbf{c})\mathrm{d}\mathbf{c}$ by changing $\mathbf{b}$ to $\mathbf{x}\boldsymbol{\theta} + \boldsymbol{\Sigma}\mathbf{c}$ where $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \mathbf{V}$

$= \int \Pr(\mathbf{W} \leq \mathrm{s_{1,1}}(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\Sigma}\mathbf{C})\boldsymbol{\gamma}_\mathrm{B})|\mathbf{X} = \mathbf{x}, \mathbf{C} = \mathbf{c})\phi(\mathbf{c})\mathrm{d}\mathbf{c}$ where $\mathbf{W}$ is a standard normal random variable and is independent of $\mathbf{C}$

$= \Pr(\mathbf{W} \leq \mathrm{s_{1,1}}(\mathbf{X}\boldsymbol{\gamma}_\mathrm{x} + (\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\Sigma}\mathbf{C})\boldsymbol{\gamma}_\mathrm{B})|\mathbf{X} = \mathbf{x})$ by the law of total probability

$= \Pr(-\mathrm{s_{1,1}}(\mathbf{X}\boldsymbol{\gamma}_\mathrm{X} + \mathbf{X}\boldsymbol{\theta}\boldsymbol{\gamma}_\mathrm{B}) \leq \mathrm{s_{1,1}}\boldsymbol{\Sigma}\mathbf{C}\boldsymbol{\gamma}_\mathrm{B} - \mathbf{W}|\mathbf{X} = \mathbf{x})$

Let $\mathbf{Z} = s_{1,1}\boldsymbol{\Sigma}\mathbf{C}\boldsymbol{\gamma}_B - \mathbf{W}$. Then $\mathbf{Z} \sim \mathrm{N}(0, \mathrm{s_{1,1}^2}\boldsymbol{\gamma}_\mathrm{B}^\mathrm{T}\mathbf{V}\boldsymbol{\gamma}_\mathrm{B} + 1)$ by Delta method

Then line seven $= \Pr(-\mathrm{s}_{1,1}(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}} + \mathbf{X}\boldsymbol{\theta}\boldsymbol{\gamma}_{\boldsymbol{B}}) \leq \mathbf{Z}|\mathbf{X} = \mathbf{x})$

$= \Pr(-\frac{\mathrm{s}_{1,1}(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}}+\mathbf{X}\boldsymbol{\theta}\boldsymbol{\gamma}_{\mathrm{B}})}{\sqrt{s_{1,1}^2\boldsymbol{\gamma}_B^T\mathbf{V}\boldsymbol{\gamma}_B+1}} \leq \frac{\mathbf{Z}}{\sqrt{s_{1,1}^2\boldsymbol{\gamma}_B^T\mathbf{V}\boldsymbol{\gamma}_B+1}}|\mathbf{X} = \mathbf{x})$

$= \Phi(\frac{\mathrm{s}_{1,1}(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}}+\mathbf{X}\boldsymbol{\theta}\boldsymbol{\gamma}_{\boldsymbol{B}})}{\sqrt{s_{1,1}^2\boldsymbol{\gamma}_B^T\mathbf{V}\boldsymbol{\gamma}_B+1}}|\mathbf{X} = \mathbf{x})$

$\approx \mathrm{H}(\frac{(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}}+\mathbf{X}\boldsymbol{\theta}\boldsymbol{\gamma}_{\mathrm{B}})}{\sqrt{s_{1,1}^2\boldsymbol{\gamma}_B^T\mathbf{V}\boldsymbol{\gamma}_B+1}}|\mathbf{X} = \mathbf{x})$

3. Web Appendix C

Relationship Equations for Binary B

When B is a binary variable, based on the Bayes theorem, there is a relationship equation connecting $\Pr(\mathrm{Y} = 1|\mathrm{X}), \Pr(\mathrm{Y} = 1|\mathrm{X}, \mathrm{B})$ and $\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y})$, regardless of the type of variable B is (Grill et al., 2015; Satten and Kupper, 1993).

$$\frac{\Pr(\mathrm{Y} = 1|\mathrm{X}, \mathrm{B})}{\Pr(\mathrm{Y} = 0|\mathrm{X}, \mathrm{B})} = \frac{\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y} = 1)}{\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y} = 0)} \cdot \frac{\Pr(\mathrm{Y} = 1|\mathrm{X})}{\Pr(\mathrm{Y} = 0|\mathrm{X})} \tag{7}$$

Re-arranging (7) and take the log on both sides, we have:

$$\log\left\{\frac{\Pr(\mathrm{Y} = 1|\mathrm{X})}{\Pr(\mathrm{Y} = 0|\mathrm{X})}\right\} = \log\left\{\frac{\Pr(\mathrm{Y} = 1|\mathrm{X}, \mathrm{B})}{\Pr(\mathrm{Y} = 0|\mathrm{X}, \mathrm{B})}\right\} + \log\left\{\frac{\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y} = 0)}{\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y} = 1)}\right\} \tag{8}$$

Equation (8) indicates that when B is binary, we need to define a model for $\mathrm{B}|\mathrm{X}, \mathrm{Y}$ instead of a model for $\mathrm{B}|\mathrm{X}$. Assume $\mathrm{logit}(\Pr(\mathrm{B} = 1|\mathrm{X}, \mathrm{Y})) = \sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}} + \phi_{\mathrm{p}+1}\mathrm{Y}$. So $\mathrm{f}(\mathrm{B}|\mathrm{X}, \mathrm{Y}) = \frac{\mathrm{e}^{(\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1}\mathrm{Y})\mathrm{B}}}{1+\mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1}\mathrm{Y}}}$. By looking at the log odds ratio in equation (8), we find that the left hand side is $\sum_{\mathrm{j}=0}^{\mathrm{p}} \beta_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}$, a linear combination of βs. So equation (8) can be written as:

$$\beta_0+\beta_1\mathrm{X}_1+...+\beta_{\mathrm{p}}\mathrm{X}_{\mathrm{p}} = \gamma_0+\gamma_1\mathrm{X}_1+...+\gamma_{\mathrm{p}}\mathrm{X}_{\mathrm{p}}+\gamma_{\mathrm{p}+1}\mathrm{B}+\log\left\{\frac{\mathrm{e}^{(\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}})\mathrm{B}}}{1+\mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}}} \cdot \frac{1+\mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1}}}{\mathrm{e}^{(\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1})\mathrm{B}}}\right\} \tag{9}$$

The right hand side of this expression can be rewritten as

$\gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X}_{\mathrm{p}} + \gamma_{\mathrm{p}+1}\mathrm{B} + \log\left\{\frac{1+\mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1}}}{1+\mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}}} \cdot \frac{1}{\mathrm{e}^{\phi_{\mathrm{p}+1}\mathrm{B}}}\right\}$

$= \gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X}_{\mathrm{p}} + \gamma_{\mathrm{p}+1}\mathrm{B} - \phi_{\mathrm{p}+1}\mathrm{B} + \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}+\phi_{\mathrm{p}+1}}\right\} - \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j}=0}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X}_{\mathrm{j}}}\right\}$

This expression can be approximated using Taylor series expansions to obtain an expression which is linear in the X_j's. A number of different ways of approximating it are possible. Below we write out the expressions for expansion about $\phi_0 = \phi_1 = ... = \phi_{p+1} = 0$. Other expansions about $\phi_1 = ... = \phi_{p+1} = 0$ and about the unconstrained MLE's

were also considered. The second approximation did lead to slightly improved results in some cases. The third approximation did not lead to a satisfactory linearization. Using the expansion about $\phi_0 = \phi_1 = ... = \phi_{p+1} = 0$ we obtain

$$\gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X_p} + \gamma_{\mathrm{p+1}}\mathrm{B} - \phi_{\mathrm{p+1}}\mathrm{B} + \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}}}\right\} - \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j}}\right\}$$
$$= \gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X_p} + \gamma_{\mathrm{p+1}}\mathrm{B} + \log\left\{\frac{1+\mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}}}}{1+\mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j}}} \cdot \frac{1}{\mathrm{e}^{\phi_{\mathrm{p+1}}\mathrm{B}}}\right\}$$
$$= \gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X_p} + \gamma_{\mathrm{p+1}}\mathrm{B} - \phi_{\mathrm{p+1}}\mathrm{B} + \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}}}\right\} - \log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j}}\right\}$$
$$\approx \gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X_p} + \gamma_{\mathrm{p+1}}\mathrm{B} - \phi_{\mathrm{p+1}}\mathrm{B} + [\log 2 + \tfrac{1}{2}(\textstyle\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}}) + \tfrac{1}{8}(\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}})^2$$
$$+ \mathrm{O}((\textstyle\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}})^3)] - [\log 2 + \tfrac{1}{2}(\textstyle\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j}) + \tfrac{1}{8}(\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j})^2 + \mathrm{O}((\textstyle\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j})^3)]$$
$$\approx \gamma_0 + ... + \gamma_{\mathrm{p}}\mathrm{X_p} + \gamma_{\mathrm{p+1}}\mathrm{B} - \phi_{\mathrm{p+1}}\mathrm{B} + \tfrac{1}{2}\phi_{\mathrm{p+1}} + \tfrac{1}{8}(\textstyle\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}})^2 - \tfrac{1}{8}(\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j})^2$$
$$= \gamma_0 + \tfrac{1}{2}\phi_{\mathrm{p+1}} + \tfrac{1}{4}\phi_0\phi_{\mathrm{p+1}} + \tfrac{1}{8}\phi_{\mathrm{p+1}}^2 + \textstyle\sum_{\mathrm{j=1}}^{\mathrm{p}}(\gamma_{\mathrm{j}} + \frac{1}{4}\phi_{\mathrm{j}}\phi_{\mathrm{p+1}})\mathrm{X_j} + (\gamma_{\mathrm{p+1}} - \phi_{\mathrm{p+1}})\mathrm{B}$$

The third last equation is by Taylor series expansions of $\log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j} + \phi_{\mathrm{p+1}}}\right\}$ and $\log\left\{1 + \mathrm{e}^{\sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathrm{X_j}}\right\}$ at point 0, respectively. By matching the coefficient of each variable on the left hand side and the right hand side of the equation, we find an approximate relationship between $\boldsymbol{\gamma}, \boldsymbol{\phi}$ and $\boldsymbol{\beta}$, when B is a binary variable:

$$\begin{cases} \beta_0 \approx \gamma_0 + \frac{1}{2}\phi_{p+1} + \frac{1}{4}\phi_0\phi_{p+1} + \frac{1}{8}\phi_{p+1}^2 \\ \beta_j \approx \gamma_j + \frac{1}{4}\phi_j\phi_{p+1}, j = 1, ..., p \\ \gamma_{p+1} = \phi_{p+1} \end{cases} \tag{10}$$

Further Details of the Statistical Method for Informative Full Bayes for Binary B

The likelihood is first reparametrized in terms of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and includes the Jacobian matrix $\mathbf{M}$. We specify independent weakly informative Cauchy priors for $\boldsymbol{\gamma}$ by introducing latent variables k and use the constraints directly as priors for $\boldsymbol{\beta}$. Then we can rewrite the joint distribution in terms of $\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{k}^2$ as $\mathrm{p}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{k}^2|\mathrm{Y}, \mathbf{X}, \mathrm{B}) \propto \Bigg\{ \prod_{\mathrm{i=1}}^{\mathrm{n}} \frac{\exp((\sum_{\mathrm{j=0}}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p+1}}\mathrm{B_i})\mathrm{Y_i})}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p+1}}\mathrm{B_i})} \cdot$

$$\left[\frac{w_{i,\bar{\boldsymbol{\beta}}}}{1+\exp\left(-(\frac{4\beta_0 - 4\gamma_0 - 2\gamma_{p+1} - \frac{1}{2}\gamma_{p+1}^2}{\gamma_{p+1}} + \sum_{j=1}^{p} X_{ij}\frac{4(\beta_j - \gamma_j)}{\gamma_{p+1}})\right)} + \frac{1 - w_{i,\bar{\boldsymbol{\beta}}}}{1+\exp\left(-(\frac{4\beta_0 - 4\gamma_0 - 2\gamma_{p+1} - \frac{1}{2}\gamma_{p+1}^2}{\gamma_{p+1}} + \sum_{j=1}^{p} X_{ij}\frac{4(\beta_j - \gamma_j)}{\gamma_{p+1}} + \gamma_{p+1})\right)}\right]^{B_i} \cdot$$

$$\left[\frac{w_{i,\bar{\boldsymbol{\beta}}}}{1+\exp\left(\frac{4\beta_0 - 4\gamma_0 - 2\gamma_{p+1} - \frac{1}{2}\gamma_{p+1}^2}{\gamma_{p+1}} + \sum_{j=1}^{p} X_{ij}\frac{4(\beta_j - \gamma_j)}{\gamma_{p+1}}\right)} + \frac{1 - w_{i,\bar{\boldsymbol{\beta}}}}{1+\exp\left(\frac{4\beta_0 - 4\gamma_0 - 2\gamma_{p+1} - \frac{1}{2}\gamma_{p+1}^2}{\gamma_{p+1}} + \sum_{j=1}^{p} X_{ij}\frac{4(\beta_j - \gamma_j)}{\gamma_{p+1}} + \gamma_{p+1}\right)}\right]^{(1-B_i)} \Bigg\} \cdot$$

$$\pi(\boldsymbol{\beta}) \cdot \left\{ \prod_{j=0}^{p+1} \frac{1}{\sqrt{2\pi k_j^2}} \exp(-\frac{\gamma_j^2}{2k_j^2}) \right\} \cdot \left\{ \prod_{j=0}^{p+1} \pi(k_j^2) \right\} \cdot |\mathbf{M}|$$

After some algebraic calculations, the full conditional distributions of $k_0^2, k_1^2, \ldots, k_{p+1}^2$ are inverse chi-square, each with distribution function $\text{Inv}-\chi^2(2, \frac{1}{2}(\text{s}_\text{j}^2+\gamma_\text{j}^2)), \text{s}_0 = 10, \text{s}_1 = \cdots = \text{s}_{\text{p+1}} = 2.5, \text{j} = 0, \ldots, \text{p}+1$. The full conditional distributions of $\gamma_0, \ldots, \gamma_{p+1}$ and the full conditional distribution of $\beta_0, \ldots, \beta_p$ do not have closed form expressions. A Metropolis-Hastings sampling algorithm is needed.

4. Web Appendix D

Standard Error Calculations

Bootstrap Estimate of the Standard Error for the Constrained ML Estimate When B *Is Normal*

We would like to obtain a bootstrap estimate of the constrained ML estimator's standard error. In our study, we implement a parametric bootstrap as follows:

- Estimate the regression coefficients $\gamma_0, \ldots, \gamma_{p+1}$ and $\theta_0, \ldots, \theta_p$ by the constrained ML method for the original sample

- Calculate the fitted outcome $\hat{\text{B}}_\text{i}$ and residual $\text{E}_\text{i,B}$ for each observation: $\hat{\text{B}}_\text{i} = \hat{\theta}_0 + \hat{\theta}_1\text{X}_\text{i1} + \cdots + \hat{\theta}_\text{p}\text{X}_\text{ip}$ and $\text{E}_\text{i,B} = \text{B}_\text{i} - \hat{\text{B}}_\text{i}$

- Take bootstrap samples of the residual (sample with replacement), $\tilde{\mathbf{e}}_\mathbf{b} = [\tilde{\text{E}}_\text{b,1,B}, \ldots, \tilde{\text{E}}_\text{b,n,B}]^\text{T}, \text{b} = 1, \ldots, \text{S}$, calculate bootstrapped $\mathbf{B}$ values $\tilde{\mathbf{B}}_\mathbf{b} = [\tilde{\text{B}}_\text{b1}, \ldots, \tilde{\text{B}}_\text{bn}]^\text{T}$, where $\tilde{\text{B}}_\text{bi} = \hat{\text{B}}_\text{i} + \tilde{\text{E}}_\text{b,i,B}$

- Calculate bootstrap $\mathbf{Y}$ values: $\tilde{\text{Y}}_\text{bi} \sim \text{Bernoulli}(\tilde{\text{P}}_\text{bi})$ where $\tilde{\text{P}}_\text{bi} = \Pr(\text{Y} = 1|\text{X}_\text{i}, \tilde{\text{B}}_\text{bi}, \hat{\boldsymbol{\gamma}})$

- Regress $\tilde{\mathbf{Y}}_\mathbf{b}$ on the fixed $\mathbf{X}$ design matrix and bootstrap samples $\tilde{\mathbf{B}}_\mathbf{b}$ to obtain bootstrap estimates of regression coefficients by the constrained ML method: $\tilde{\gamma}_{b,0}, \ldots, \tilde{\gamma}_{b,p+1}$

- The $\tilde{\boldsymbol{\gamma}}_\mathbf{b}$ can be used to construct bootstrap standard error: $\tilde{\sigma}_{.,j} = (\frac{\sum_{b=1}^{S}(\tilde{\gamma}_{b,j} - \bar{\tilde{\gamma}}_{.,j})^2}{S-1})^{1/2}$, $j = 0, \ldots, p+1$, in the usual bootstrap manner as described in Efron and Tibshirani (1986).

Bootstrap Estimate of the Standard Error for the Constrained ML Estimate When B *Is Binary*

- Estimate $\gamma_0, \ldots, \gamma_{p+1}$ and $\phi_0, \ldots, \phi_{p+1}$ by the constrained ML method for the original sample

- Calculate bootstrap $\mathbf{B}$ values: $\tilde{\mathrm{B}}_{\mathrm{bi}} \sim \mathrm{Bernoulli}(\hat{\mathrm{P}}_{\mathrm{B_i}}), \mathrm{b} = 1, \ldots, \mathrm{S}$, where $\hat{\mathrm{P}}_{\mathrm{B_i}} = \Pr(\mathrm{B} = 1|\mathrm{X_i}, \mathrm{Y_i}, \hat{\boldsymbol{\phi}}, \bar{\boldsymbol{\beta}})$

- Calculate bootstrap $\mathbf{Y}$ values: $\tilde{\mathrm{Y}}_{\mathrm{bi}} \sim \mathrm{Bernoulli}(\tilde{\mathrm{P}}_{\mathrm{bi}})$ where $\tilde{P}_{bi} = \Pr(\mathrm{Y} = 1|\mathrm{X_i}, \tilde{\mathrm{B}}_{\mathrm{bi}}, \hat{\boldsymbol{\gamma}})$

- Regress $\tilde{\mathbf{Y}}_{\mathbf{b}}$ on the fixed $\mathbf{X}$ and bootstrap samples $\tilde{\mathbf{B}}_{\mathbf{b}}$ to obtain bootstrap estimates of regression coefficients by the constrained ML method: $\tilde{\gamma}_{b,0}, \ldots, \tilde{\gamma}_{b,p+1}$

- Construct bootstrap standard error: $\tilde{\sigma}_{.,\mathrm{j}} = (\frac{\sum_{\mathrm{b}=1}^{\mathrm{S}} (\tilde{\gamma}_{\mathrm{b,j}} - \bar{\tilde{\gamma}}_{.,\mathrm{j}})^2}{\mathrm{S}-1})^{1/2}$, $\mathrm{j} = 0, \ldots, \mathrm{p}+1$.

Comparison of the average bootstrap point estimates and standard errors to the Monte Carlo average and standard deviation are provided in Table S3. The bootstrap mean and the Monte Carlo mean appear to be quite similar. The bootstrap estimated standard error is too large in the first scenario for the constrained ML method. However, with the Firth correction the standard errors from the bootstrap method match the empirical standard deviations, in both scenarios.

Standard Error Calculations for other Estimators

For the direct regression and direct regression plus Firth correction the standard errors are based on the asymptotic formulas. For the non-informative Bayes, the informative Bayes and the transformation approach the standard deviation of the posterior draws are used as the standard errors. For the Chatterjee et al. (2016) method the standard errors are based on bootstrap estimates as described above.

5. Web Appendix D

Simulation Results for Binary B

Similar to the study presented in Table 1 in the main paper we conducted a simulation study for binary B. This simulation scenario has three predicting variables X_1, X_2, B where B is binary. There are 75 observations in each dataset. 500 datasets are generated. Y_i is Bernoulli distributed with $\text{logit}(\Pr(Y_i = 1|X_{i1}, X_{i2}, B_i)) = 2 + 4X_{i1} + 4X_{i2} + 2B_i$. X_{i1}, X_{i2} are independently and identically distributed on $U(-0.75, 0.25)$ and B_i is simulated from $\text{logit}(\Pr(B_i = 1|X_{i1}, X_{i2}) = 1 + X_{i1} + X_{i2}$. A logistic regression based on a large dataset of 10000 subjects gives estimates for the model $\text{logit}(\Pr(Y = 1|\mathbf{X})) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. The estimates and standard errors are $\bar{\beta}_0 = 2.97, \bar{S}_0 = 0.06, \bar{\beta}_1 = 3.87, \bar{S}_1 = 0.10, \bar{\beta}_2 = 3.68, \bar{S}_2 = 0.10$.

Table S4a. summarizes the simulation results for this simulation scenario, where B is binary. The constrained methods in this simulation scenario exhibit greater improvement in estimating efficiency for the coefficients of $\mathbf{X}$ than in the simulation scenarios with Gaussian B. The constrained ML with Firth penalty and transformation approach can improve the relative efficiency of parameters γ_1 and γ_2 by more than 600%. For Brier score and AUC all the methods that use the external information are similar. They are slightly better than the methods that don't use the external information, and almost as good as the best possible value. They are all better than not using B. In this simulation scenario, the informative full Bayes method is more computationally intensive since both the conditional distributions of $\boldsymbol{\gamma}$ and the conditional distributions of $\boldsymbol{\beta}$ do not have closed form expressions. Drawing samples based on a Metropolis-Hasting sampling algorithm in this case is computationally demanding.

Additional Simulation Results for Gaussian B

Additional simulations were conducted to investigate a broader range of scenarios in which models may be misspecified, or covariates may not be necessary, or where covariates are highly correlated, or where there is a larger number of covariates. The scenarios considered are adaptations of the scenario in Table 1 (in main text) in which B is Gaussian. All the datasets are of size 55, and the results are based on 500 simulated datasets.

For each scenario the intercepts in the data generating model were chosen to give the proportion of observations with $Y = 1$ as close to 0.5. The scenarios are described below. The results are consistent with those in Table 1 (in main text), for the estimates of γ's the methods that use the external information have less variability. For the measures of predictive ability the methods are able to demonstrate some gain through using B compared to the established model, that the more sophisticated methods are slightly better than the simple methods (direct regression and simple logistic($\bar{p}$)). Overall there is no clear best method amongst the sophisticated methods although the constrained MLE + Firth and the transformation approach appear to perform well. The Chatterjee et al. (2016) method gives as good a performance as any other method that uses external information as measured by AUC, but does tend to give very slightly worse performance as measured by the Brier score, and does tend to give more variable predicted probabilities than the other methods.

Additional Simulation Results: X_1 not associated with B

$\mathrm{Y_i}$ is Bernoulli distributed with $\mathrm{logit}(\Pr(\mathrm{Y_i} = 1|\mathrm{X_{i1}}, \mathrm{X_{i2}}, \mathrm{B_i})) = 1.75 + 3\mathrm{X_{i1}} + 3\mathrm{X_{i2}} + 2\mathrm{B_i}$. $\mathrm{X_{i1}}, \mathrm{X_{i2}}$ are independently and identically distributed on $\mathrm{U}(-0.75, 0.25)$ and $\mathrm{B_i}$ is simulated as $\mathrm{B_i} = 0.5\mathrm{X_{i2}} + \mathrm{N}(0, 0.75^2)$. The results are shown in Table S4b.

Additional Simulation Results: B and X_1 are highly correlated

$\mathrm{Y_i}$ is Bernoulli distributed with $\mathrm{logit}(\Pr(\mathrm{Y_i} = 1|\mathrm{X_{i1}}, \mathrm{X_{i2}}, \mathrm{B_i})) = 2.75 + 3\mathrm{X_{i1}} + 3\mathrm{X_{i2}} + 2\mathrm{B_i}$. $\mathrm{X_{i1}}, \mathrm{X_{i2}}$ are independently and identically distributed on $\mathrm{U}(-0.75, 0.25)$ and $\mathrm{B_i}$ is simulated as $\mathrm{B_i} = 2\mathrm{X_{i1}} + 0.5\mathrm{X_{i2}} + \mathrm{N}(0, 0.75^2)$. The correlation between B and X_1 is 0.6. The results are shown in Table S4c.

Additional Simulation Results: Model for $B|X$ is misspecified

$\mathrm{Y_i}$ is Bernoulli distributed with $\mathrm{logit}(\Pr(\mathrm{Y_i} = 1|\mathrm{X_{i1}}, \mathrm{X_{i2}}, \mathrm{B_i})) = 0.14 + 3\mathrm{X_{i1}} + 3\mathrm{X_{i2}} + 2\mathrm{B_i}$. $\mathrm{X_{i1}}, \mathrm{X_{i2}}$ are independently and identically distributed on $\mathrm{U}(-0.75, 0.25)$ and $\mathrm{B_i}$ is simulated as $\mathrm{B_i} = 1.0 + \mathrm{X_{i1}} - 0.5\mathrm{X_{i2}}^2 + \mathrm{N}(0, 0.75^2)$. The results are shown in Table S4d.

Additional Simulation Results: Larger number of X variables

$\mathrm{Y_i}$ is Bernoulli distributed with $\mathrm{logit}(\Pr(\mathrm{Y_i} = 1|\mathrm{X_{i1}}, \mathrm{X_{i2}}, \mathrm{X_{i3}}, \mathrm{X_{i4}}, \mathrm{B_i})) = 1.55 + \mathrm{X_{i1}} + \mathrm{X_{i2}} + \mathrm{X_{i3}} + \mathrm{X_{i2}} + 2\mathrm{B_i}$. $\mathrm{X_{i1}}, \mathrm{X_{i2}}, \mathrm{X_{i3}}, \mathrm{X_{i4}}$, are independently and identically distributed on $\mathrm{U}(-0.75, 0.25)$ and $\mathrm{B_i}$ is simulated as $\mathrm{B_i} = 0.01\mathrm{X_{i1}} + 0.05\mathrm{X_{i2}} + 0.5\mathrm{X_{i3}} + 0.5\mathrm{X_{i4}} + \mathrm{N}(0, 0.75^2)$. The results are shown in Table S4e.

6. Web Appendix F

The Impact of Varying the Tuning Parameter d in Constrained Maximum Likelihood Method for the Prostate Cancer Example

Table S5. show the results for four different values of d for the constrained ML method, and how they compare with direct regression and the Chatterjee et al. (2016) method. As expected large d gives very similar results to direct regression, and very small d gives similar results to the Chatterjee et al. (2016) method.

7. Web Appendix G

Details of Computational Implementation, Algorithms, Software and R Functions Used

Table S6. provides summary information on how the different methods were implemented. Time refers to computation time for 500 simulated datasets in Table 1 in main text.

Table S3. Simulation results of parametric bootstrap: we report the ratio of average bootstrap mean and Monte Carlo mean $(\frac{1}{500}\sum_{\mathrm{m}=1}^{500}\bar{\tilde{\gamma}}_{\mathrm{m,j}})/(\frac{1}{500}\sum_{\mathrm{m}=1}^{500}\hat{\gamma}_{\mathrm{m,j}})$ and the ratio of average bootstrap standard error and Monte Carlo standard deviation $(\frac{1}{500}\sum_{\mathrm{m}=1}^{500}\tilde{\sigma}_{\mathrm{m,j}})/\sqrt{V(\hat{\gamma}_j)}$ of each regression coefficient

Method	Ratio	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
First simulation scenario. B Gaussian				
Constrained ML	Avg.Boot.Mean/MC.Mean	1.10	1.10	1.23
	Avg.Boot.SE/MC.SD	1.74	1.81	1.89
Constrained ML + Firth	Avg.Boot.Mean/MC.Mean	0.98	0.99	0.99
	Avg.Boot.SE/MC.SD	1.03	1.05	1.10
Second simulation scenario. B binary				
Constrained ML	Avg.Boot.Mean/MC.Mean	1.13	1.12	1.03
	Avg.Boot.SE/MC.SD	1.13	1.00	0.96
Constrained ML + Firth	Avg.Boot.Mean/MC.Mean	1.07	1.06	0.89
	Avg.Boot.SE/MC.SD	0.95	0.94	0.97

Table S4a. Simulation results for binary B: for each method, we report mean (relative efficiency w.r.t. direct regression), MSE, average Brier score, average AUC, average $\hat{p}$ (SD) and computing time for 500 datasets of size 75

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	Scaled Brier Score	AUC	$\hat{p}$ mean(SD)	Time
True value	4	4	2	0.621	0.871	0.69 (0.300)	-
Established model using known $\bar{\beta}$	-	-	-	0.763	0.800	0.68 (0.256)	-
Direct regression	4.49 (1)	4.40 (1)	2.22 (1)	0.667	0.861	0.69 (0.309)	1.2
MSE	3.48	3.22	0.82				
Direct regression + Firth	3.98 (1.49)	3.90 (1.50)	2.00 (1.46)	0.660	0.862	0.68 (0.297)	3.3
MSE	2.18	2.05	0.53				
Non-informative Bayes	3.78 (1.75)	3.70 (1.78)	1.92 (1.66)	0.657	0.861	0.68 (0.279)	4.4
MSE	1.90	1.81	0.47				
Constrained ML	4.08 (5.15)	3.94 (3.90)	2.13 (1.03)	0.646	0.868	0.67 (0.315)	43.9
MSE	0.64	0.82	0.79				
Constrained ML + Firth	3.93 (12.14)	3.77 (11.27)	1.80 (1.96)	0.641	0.867	0.67 (0.306)	77.2
MSE	0.27	0.32	0.44				
Informative full Bayes	3.91 (9.03)	3.75 (8.62)	1.95 (1.57)	0.636	0.866	0.69 (0.290)	30939.1
MSE	0.39	0.36	0.61				
Transformation	3.91(8.92)	3.76(9.43)	1.96(1.63)	0.636	0.866	0.69 (0.278)	550.03
MSE	0.35	0.38	0.48				
Chatterjee et al.	4.21 (5.55)	4.03 (4.79)	2.21 (1.06)	0.637	0.867	0.69 (0.305)	59.5
MSE	0.81	0.69	1.09				
Simple logistic $(\bar{p}, B)$	4.47 (1.4)	4.25 (1.46)	2.16 (1.11)	0.653	0.866	0.69 (0.305)	1.2
MSE	2.85	2.28	1.01				

Table S4b. Same setting as main text Table 1, continuous B with X_1 not associated with B: for each method, we report mean (relative efficiency w.r.t. direct regression), MSE, average Brier score, average AUC, average $\hat{p}$ (SD) and computing time for 500 datasets of size 55

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	**Scaled Brier Score**	**AUC**	**$\hat{p}$ mean(SD)**	**Time**
True value	3	3	2	0.589	0.871	0.48 (0.331)	-
Established model using known $\bar{\beta}$	-	-	-	0.785	0.768	0.49 (0.229)	-
Direct regression	4.53 (1)	5.37 (1)	2.64 (1)	0.643	0.859	0.48 (0.342)	1.4
MSE	528.27	1741.87	60.29				
Direct regression + Firth	3.02 (211.71)	3.07 (592.87)	1.96 (117.56)	0.636	0.859	0.48 (0.323)	3.5
MSE	2.48	2.93	0.51				
Non-informative Bayes	2.83 (257.37)	2.9 (787.48)	1.97 (143.42)	0.635	0.859	0.48 (0.305)	3.8
MSE	2.07	2.22	0.42				
Constrained ML	3.11 (456.97)	2.99 (1922.70)	2,25 (80.54)	0.613	0.866	0.48 (0.333)	90.7
MSE	1.16	0.90	0.80				
Constrained ML + Firth	2.83 (780.41)	2.78 (3448.79)	1.91 (143.32)	0.607	0.866	0.49 (0.318)	81.2
MSE	0.70	0.55	0.43				
Informative full Bayes	2.84 (715.76)	2.77 (2810.88)	2.17 (114.75)	0.611	0.865	0.48 (0.319)	7349.5
MSE	0.76	0.67	0.55				
Transformation	2.84 (817.69)	2.84 (3411.74)	1.92 (111.34)	0.608	0.866	0.49 (0.307)	1096.2
MSE	0.67	0.53	0.54				
Chatterjee et al.	3.19 (375.1)	3.09 (1213.92)	2.29 (66.98)	0.615	0.866	0.48 (0.335)	52.2
MSE	1.44	1.44	0.98				
Simple logistic $(\bar{p}, B)$	2.89 (230.41)	3.80 (440.16)	2.17 (80.91)	0.636	0.862	0.48 (0.340)	1.2
MSE	2.29	4.59	0.77				

Table S4c. Same setting as main text Table 1, continuous B, B and X_1 are highly correlated: for each method, we report mean (relative efficiency w.r.t. direct regression), MSE, average Brier score, average AUC, average $\hat{p}$ (SD) and computing time for 500 datasets of size 55

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	**Scaled Brier Score**	**AUC**	**$\hat{p}$ mean(SD)**	**Time**
True value	3	3	2	0.454	0.923	0.49 (0.369)	-
Established model using known $\bar{\beta}$	-	-	-	0.634	0.851	0.50 (0.305)	-
Direct regression	5.23 (1)	6.93 (1)	4.39 (1)	0.509	0.912	0.49 (0.380)	1.9
MSE	793.88	2448.72	895.3				
Direct regression + Firth	3.00 (237.26)	3.00 (701.62)	1.99 (1004.13)	0.502	0.913	0.49 (0.361)	2.9
MSE	3.33	3.47	0.89				
Non-informative Bayes	2.79 (286.73)	2.73 (922.46)	2.12 (986.41)	0.497	0.914	0.49 (0.346)	4.1
MSE	2.80	2.71	0.92				
Constrained ML	2.86 (560.26)	3.24 (1480.73)	2.39 (275.80)	0.476	0.919	0.49 (0.372)	67.5
MSE	1.43	1.70	3.37				
Constrained ML + Firth	2.91 (967.56)	2.98 (3843.37)	1.92 (1199.52)	0.472	0.919	0.49 (0.36)	102.2
MSE	0.82	0.63	0.75				
Informative full Bayes	2.52 (749.66)	2.99 (2645.46)	2.50 (1046.34)	0.472	0.919	0.49 (0.369)	10850.3
MSE	1.26	0.90	1.09				
Transformation	2.97 (914.56)	3.04 (3340.43)	2.05 (972.02)	0.472	0.918	0.49 (0.356)	1647.3
MSE	0.86	0.73	0.92				
*Chatterjee et al.	2.91 (353.35)	3.37 (924.68)	2.47 (545.14)	0.479	0.918	0.49 (0.376)	57.9
MSE	2.24	2.77	1.86				
Simple logistic $(\bar{p}, B)$	4.74 (3.46)	2.76 (31.32)	3.36 (1.22)	0.503	0.913	0.50 (0.370)	1.4
MSE	231.05	77.74	733.51				

* The method of Chatterjee et al. did not converge for 59 of the 500 datasets. The results shown for Chatterjee et al. method were calculated from estimates based on the 441 datasets in which the Chatterjee et al. method does converge

Table S4d. Same setting as main text Table 1, continuous B, model for $B|X$ is misspecified: for each method, we report mean (relative efficiency w.r.t. direct regression), MSE, average Brier score, average AUC, average $\hat{p}$ (SD) and computing time for 500 datasets of size 55

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	**Scaled Brier Score**	**AUC**	**$\hat{p}$ mean(SD)**	**Time**
True value	3	3	2	0.529	0.893	0.49 (0.344)	-
Established model using known $\bar{\beta}$	-	-	-	0.747	0.792	0.50 (0.260)	-
Direct regression	3.61 (1)	3.51 (1)	2.34 (1)	0.585	0.882	0.49 (0.355)	1.6
MSE	4.07	4.09	1.22				
Direct regression + Firth	3.08 (1.62)	2.97 (1.73)	1.95 (1.89)	0.579	0.882	0.49 (0.336)	4.0
MSE	2.29	2.21	0.59				
Non-informative Bayes	2.93 (1.84)	2.77 (2.08)	2.01 (2.00)	0.576	0.883	0.49 (0.319)	3.8
MSE	2.01	1.89	0.55				
Constrained ML	2.97 (4.48)	3.06 (3.94)	2.27 (1.20)	0.551	0.889	0.48 (0.346)	66.0
MSE	0.82	0.97	0.99				
Constrained ML + Firth	2.84 (7.18)	2.82 (6.82)	1.91 (2.17)	0.549	0.888	0.48 (0.332)	85.8
MSE	0.54	0.59	0.52				
Informative full Bayes	2.80 (5.20)	2.89 (5.04)	2.30 (1.43)	0.548	0.889	0.48 (0.337)	10238.4
MSE	0.75	0.77	0.86				
Transformation	2.90 (7.12)	2.83 (6.95)	1.95 (1.89)	0.552	0.887	0.49 (0.324)	969.2
MSE	0.53	0.58	0.59				
Chatterjee et al.	3.09 (3.35)	3.16 (2.30)	2.33 (1.02)	0.554	0.888	0.48 (0.349)	41.1
MSE	1.11	1.69	1.18				
Simple logistic $(\bar{p}, B)$	3.91 (1.43)	2.71 (3.09)	2.17 (1.37)	0.576	0.885	0.49 (0.350)	0.9
MSE	3.39	1.32	0.83				

Table S4e. Same setting as main text Table 1, continuous B, larger number of X variables: for each method, we report mean (relative efficiency w.r.t. direct regression), MSE, average Brier score, average AUC, average $\hat{p}$ (SD) and computing time for 500 datasets of size 55

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	**Scaled Brier Score**	**AUC**	**$\hat{p}$ mean(SD)**	**Time**
True value	1	1	1	1	2	0.692	0.823	0.51 (0.299)	-
Established model using known $\bar{\beta}$	-	-	-	-	-	0.937	0.652	0.49 (0.154)	-
Direct regression	1.21 (1)	1.17 (1)	1.15 (1)	1.20 (1)	2.48 (1)	0.787	0.797	0.51 (0.325)	2.1
MSE	2.31	2.77	3.04	2.58	1.34				
Direct regression + Firth	1.00 (1.58)	0.96 (1.62)	0.95 (1.73)	1.00 (1.62)	2.00 (2.14)	0.768	0.797	0.50 (0.299)	5.8
MSE	1.43	1.70	1.75	1.56	0.52				
Non-informative Bayes	0.97 (1.76)	0.92 (1.81)	0.94 (1.95)	1.00 (1.79)	2.16 (1.96)	0.755	0.800	0.50 (0.284)	4.0
MSE	1.29	1.52	1.55	1.42	0.59				
Constrained ML	1.15 (1.95)	1.04 (1.63)	1.05 (1.36)	1.08 (1.10)	2.49 (0.91)	0.755	0.809	0.50 (0.319)	845.2
MSE	1.18	1.69	2.22	2.30	1.46				
Constrained ML + Firth	0.97 (3.57)	0.96 (3.90)	0.92 (4.03)	0.93 (1.96)	2.00 (2.01)	0.745	0.807	0.50 (0.294)	535.2
MSE	0.63	0.71	0.76	1.30	0.55				
Informative full Bayes	1.04 (3.74)	0.97 (4.08)	0.92 (5.11)	0.81 (3.76)	2.69 (1.47)	0.725	0.816	0.50 (0.317)	11249.8
MSE	0.61	0.67	0.60	0.71	1.23				
Transformation	0.86 (5.45)	0.81 (6.00)	0.76 (6.68)	0.75 (5.08)	2.03 (1.86)	0.721	0.814	0.50 (0.273)	700.5
MSE	0.43	0.49	0.51	0.56	0.60				
Chatterjee et al.	1.13 (2.56)	1.10 (2.78)	1.09 (3.11)	0.95 (2.62)	2.46 (1.03)	0.735	0.814	0.50 (0.317)	51.7
MSE	0.90	1.00	0.98	1.77	1.31				
Simple logistic $(\bar{p}, B)$	0.71 (8.98)	0.71 (10.9)	1.33 (3.41)	1.24 (3.29)	2.21 (1.73)	0.734	0.816	0.50 (0.300)	1.8
MSE	0.34	0.34	1.00	0.83	0.69				

Table S5. The impact of varying the scaling quantity d in constrained maximum likelihood method for the prostate cancer example, and the comparison with direct regression and the Chatterjee et al. method

Model		**PSA**	**Age**	**DRE findings**	**Prior biopsy history**	**Race**		**AUC**	**Brier Score**
Original PCPThg		1.29 (0.09)	0.031 (0.012)	1.00 (0.17)	-0.36 (0.18)	0.96 (0.27)	-	0.707	0.933
Estimated PCPThg		1.06 (0.18)	0.033 (0.012)	1.15 (0.26)	-1.44 (0.27)	0.44 (0.29)	-	0.716	0.975
Expanded model with PCA3 score							**PCA3**		
Direct regression		1.00 (0.19)	0.009 (0.013)	1.07 (0.27)	-1.30 (0.28)	0.04 (0.31)	0.56 (0.08)	0.767	0.950
Constrained ML	d=1	1.20 (0.09)	0.010 (0.005)	1.08 (0.17)	-0.55 (0.13)	0.30 (0.12)	0.59 (0.08)	0.766	0.953
	d=2	1.11 (0.10)	0.003 (0.005)	1.13 (0.24)	-0.73 (0.14)	0.041 (0.12)	0.58 (0.08)	0.767	0.951
	d=10	1.00 (0.18)	0.009 (0.012)	1.07 (0.27)	-1.30 (0.28)	0.038 (0.31)	0.57 (0.08)	0.767	0.950
	d=0.1	1.33 (0.07)	0.009 (0.004)	0.94 (0.10)	-0.40 (0.08)	0.61 (0.10)	0.59 (0.09)	0.761	0.888
Chatterjee et al. method		1.22 (0.08)	0.007 (0.005)	0.86 (0.10)	-0.20 (0.08)	0.58 (0.11)	0.56 (0.097)	0.759	0.888
Expanded model with binary T2: ERG							**T2: ERG**		
Direct regression		1.01 (0.18)	0.032 (0.012)	1.03 (0.26)	-1.44 (0.28)	0.57 (0.29)	0.77 (0.20)	0.745	0.929
Constrained ML	d=1	1.14 (0.07)	0.032 (0.004)	1.06 (0.15)	-0.52 (0.11)	0.81 (0.19)	0.74 (0.22)	0.742	0.928
	d=2	1.04 (0.09)	0.02 (0.01)	1.09 (0.21)	-0.69 (0.12)	0.55 (0.21)	0.73 (0.21)	0.744	0.924
	d=10	1.01 (0.19)	0.03 (0.01)	1.04 (0.27)	-1.44 (0.27)	0.56 (0.29)	0.75 (0.21)	0.744	0.929
	d=0.1	1.24 (0.03)	0.03 (0.004)	0.92 (0.05)	-0.36 (0.04)	1.10 (0.05)	0.72 (0.24)	0.735	0.902
Chatterjee et al. method		1.25 (0.03)	0.029 (0.002)	0.85 (0.05)	-0.37 (0.04)	1.06 (0.05)	0.77 (0.27)	0.736	0.911

Table S6. Details of computational implementation, algorithms, software and functions used

Method	**Numerical Algorithm**	**Software/Code**	**Time**
Direct regression	Maximum likelihood estimation	R build-in package stats function glm	1.3
Direct regression + Firth	Implement Firth's penalized likelihood on the basis of direct regression	R package logistf function logistif [1]	2.4
Non-informative Bayes	The regression of Y on X and B is based on Bayesian logistic regression with weakly informative Cauchy prior. The posterior draws are obtained based on an approximate EM algorithm	R package arm function bayesglm [2]	3.6
Constrained ML	Use Lagrange multipliers to optimize the joint log-likelihood under a set of constraints, with specified wide bounds for each parameter	R package Rsolnp function solnp [3]	44.9
Constrained ML + Firth	Adding a Firth penalty term to the objective function on the basis of constrained ML	R package Rsolnp function solnp	78.2
Informative full Bayes	A full Bayes algorithm that starts with re-parameterize the joint likelihood using a Jacobian transformation and then implements the full Bayesian inference based on the joint likelihood with the constraints directly used as priors. A random walk Metropolis-Hastings sampling algorithm is conducted to obtain draws from the posterior distribution	Self-written R code for Jacobian transformation, the implementation of full Bayesian inference based on the joint likelihood and the Metropolis-Hastings sampling algorithm	9097.6
Transformation	An approximate Bayes approach that starts with using draws from non-informative Bayes and standard Bayes. These draws are transformed into constrained draws by solving an optimization problem, which is to minimize the normalized Euclidean distance of the draws from the constrained space	Self-written R code for constructing the optimization problem, and simplifying the multi-dimensional optimization problem into one-dimensional optimization problem. Use R package arm for the non-informative Bayes and function optimize in R build-in package stats to solve the one-dimensional optimization problem	888.2
Chatterjee et al. method	Constrained maximum likelihood estimation using Lagrange multipliers	Self-written Newton Raphson algorithm	43.2

[1] Heinze,G., Ploner, M (2016). logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.22

[2] Gelman, A., Su, Y., Yajima, M., Hill, J., Pittau, M.G., Kerman, J., Zheng, T., Dorie, V. (2016). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.9-3

[3] Ghalanos, A., Theussl, S. Rsolnp (2013). Rsolnp: General Non-Linear Optimization. R package version 1.15

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006) *Measurement Error in Nonlinear Models: a modern perspective, Second edition*. Boca Raton, Florida: Chapman & Hall /CRC.

Chatterjee, N., Chen, Y.-H., Maas, P. and Carroll, R. J. (2016) Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, **111**, 107–117.

Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54–75.

Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.-S. (2008) A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, **2**, 1360–1383.

Grill, S., Fallah, M., Leach, R. J., Thompson, I. M., Hemminki, K. and Ankerst, D. P. (2015) A simple-to-use method incorporating genomic markers into prostate cancer risk prediction tools facilitated future validation. *Journal of Clinical Epidemiology*, **68**, 563–573.

Monahan, J. and Stefanski, L. A. (1992) *Normal scale mixture approximations to F*(z) and computation of the logistic-normal integral. in Handbook of the logistic distribution*. New York: CRC Press.

Satten, G. A. and Kupper, L. L. (1993) Inferences about exposure-disease associations using probability-of- exposure information. *Journal of the American Statistical Association*, **88**, 200–208.