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

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

University of Michigan, Ann Arbor, Michigan, USA

†

Summary. We consider a situation where there is rich historical data available for the coefficients and their standard errors in an established regression model describing the association between a binary outcome variable Y and a set of predicting factors $\mathbf{X}$, from a large study. We would like to utilize this summary information for improving estimation and prediction in an expanded model of interest, $\mathrm{Y}|\mathbf{X},\mathrm{B}$. The additional variable B is a new biomarker, measured on a small number of subjects in a new dataset. We develop and evaluate several approaches for translating the external information into constraints on regression coefficients in a logistic regression model of $\mathrm{Y}|\mathbf{X},\mathrm{B}$. Borrowing from the measurement error literature we establish an approximate relationship between the regression coefficients in the models $\Pr(\mathrm{Y}=1|\mathbf{X},\boldsymbol{\beta})$, $\Pr(\mathrm{Y}=1|\mathbf{X},\mathrm{B},\boldsymbol{\gamma})$ and $\mathrm{E}(\mathrm{B}|\mathbf{X},\boldsymbol{\theta})$ for a Gaussian distribution of B. For binary B we propose an alternate expression. The simulation results comparing these methods indicate that historical information on $\Pr(\mathrm{Y}=1|\mathbf{X},\boldsymbol{\beta})$ can improve the efficiency of estimation and enhance the predictive power in the regression model of interest $\Pr(\mathrm{Y}=1|\mathbf{X},\mathrm{B},\boldsymbol{\gamma})$. We illustrate our methodology by enhancing the High-grade Prostate Cancer Prevention Trial Risk Calculator, with two new biomarkers prostate cancer antigen 3 and TMPRSS2:ERG.

Keywords: Bayesian methods; Constrained estimation; Logistic regression; Prediction models

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

1. Introduction

Risk prediction models for different binary disease endpoints are abundant in the clinical and epidemiological literature. Examples of established models are the breast cancer risk calculator (Gail et al., 1989) and the Framingham risk score (D'Agostino et al., 2001) which can be used to assess an individual's risk of experiencing a future health event and to make decisions concerning screening and prophylactic prevention. As a motivating example in this paper, the Prostate Cancer Prevention Trial Risk Calculator (Thompson et al., 2006) is an online assessment tool which provides personalized risk estimate of detecting prostate cancer based on risk factors such as age, prostate-specific antigen (PSA) and digital rectal examination (DRE) findings.

While these established models are often based on standard epidemiologic and behavioral risk factors, wider availability of high throughput data and novel assay technologies are generating candidate biomarkers for potential inclusion in existing prediction models. It's very likely that the new biomarkers are assessed only on subjects in a study of moderate size and cannot be measured on the much larger population used for the well-established model. Investigators could directly estimate the expanded model in the new dataset, but results from this expanded prediction model based solely on a limited number of subjects could be highly variable. It is natural to consider using the information from the well-established model to increase the accuracy of the expanded model.

Substantial research has been done on the problem of enhancing risk prediction models with supplemental external information. The external information may be used to combine estimates from previous studies with the regression coefficients estimated in the new dataset. Steyerberg et al. (2000) described a method to adjust the multivariate logistic regression model's coefficients estimated in a dataset based on univariate regression models' coefficients in the literature. Newcombe et al. (2012) presented two possible approaches incorporating the effect estimates of a set of predictors: the first one was by adding a composite weighted risk score based on these estimates and the second one was by specifying informative priors for the coefficients of these variables in a Bayesian logistic regression model. Chatterjee et al. (2016) developed a general method for incorporating external coefficients, derived from constrained estimating equations. Other related approaches used constrained maximum likelihood and empirical likelihood (Imbens and Lancaster, 1994; Qin, 2000; Qin et al., 2015). Cheng et al. (2018) developed and compared a number of approaches for the situation when the outcome variable is continuous. They established exact relationships between the parameters in the model of interest that includes the new biomarker and the parameters in the established model, then proposed both frequentist and Bayesian approaches. In the current paper we adapt and extend the approaches to the situation when the outcome variable is binary.

There are also a number of simple approaches. For the Gail model, Mealiffe et al. (2010) computed a multiplicative risk score based on previously published odds ratios of newly discovered biomarkers. They then multiplied the Gail risk estimate and the multiplicative risk score to give a combined risk score. Grill et al. (2015) proposed a simple method of incorporating new markers via Bayes Theorem. They updated the posterior odds of getting cancer based on both standard risk factors and new markers by using the likelihood ratio incorporating dependence between the two sets of risk factors to adjust the prior odds of getting cancer based on standard risk alone. Grill et al. (2017) assessed the performance of a set of likelihood ratio approaches as well as the approach proposed in Chatterjee et al. (2016).

We consider a situation where the outcome is a binary indicator of disease and the well-established model is described in a published article, in which the estimated regression coefficients and their standard errors are presented. The expanded model includes one additional biomarker as a potential predictor. To introduce notation, let Y denote the binary outcome, **X** is a set of p standard risk factors and B is a new biomarker. The association between Y and **X** is described through the following logistic model:

$$\mathrm{logit}(\Pr(\mathrm{Y}=1|\mathbf{X})) = \mathbf{X}\boldsymbol{\beta} = \beta_0 + \beta_1\mathrm{X}_1 + \cdots + \beta_\mathrm{p}\mathrm{X}_\mathrm{p} \tag{1}$$

We assume we have available summary-level information on the estimated regression coefficients $\bar{\boldsymbol{\beta}}$ and their standard errors $\bar{\mathbf{S}}$ in model (1). Based on the work that went into establishing this model, we will assume that all the **X**'s are deemed to be important and need to be included in any model, and further that the above form provides at least a good approximation to the distribution of Y given $\mathbf{X}$.

The model of primary interest is one that describes the joint effect of $\mathbf{X}, \mathrm{B}$ on Y:

$$\mathrm{logit}(\Pr(\mathrm{Y}=1|\mathbf{X},\mathrm{B})) = \mathbf{X}\boldsymbol{\gamma}_{\mathrm{X}} + \mathrm{B}\gamma_{\mathrm{B}} = \gamma_0 + \gamma_1\mathrm{X}_1 + \cdots + \gamma_{\mathrm{p}}\mathrm{X}_{\mathrm{p}} + \gamma_{\mathrm{p}+1}\mathrm{B} \qquad (2)$$

Our goal is to obtain the best estimate we can of the γ's in a model of this form, making use of all the available information from the established model and the small dataset.

Another model that can be estimated from the current small dataset is:

$$\mathrm{E}(\mathrm{B}|\mathbf{X}) = \mathrm{g}^{-1}(\mathbf{X}\boldsymbol{\theta}) = \mathrm{g}^{-1}(\theta_0 + \theta_1\mathrm{X}_1 + \cdots + \theta_{\mathrm{p}}\mathrm{X}_{\mathrm{p}}) \qquad (3)$$

where g is the link function, which is the identity function $\mathrm{g}(\mathrm{y}) = \mathrm{y}$ for Gaussian B and the logit function $\mathrm{g}(\mathrm{y}) = \log(\mathrm{y}/(1-\mathrm{y}))$ for binary B. We propose to formulate the problem in an inferential framework where the historical information is translated in terms of non-linear constraints on the regression parameters. The distribution of B will greatly affect how we translate the historical information into constraints on the regression parameters. We consider the cases that B is either Gaussian or binary.

The following is the structure of the remainder of this article: in Section 2 we describe the Prostate Cancer Prevention Risk Calculator and the available data including the new biomarkers that might be able to enhance this calculator. In Section 3, we establish a relationship equation between the regression coefficients of models (1) - (3) when B is Gaussian. In Section 4, we consider the situation when B is binary and derive the corresponding constrained solutions. We present a simulation study in Section 5. In Section 6 we demonstrate the proposed approaches for the High-grade Prostate Cancer Prevention Trial Risk Calculator. Concluding remarks are presented in Section 7.

2. A Motivating Example: Prostate Cancer Risk Prediction

The Prostate Cancer Prevention Trial (PCPT) was a phase III randomized placebo-controlled trial of drug finasteride for the prevention of carcinoma of the prostate. The PCPT randomly assigned about 18882 men who were at least 55 years old and did not have prostate cancer to either finasteride or placebo for 7 years. At the end of the 7 years of the study, all men who had not been diagnosed with prostate cancer during the trial were asked to undergo an end-of-study prostate biopsy. The biopsy result could be no cancer, low-grade cancer or high-grade cancer, which was defined as Gleason score of 7 or higher. Variables collected in this trial included family history of prostate cancer, age, race, previous biopsy result, PSA and digital rectal examination.

The use of PSA to screen for prostate cancer (PCa) had been controversial because the test has low specificity and can lead to overtreatment. Therefore, improved tests that use additional information are needed. The Prostate Cancer Prevention Trial Risk Calculator (PCPTrc) for prostate cancer, and a separate calculator for high-grade prostate cancer (PCPThg) (Thompson et al., 2006) were the first online prostate cancer risk assessment tools to allow an individual to assess his risk for prostate cancer. These calculators are well established and are frequently used. These calculators were developed from 5519 men in the placebo group of the PCPT who underwent prostate biopsy. The PCPThg calculator (version 1.0) predicts the chance of high-grade prostate cancer based on PSA level, age, DRE findings, prior biopsy result and race:

$$\mathrm{log}(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}) = -6.25+0.03\mathrm{age_i}+0.96\mathrm{race_i}+1.29\mathrm{log}(\mathrm{PSA_i})+1.00\mathrm{DRE_i}-0.36\mathrm{biopsy_i} \quad (4)$$

where p_i is the probability of observing high grade prostate cancer for subject i. If we plug in a person's age, race, PSA level, DRE result and previous biopsy information, we can estimate the probability of detecting high-grade prostate cancer. The estimated logistic models coefficients and the 95% confidence intervals are available in Thompson et al. (2006). The estimated coefficients and covariance-variance matrices were also accessible as a R code document at (http://deb.uthscsa.edu/URORiskCalc/Pages/calcs.jsp).

The PCPT risk calculators are based on standard clinical, demographic and epidemiologic variables. None of the variables are related to the molecular mechanisms of carcinogenesis or prostate cancer disease progression. It is plausible to think that including other variables that are more related to the biology of cancer would lead to improved ability to detect PCa. Prostate cancer antigen 3 (PCA3) and TMPRSS2:ERG (T2:ERG) gene fusions are two prostate cancer biomarkers which have been shown to have better specificity for early detection of PCa than PSA (Truong et al., 2013; Tomlins et al., 2015). Their transcripts are detectable and quantifiable in urine collected after digital rectal examination. To investigate whether PCA3 and T2:ERG could be combined with the PCPThg calculator to give more accurate risk prediction, Tomlins et al. (2015) undertook a study in 679 men, in whom all the PCPThg calculator variables and both a PCA3 score and a T2:ERG score were measured. In this dataset the proportion with high grade PCa is 26.4%. An independent validation study of 1218 men was also available. In this dataset the proportion with high grade PCa is 18.3%.

Tomlins et al. (2015) expanded the PCPThg model by incorporating PCA3 as an additional risk factor. They used the predicted risk score from the PCPThg (i.e., $\hat{\Pr}(\mathrm{Y_i} = 1|\mathrm{X_i}, \bar{\beta}_{\mathrm{PCPThg}}) \times 100$) directly as a predicting variable and estimated the joint effect of the PCPThg risk score and the PCA3 value on the probability of high-grade PCa. They estimated the new model in the training dataset, and found that when applied to the validation dataset the AUC increased from 0.707 for the PCPThg model to 0.752 for their model. They also constructed another expanded PCPThg model by incorporating T2:ERG and showed that the AUC increased from 0.707 to 0.754. We would like to propose more sophisticated statistical approaches that could potentially provide further improvement compared to these results.

3. Statistical Approaches

3.1. Logistic Regression Approximation of the Marginal $\Pr(\mathrm{Y} = 1|\mathbf{X})$

A difficulty in translating the summary information from modeling $\Pr(\mathrm{Y} = 1|\mathbf{X})$ to modeling $\Pr(\mathrm{Y} = 1|\mathbf{X}, \mathrm{B})$ is that a logistic model $\mathrm{logit}(\Pr(\mathrm{Y} = 1|\mathbf{X}, \mathrm{B}))$ does not reduce to a logistic model $\mathrm{logit}(\Pr(\mathrm{Y} = 1|\mathbf{X}))$ when marginalized over the distribution of B. To connect the regression coefficients in models (1), (2) and (3), we need to approximate $\mathrm{logit}(\Pr(\mathrm{Y} = 1|\mathbf{X}))$ written in terms of parameters $\boldsymbol{\gamma}, \boldsymbol{\theta}$ and variables $\mathbf{X}$, and equate that to $\mathrm{logit}(\Pr(\mathrm{Y} = 1|\mathbf{X})) = \mathbf{X}\boldsymbol{\beta}$. To achieve this, we consider the following integral:

$$\begin{aligned}\Pr(\mathrm{Y} = 1|\mathbf{X}) &= \int \Pr(\mathrm{Y} = 1|\mathbf{X}, \mathrm{B})\mathrm{P}(\mathrm{B}|\mathbf{X})\mathrm{dB} \\ &= ((2\pi)^{1/2}\sigma_2)^{-1} \int \mathrm{H}(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}} + \mathrm{B}\boldsymbol{\gamma}_{\mathrm{p+1}}) \exp\Big(- \frac{(\mathrm{B} - \mathbf{X}\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{B} - \mathbf{X}\boldsymbol{\theta})}{2\sigma_2^2}\Big)\mathrm{dB}\end{aligned} \tag{5}$$

where $\mathrm{H(v)} = (1+\exp(-\mathrm{v}))^{-1}$, and $\mathrm{B}|\mathbf{X}$ follows a Gaussian distribution $\mathrm{N}(\mathbf{X}\boldsymbol{\theta}, \sigma_2^2)$. This integral in (5) does not have a closed-form solution and approximations are necessary.

By a normal scale mixture representation of the logistic distribution function and computation of the logistic-normal integral (Monahan and Stefanski, 1992), we can find an approximated equation: $\mathrm{Pr}(\mathrm{Y}=1|\mathbf{X}) \approx \mathrm{H}\left(\frac{\mathbf{X}\boldsymbol{\gamma}_{\mathrm{x}}+(\mathbf{X}\boldsymbol{\theta})\boldsymbol{\gamma}_{\mathrm{p}+1}}{(1+\gamma_{\mathrm{p}+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}}\right)$. The derivation of the approximation is given in Supplementary Material Appendix B. Using this approximation, we find an approximate relationship between $\boldsymbol{\gamma}, \boldsymbol{\theta}$ and $\boldsymbol{\beta}$:

$$\beta_j \approx (\gamma_j + \gamma_{p+1}\theta_j)/((1+\gamma_{p+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}), j = 0, \ldots, p. \tag{6}$$

3.2. Firth Correction in Logistic Regression

The Firth correction (Firth, 1993) is a general approach to reduce bias in maximum likelihood estimation by maximizing a penalized log-likelihood function, where the penalty is $\frac{1}{2}|\mathbf{I}|$ and $\mathbf{I}$ is the information matrix. In logistic regression, standard maximum likelihood estimates often experience serious bias or even non-existence due to separability and multicollinearity, and the Firth correction is suggested (Heinze and Schemper, 2002) as a way to improve the estimates. In our constrained solution, we add the Firth correction to stabilize the estimates from standard logistic regression.

3.3. Unconstrained Solutions

3.3.1. Direct Regression

Without constraints, the unknown parameters $\boldsymbol{\gamma}$ in model (2) can be estimated by maximizing the likelihood. Specifically, the estimate solves

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{\mathrm{i}=1}^{\mathrm{n}} [\mathrm{Y_i}(\sum_{\mathrm{j}=0}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p}+1}\mathrm{B_i}) - \log(1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p}+1}\mathrm{B_i}))] \right\} \tag{7}$$

In addition, we implement Firth's penalized likelihood approach using R package **logistf**. We use least squares to estimate $\boldsymbol{\theta}$ in model (3).

3.3.2. Standard Bayes

Draws for the posterior distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are obtained using flat conjugate priors for $\boldsymbol{\theta}$ and weakly informative Cauchy distribution priors for $\boldsymbol{\gamma}$, as described in the Supplementary Materials Appendix A.

3.4. Constrained Solutions

3.4.1. Constrained Maximum Likelihood

The constrained maximum likelihood (constrained ML) estimation maximizes the joint log-likelihood under the set of constraints generated based on the approximate relationship equations in (6). We will require the parameter estimates for $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ to result in the derived value of $\boldsymbol{\beta}$ being within d standard errors of the old point estimates:

$$\begin{aligned}&\min_{\boldsymbol{\gamma},\boldsymbol{\theta}}\Big\{\sum_{i=1}^{n}[-\mathrm{Y_i}(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})+\log(1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i}))]\\&\qquad+\sum_{\mathrm{i=1}}^{\mathrm{n}}\frac{(\mathrm{B_i}-\sum_{\mathrm{j=0}}^{\mathrm{p}}\theta_{\mathrm{j}}\mathrm{X_{ij}})^2}{2\hat{\sigma}_2^2}\Big\}\\&\text{s.t.}(\gamma_j+\gamma_{p+1}\theta_j)/(1+\gamma_{p+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}}\in[\bar{\beta}_j-d\bar{S}_j,\bar{\beta}_j+d\bar{S}_j],j=0,\ldots,p\end{aligned}\tag{8}$$

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. From simulations, we find that fixing d as $d = 1$ leads to decent properties of the estimates of $\boldsymbol{\gamma}$. A modified version that includes the Firth correction is also considered. Further details of these methods are provided in the Supplementary Materials Appendix A. We use the bootstrap as described in Supplementary Material Appendix D to estimate the standard errors.

3.4.2. Informative Full Bayes

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

$$\begin{aligned}&p(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2|\text{data})\propto\mathrm{L(Y|}\mathbf{X},\mathrm{B},\boldsymbol{\gamma})\cdot\mathrm{L(B|}\mathbf{X},\boldsymbol{\theta},\sigma_2^2)\cdot\pi(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2)\\&=\Big\{\prod_{\mathrm{i=1}}^{\mathrm{n}}\frac{\exp((\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})\mathrm{Y_i})}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})}\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\big(-\frac{1}{2\sigma_2^2}(\mathrm{B_i}-\sum_{\mathrm{j=0}}^{\mathrm{p}}\theta_{\mathrm{j}}\mathrm{X_{ij}})^2\big)\Big\}\cdot\pi(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma_2^2)\end{aligned}\tag{9}$$

The logistic regression approximation result (6) 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 re-parametrize (9) in terms of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and σ_2^2, and include a Jacobian transformation matrix denoted by $\mathbf{J}$, where $|\mathbf{J}| = \frac{1}{|\gamma_{p+1}^{p+1}|}(1 + \frac{\gamma_{p+1}^2\sigma_2^2}{1.7^2})^{\frac{p+1}{2}}$. We use a non-informative prior inverse-gamma for σ_2^2 and independent weakly informative Cauchy priors for $\boldsymbol{\gamma}$ (Gelman et al., 2008). For parameters $\boldsymbol{\beta}$, we use the constraints as priors:

$$\beta_\mathrm{j} = (\gamma_\mathrm{j} + \gamma_\mathrm{p+1}\theta_\mathrm{j})/(1 + \gamma_\mathrm{p+1}^2\sigma_2^2/1.7^2)^{\frac{1}{2}} \sim \mathrm{N}(\bar{\beta}_\mathrm{j}, \bar{\mathrm{S}}_\mathrm{j}^2), \mathrm{j} = 0, \ldots, \mathrm{p} \qquad (10)$$

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

Further details of the priors and the implementation of a Metropolis-Hastings algorithm are given in the Supplementary Materials Appendix B. We note that in the algorithm the full conditional distributions of $\gamma_0, \ldots, \gamma_{p+1}$ and σ_2^2 do not have closed form expressions, and because of the non-linear relationship between the parameters, the Metropolis-Hasting algorithm cannot be performed efficiently and thus it is computationally slow to obtain uncorrelated draws from the posterior distributions.

3.4.3. Transformation Approach

As the informative full Bayes is computationally intensive, we propose an approximate Bayesian approach that can produce draws that fall into the constrained space but reduces the computational burden of the informative Bayes method. The motivation for this stems from the Bayesian transformation approach incorporating monotone or unimodal constraints in posterior inference as proposed in Gunn and Dunson (2005), which we modify to the scenario of a regression model with external coefficient information.

Suppose the draws from non-informative standard Bayes method as described in Section 3.3.2 are $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$. The corresponding posterior covariance matrices are $\boldsymbol{\Sigma_\gamma}, \boldsymbol{\Sigma_\theta}$. We extract the posterior variances from $\boldsymbol{\Sigma_\gamma}, \boldsymbol{\Sigma_\theta}$ and denote them by $s_{\gamma_0}^2, \ldots, s_{\gamma_p}^2, s_{\gamma_{p+1}}^2, s_{\theta_0}^2, \ldots, s_{\theta_p}^2$. The OLS residual variance from fitting $\mathrm{E(B|}\mathbf{X})$ is $\hat{\sigma}_2^2$. Then a constrained draw $\boldsymbol{\gamma}^\star, \boldsymbol{\theta}^\star$ is obtained from an unconstrained draw by solving the optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\gamma}^\star, \boldsymbol{\theta}^\star} \quad & \{\mathrm{d}_\mathrm{NED}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}^\star) + \mathrm{d}_\mathrm{NED}^2(\boldsymbol{\theta}, \boldsymbol{\theta}^\star)\} = \sum_\mathrm{j=0}^\mathrm{p+1} \frac{(\gamma_\mathrm{j} - \gamma_\mathrm{j}^\star)^2}{\mathrm{s}_{\gamma_\mathrm{j}}^2} + \sum_\mathrm{k=0}^\mathrm{p} \frac{(\theta_\mathrm{k} - \theta_\mathrm{k}^\star)^2}{\mathrm{s}_{\theta_\mathrm{k}}^2} \\ \text{s.t.} \quad & (\gamma_j^\star + \gamma_{p+1}^\star\theta_j^\star)/(1 + \gamma_{p+1}^{\star 2}\hat{\sigma}_2^2/1.7^2)^{\frac{1}{2}} \in [\bar{\beta}_j - d\bar{S}_j, \bar{\beta}_j + d\bar{S}_j], j = 0, \ldots, p \end{aligned} \qquad (11)$$

where $\mathrm{d_{NED}}$ stands for normalized Euclidean distance. For the transformation of a single draw, we generate d from a half normal distribution: $d \sim |\mathrm{N}(0,1)|$. The intuition behind this transformation procedure is that it will produce values $\boldsymbol{\gamma}^{\star}, \boldsymbol{\theta}^{\star}$ subject to the box constraints that are closest in normalized distance to the unconstrained values $\boldsymbol{\gamma}, \boldsymbol{\theta}$.

The transformation is computationally efficient since we have a fast algorithm to solve the optimization problem in (11). We fix $\gamma^{\star}_{p+1}$ and divide the minimization function (11) into $p+1$ two-dimensional constrained minimization problems in which the solutions can be re-expressed as functions of $\gamma^{\star}_{p+1}$. After that, the whole minimization problem is reduced to an easy to solve one-dimensional optimization problem in $\gamma^{\star}_{p+1}$. The constrained draws produced by the transformation approach are not draws from the true posterior distribution, however, we found in a limited number of simulations that they are reasonable approximations that can be generated much faster.

3.4.4. Constrained Approach of Chatterjee et al (2016)

For comparison we include a constrained maximum likelihood method that uses the integrated score equations (Chatterjee et al., 2016). The method assumes the model for $\mathrm{Y}|\mathbf{X}, \mathrm{B}$ is correct, it does not make any explicit assumptions about the distribution of $\mathrm{B}|\mathbf{X}$, but it does require the distribution of $\mathbf{X}$ to be the same in the current sample as in the data that was used to develop the model for $\mathrm{Y}|\mathbf{X}$. The method uses only the point estimates $\bar{\beta}$ and does not take into account the standard errors of those estimates.

3.4.5. Logistic Regression Plug-in Method

We also included a simple method which consists of obtaining predicted probabilities by fitting a logistic regression model with two covariates, B and $\log(\mathrm{\bar{p}_i}/(1-\mathrm{\bar{p}_i})$, where $\bar{p}_i$ is the prediction from the established model for $\mathrm{Y}|\mathbf{X}$. It is easy to show that this method does give a final model for $\mathrm{Y}|\mathbf{X}, \mathrm{B}$ that has a logistic link function and is linear in $\mathbf{X}$ and B, and with some algebra the estimates of γ can be obtained.

4. Statistical Approaches when B is Binary

4.1. The Approximate Relationship Equation When B is Binary

If B is a binary variable, the logistic regression approximation in Section 3 does not hold and the approximate relationship in equation (6) is not applicable. However, by Bayes theorem, there is a relationship equation connecting $\mathrm{Pr(Y=1|\mathbf{X}), Pr(Y=1|\mathbf{X},B)}$ and $\mathrm{f(B|\mathbf{X},Y)}$ (Grill et al., 2015; Satten and Kupper, 1993):

$$\frac{\mathrm{Pr(Y=1|\mathbf{X},B)}}{\mathrm{Pr(Y=0|\mathbf{X},B)}} = \frac{\mathrm{f(B|\mathbf{X},Y=1)}}{\mathrm{f(B|\mathbf{X},Y=0)}} \cdot \frac{\mathrm{Pr(Y=1|\mathbf{X})}}{\mathrm{Pr(Y=0|\mathbf{X})}} \tag{12}$$

Thus, when B is binary, we need to define a model for $\mathrm{B|\mathbf{X},Y}$ instead of a model for $\mathrm{B|\mathbf{X}}$. Assume $\mathrm{logit(Pr(B=1|\mathbf{X},Y))} = \sum_{\mathrm{j=0}}^{\mathrm{p}} \phi_{\mathrm{j}}\mathbf{X}_{\mathrm{j}} + \phi_{\mathrm{p+1}}\mathrm{Y}$. Some algebraic simplifications of equation (12) followed by a Taylor series expansion (as shown in Supplementary Material Appendix C) result in an approximate relationship equation: $\beta_0 + \beta_1\mathrm{X}_1 + \cdots + \beta_{\mathrm{p}}\mathrm{X}_{\mathrm{p}} \approx \gamma_0 + \frac{1}{2}\phi_{\mathrm{p+1}} + \frac{1}{4}\phi_0\phi_{\mathrm{p+1}} + \frac{1}{8}\phi_{\mathrm{p+1}}^2 + \sum_{\mathrm{j=1}}^{\mathrm{p}}(\gamma_{\mathrm{j}} + \frac{1}{4}\phi_{\mathrm{j}}\phi_{\mathrm{p+1}})\mathrm{X}_{\mathrm{j}} + (\gamma_{\mathrm{p+1}} - \phi_{\mathrm{p+1}})\mathrm{B}$. Then the approximate relationship between $\boldsymbol{\gamma}, \boldsymbol{\phi}$ and $\boldsymbol{\beta}$ is:

$$\begin{cases} \beta_0 \approx \gamma_0 + \frac{1}{2}\phi_{\mathrm{p+1}} + \frac{1}{4}\phi_0\phi_{\mathrm{p+1}} + \frac{1}{8}\phi_{\mathrm{p+1}}^2 \\ \beta_j \approx \gamma_j + \frac{1}{4}\phi_j\phi_{p+1}, j = 1, \ldots, p \\ \gamma_{p+1} = \phi_{p+1} \end{cases} \tag{13}$$

4.2. Unconstrained and Constrained Solutions

The two unconstrained solutions, direct regression and standard Bayes can be performed in the same way as described in Section 3 regardless of the distribution of B.

To develop a constrained solution, we need to first define the likelihood function $\mathrm{L(B|\mathbf{X})}$. It can be shown that $\mathrm{Pr(B_i=1|X_i)}$ is a weighted average of $\frac{\exp(\sum_{j=0}^{p} X_{ij}\phi_j)}{1+\exp(\sum_{j=0}^{p} X_{ij}\phi_j)}$ and $\frac{\exp(\sum_{j=0}^{p} X_{ij}\phi_j+\phi_{p+1})}{1+\exp(\sum_{j=0}^{p} X_{ij}\phi_j+\phi_{p+1})}$. We use estimates of the weights given by $w_{i,\bar{\boldsymbol{\beta}}} = \frac{1}{1+\exp(\mathbf{X}_i\bar{\boldsymbol{\beta}})}$ and $1 - w_{i,\bar{\boldsymbol{\beta}}} = \frac{\exp(\mathbf{X}_i\bar{\boldsymbol{\beta}})}{1+\exp(\mathbf{X}_i\bar{\boldsymbol{\beta}})}$ where $\bar{\boldsymbol{\beta}}$ are the values for $\boldsymbol{\beta}$ from the established model. Then $\mathrm{L(B|\mathbf{X})}$ can be written as: $\mathrm{L(B|\mathbf{X})} = \prod_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{L(B_i|\mathbf{X_i}}, \boldsymbol{\phi}) = \prod_{\mathrm{i=1}}^{\mathrm{n}} \big(\frac{\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}})}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}})} \cdot \mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}} + \frac{\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})} \cdot (1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})\big)^{\mathrm{B_i}} \cdot \big(\frac{1}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}})} \cdot \mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}} + \frac{1}{1+\exp(\sum_{\mathrm{j=0}}^{\mathrm{p}} \mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})} \cdot (1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})\big)^{(1-\mathrm{B_i})}$.

4.2.1. Constrained Maximum Likelihood

The constrained ML estimation optimizes the following joint log-likelihood $\mathrm{L(Y|\mathbf{X},B)L(B|\mathbf{X})}$ with a set of constraints on $\boldsymbol{\gamma}, \boldsymbol{\phi}$, namely:

$$\begin{aligned}\min_{\boldsymbol{\gamma},\boldsymbol{\phi}} \Big\{ &\sum_{\mathrm{i}=1}^{\mathrm{n}} [-\mathrm{Y_i}(\sum_{\mathrm{j}=0}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p+1}}\mathrm{B_i}) + \log(1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}} \gamma_{\mathrm{j}}\mathrm{X_{ij}} + \gamma_{\mathrm{p+1}}\mathrm{B_i}))] \\ &-\sum_{\mathrm{i}=1}^{\mathrm{n}} \Big[\mathrm{B_i}\log\Big(\frac{\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}}}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})} + \frac{\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})(1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})}\Big) \\ &+(1-\mathrm{B_i})\log\Big(\frac{\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}}}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})} + \frac{(1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})}\Big)\Big]\Big\}\end{aligned} \tag{14}$$

$$\text{s.t.}\begin{cases}\gamma_0+\frac{1}{2}\phi_{p+1}+\frac{1}{4}\phi_0\phi_{p+1}+\frac{1}{8}\phi_{p+1}^2 \in [\bar{\beta}_0-d\bar{S}_0, \bar{\beta}_0+d\bar{S}_0] \\ \gamma_j+\frac{1}{4}\phi_j\phi_{p+1} \in [\bar{\beta}_j-d\bar{S}_j, \bar{\beta}_j+d\bar{S}_j], j=1,\ldots,p \\ \gamma_{p+1}=\phi_{p+1}\end{cases}$$

We also consider a modification that adds the Firth penalty term.

4.2.2. Informative Full Bayes

Analogous to the derivation of the informative full Bayes solution described in Section 3, we first write down the product of $\mathrm{L(Y|\mathbf{X},B)}$ and $\mathrm{L(B|\mathbf{X})}$ with priors.

$$\begin{aligned}\mathrm{p}(\boldsymbol{\gamma},\boldsymbol{\phi}|\mathrm{data}) \propto \Big\{ &\prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\exp((\sum_{\mathrm{j}=0}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})\mathrm{Y_i})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})}\cdot \\ &\Big(\frac{\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}}}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})} + \frac{\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})(1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})}\Big)^{\mathrm{B_i}}\cdot \\ &\Big(\frac{\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}}}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}})} + \frac{(1-\mathrm{w}_{\mathrm{i},\bar{\boldsymbol{\beta}}})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\mathrm{X_{ij}}\phi_{\mathrm{j}}+\phi_{\mathrm{p+1}})}\Big)^{(1-\mathrm{B_i})}\Big\}\cdot\pi(\boldsymbol{\gamma},\boldsymbol{\phi})\end{aligned} \tag{15}$$

We can re-parametrize (15) in terms of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and include a Jacobian corresponding to this transformation. We denote the Jacobian matrix by $\mathbf{M}$ where $|\mathbf{M}| = |\frac{4}{\gamma_{p+1}}|^{p+1}$. We specify independent weakly informative Cauchy priors for $\boldsymbol{\gamma}$ and use the constraints directly as priors for $\boldsymbol{\beta}$. Then similar to section 3.4.2 we can rewrite the joint distribution in terms of $\boldsymbol{\gamma}, \boldsymbol{\beta}$ as $\mathrm{p}(\boldsymbol{\gamma},\boldsymbol{\beta}|\mathrm{Y},\mathbf{X},\mathrm{B}) \propto \Big\{ \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\exp((\sum_{\mathrm{j}=0}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})\mathrm{Y_i})}{1+\exp(\sum_{\mathrm{j}=0}^{\mathrm{p}}\gamma_{\mathrm{j}}\mathrm{X_{ij}}+\gamma_{\mathrm{p+1}}\mathrm{B_i})}\cdot$

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

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

$$\pi(\boldsymbol{\beta})\cdot\pi(\boldsymbol{\gamma})\cdot|\mathbf{M}|$$

The full conditional distributions of $\gamma_0,\ldots,\gamma_{p+1}$ and $\beta_0,\ldots,\beta_p$ do not have closed form expressions and a Metropolis-Hastings algorithm is used.

4.2.3. Transformation Approach

Suppose the raw draws from the non-informative Bayes method for $\mathrm{Y}|\mathbf{X},\mathrm{B}$ are $\boldsymbol{\gamma}$ and the raw draws from non-informative Bayes method for $\mathrm{B}|\mathbf{X},\mathrm{Y}$ are $\boldsymbol{\phi}$. The posterior variances are $s_{\gamma_j}^2, j=0,\ldots,p+1$ and $s_{\phi_k}^2, k=0,\ldots,p+1$. Then $\boldsymbol{\gamma}^\star,\boldsymbol{\phi}^\star$ are obtained by solving the following optimization problem:

$$\min_{\boldsymbol{\gamma}^\star,\boldsymbol{\phi}^\star}\{\mathrm{d}_{\mathrm{NED}}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}^\star)+\mathrm{d}_{\mathrm{NED}}^2(\boldsymbol{\phi},\boldsymbol{\phi}^\star)\}=\sum_{\mathrm{j=0}}^{\mathrm{p+1}}\frac{(\gamma_{\mathrm{j}}-\gamma_{\mathrm{j}}^\star)^2}{\mathrm{s}_{\gamma_{\mathrm{j}}}^2}+\sum_{\mathrm{k=0}}^{\mathrm{p+1}}\frac{(\phi_{\mathrm{k}}-\phi_{\mathrm{k}}^\star)^2}{\mathrm{s}_{\phi_{\mathrm{k}}}^2}$$

$$\text{s.t.}\begin{cases}\gamma_0^\star+\frac{1}{2}\phi_{p+1}^\star+\frac{1}{4}\phi_0^\star\phi_{p+1}^\star+\frac{1}{8}\phi_{p+1}^{\star 2}\in[\bar{\beta}_0-d\bar{S}_0,\bar{\beta}_0+d\bar{S}_0]\\ \gamma_{\mathrm{j}}^\star+\frac{1}{4}\phi_{\mathrm{j}}^\star\phi_{\mathrm{p+1}}^\star\in[\bar{\beta}_{\mathrm{j}}-\mathrm{d}\bar{\mathrm{S}}_{\mathrm{j}},\bar{\beta}_{\mathrm{j}}+\mathrm{d}\bar{\mathrm{S}}_{\mathrm{j}}],\mathrm{j}=1,\ldots,\mathrm{p}\\ \gamma_{\mathrm{p+1}}^\star=\phi_{\mathrm{p+1}}^\star\end{cases}\tag{16}$$

5. Simulation Study

To evaluate the performance of the various approaches we conduct a simulation study. The results for Gaussian B are presented here, and the results for binary B and other scenarios are presented in the Supplementary Materials Appendix E. The simulation scenario has three predicting variables, $\mathrm{X}_1,\mathrm{X}_2$ and B and the sample size of each dataset is 55. Five hundred replicate datasets are generated. $\mathrm{Y_i}$ is Bernoulli distributed with $\mathrm{logit}(\Pr(\mathrm{Y_i}=1|\mathrm{X_{i1}},\mathrm{X_{i2}},\mathrm{B_i}))=2+3\mathrm{X_{i1}}+3\mathrm{X_{i2}}+2\mathrm{B_i}$. $\mathrm{X_{i1}},\mathrm{X_{i2}}$ are independently and identically distributed on $\mathrm{U}(-0.75,0.25)$ and $\mathrm{B_i}$ is simulated as $\mathrm{B_i}=0.5\mathrm{X_{i1}}+0.5\mathrm{X_{i2}}+\mathrm{N}(0,0.75^2)$. A logistic regression based on a large dataset of 10000 subjects gives estimates for the model $\mathrm{logit}(\Pr(\mathrm{Y}=1|\mathbf{X}))=\beta_0+\beta_1\mathrm{X}_1+\beta_2\mathrm{X}_2$.

The estimates and standard errors from this fit are $\bar{\beta}_0 = 1.50, \bar{S}_0 = 0.04, \bar{\beta}_1 = 2.95, \bar{S}_1 = 0.09, \bar{\beta}_2 = 3.01, \bar{S}_2 = 0.09$.

We report three evaluation metrics related to estimation accuracy: the average of estimated coefficient, relative efficiency of estimated coefficient and mean squared error across 500 replicates. The average of the estimated coefficients is defined as: $\bar{\gamma}_j = \frac{1}{500}\sum_{m=1}^{500}\hat{\gamma}_{m,j}$; the relative efficiency of the estimated coefficients is defined as: $\frac{\mathrm{V}(\hat{\gamma}_{\mathrm{j,direct}})}{\mathrm{V}(\hat{\gamma}_{\mathrm{j,method}})}$ where $\mathrm{V}(\hat{\gamma}_{\mathrm{j,direct}}) = \frac{1}{500}\sum_{\mathrm{m=1}}^{500}(\hat{\gamma}_{\mathrm{m,j}} - \bar{\gamma}_{\mathrm{j}})^2$ estimated by direct regression; the MSE of the estimated coefficients is defined as: $\frac{1}{500}\sum_{\mathrm{m=1}}^{500}(\hat{\gamma}_{\mathrm{m,j}} - \gamma_{\mathrm{j}})^2, \mathrm{j} = 1, \ldots, \mathrm{p}+1$.

The predictive ability of logistic prediction models can be assessed using a variety of methods and metrics on a validation dataset (Steyerberg et al., 2010). In this simulation study, we assess the predictive ability of the model on a validation dataset of size 800 using the scaled Brier score $(\frac{\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{Y_i} - \hat{\mathrm{p}}_{\mathrm{i}})^2}{\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{Y_i} - \bar{\mathrm{Y}})^2})$. We assess the variability of the predictions on the validation dataset using the standard deviation of the predicted probabilities, $((1/799)\sum(\hat{p}_i - \bar{\hat{p}})^2)^{1/2}$. The discriminatory performance of the model is assessed using the area under the ROC curve (AUC). These three performance measures are also calculated for the model based on $\bar{\beta}$ which does not use B, and for the best achievable model that uses the true values of the γ's.

Table 1 presents the results. In this setting what is achievable with the true model is noticeably better, as measured by Brier score and AUC, than using the established model, and the established model does not give the correct standard deviation of the predicted probabilities. There is bias in $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\gamma}_3$ for the direct regression approach, which is reduced by including the Firth penalty. We find that the constrained methods greatly improve the estimation efficiency of the coefficients γ_1 and γ_2 of $\mathbf{X}$. The constrained methods can reduce the MSE of $\mathrm{X}_1, \mathrm{X}_2$ by 70% or more, and give 4 or 5 times more efficient point estimates of γ_1 and γ_2. In contrast, these constrained solutions can only reduce the MSE of γ_3 by 10% to 40% and generally have similar efficiency to that of direct regression plus Firth. The new methods give similar variability of the predicted probabilities compared to the true model. All the methods that use the external information give similar Brier scores and AUCs which are very similar to the best that can be achieved. They are all better than not using B at all, and slightly better than the methods that don't use the external information. In terms of computational efficiency, the informative prior Bayesian approach and the transformation approach require more time than the other methods, however the transformation approach takes about 18% the time of the informative full Bayes approach.

The results for the Chatterjee et al. (2016) method show some similarities to the other methods that use the external information, but also show some differences. The method shows a similar amount of bias or even slightly greater bias than the other methods. The method does result in some gain in efficiency in the point estimates, and also smaller MSE, compared to direct regression, but not as much gain as for the other new methods. For $\hat{\gamma}_3$ there is no gain in efficiency compared to direct regression, and even loss in efficiency compared to the direct regression with Firth correction. The method does tend to give slightly more variability to the predicted probabilities than the other methods. The Chatterjee et al. (2016) method has comparable values of AUC and Brier score as the other methods that use external information.

The simple logistic regression plug-in method is found to have better performance than direct regression, but not as good performance as the more sophisticated approaches for using the external information. It also can result in biased estimates of the γ's.

The other results presented in the Supplementary Materials give similar conclusions.

6. Application to the Prostate Cancer Data

We demonstrate our methodology by enhancing the Prostate Cancer Prevention Trial Risk Calculator for high-grade prostate cancer. Using the data from Tomlins et al. (2015) we will illustrate the methods described in this paper to develop a logistic model that includes all the PCPThg variables and PCA3. We estimate the new model from the training dataset of 679 men, incorporating the known coefficients and their standard errors from the PCPThg calculator. After a transformation ($\log_2(\text{PCA3}+1)$) the distribution of PCA3 is roughly normally distributed in both cohorts and thus the approximate relationship equations (6) are applicable. The distribution of T2:ERG looks like a truncated normal whose value is bounded below at zero, with many observations equal to zero. We dichotomized T2:ERG by splitting at the median and develop a logistic model that includes PCPThg variables and dichotomized T2:ERG. The approximate relationship equations (13) would be appropriate in this case.

These two expanded PCPThg models will be estimated by both the unconstrained methods and the constrained methods described in Section 3 and Section 4. For comparing coefficient estimation across different methods, we report the estimated coefficients and their standard errors calculated from the training dataset. For comparing prediction power, we calculate the Brier Score and the AUC based on the validation dataset. We also present the original PCPThg model and the expanded model developed by Tomlins et al. (2015). We give the calibration plots for the original PCPThg model, the expanded model by Tomlins et al. (2015), the expanded PCPThg model estimated without constraints (direct regression) and the expanded PCPThg model estimated with constraints (transformation approach). The calibration plot contains the predicted and the observed risk of high-grade cancer in 10 groups which are defined by sorting the predicted probabilities from lowest to highest and then separated into 10 groups of approximately equal size. For each group, the expected numbers of events is the sum of the predicted probability in the group. Perfect predictions should be on the 45° line.

Table 2 presents the expanded PCPThg model incorporating these two biomarkers fitted to the training dataset. For the expanded PCPThg model incorporating PCA3 score, if we compare the standard errors across different methods, it is easily seen that the constrained methods can reduce the standard errors of regression coefficients compared to direct regression. For example, the informative full Bayes solution can substantially reduce the standard errors in parameters of variables PSA (0.08 vs 0.19), age (0.008 vs 0.013), DRE findings (0.17 vs 0.27), prior biopsy history (0.16 vs 0.28) and race (0.23 vs 0.31). The constrained ML with Firth penalty can reduce the standard errors of the parameters of variables PSA, age, prior biopsy history and race by at least 50%.

Among the 1218 validation study patients, AUC for PCPThg model and the expanded PCPThg score plus PCA3 model are 0.707 and 0.752. By incorporating PCA3 score in the PCPThg model, the AUC increases to 0.767 in direct regression. However, the constrained methods do not further increase the AUC. All the new methods except for the transformation approach give predicted probabilities that are on average too high, suggesting they are not well calibrated. For calibration, as measured by the Brier score, the original PCPT calculator performs better than direct regression and of the new methods only the transformation approach gives an improved Brier score. In Figure 1 we can see that the expanded PCPThg model incorporating PCA3 tends to overestimate the risk of getting high-grade PCa among those patients with high risk. However, the overall calibration ability of the expanded PCPThg model estimated by the transformation approach still outperforms that of the original PCPThg model, the expanded PCPThg score plus PCA3 model or the expanded PCPThg model estimated by direct regression.

The expanded PCPThg model incorporating binary T2:ERG fitted to the training dataset again shows that the constrained methods can reduce the standard errors of regression coefficients compared to direct regression. In Figure 1 we can see that the expanded PCPThg model incorporating binary T2:ERG tends to overestimate the risk of getting high-grade cancer among those patients with high risk. However, the transformation approach predicts the risk well for the high risk groups.

For the prostate cancer example with either new biomarker, the Chatterjee et al. (2016) method gave parameter estimates and standard errors that are different from both the methods that don't use the external information and from the other methods that do use the external information. In general, they tend to be closer to those of the original PCPT calculator. Also the method gives a much lower predicted population proportion than all the other methods and lower than the observed prevalence of 18.3% in the validation dataset. This is possibly because the Chatterjee et al. (2016) method does not take into account the uncertainty in the estimates coefficients. It is notable, in contrast to what was seen in the simulation studies, that the Chatterjee et al. (2016) method tends to give smaller standard errors than all the other methods for the coefficients of the **X** variables, but larger standard errors for the coefficient of the new biomarker. The method does assume that the **X** distribution is the same in the original dataset and the dataset with the new biomarkers, but in fact there are considerable differences between these **X** distributions, it is unclear how this is affecting the estimates and standard errors. The results (as shown in Appendix F of the Supplementary materials) for the constrained MLE, with a range of values for d demonstrate that the choice of d can be quite impactful, both on the estimates and on their standard errors. As expected the results for $d = 0.1$ are quite close to the Chatterjee et al. (2016) method, because neither method incorporates the uncertainty in the parameter estimates.

7. Discussion

We propose several strategies for translating the external coefficient information obtained from outside the dataset into constraints on regression coefficients in the setting of a logistic regression model describing $\mathrm{Pr(Y = 1|X, B)}$. Simulation studies show that the external coefficient information from the established model can help improve the efficiency of estimation and enhance the predictive power in the expanded model.

In terms of computational efficiency, in simulation studies the transformation approach shows advantage over the informative full Bayes because in the transformation approach the raw draws are first obtained in a fast way and then transformed into draws that obey the constraints based on an efficient optimization algorithm, while the informative full Bayes solution produces constrained draws inefficiently. When the dimensionality of the predictors $\mathbf{X}$ increases, the computational cost of the transformation approach solution will not increase much because the high-dimensional optimization problem will always reduce to a one-dimensional optimization problem based on our algorithm regardless of the dimensionality of the predictor space. Furthermore, the correlation among the samples in the Markov chain for the informative full Bayes approach is very high and effective samples are harder to obtain when the dimensionality increases (additional simulation results that validate this finding are not shown). As a consequence, the discrepancy of these two constrained solutions in computational cost will be more apparent in higher dimensions. In general the Bayesian approaches are much more computationally time-consuming than the other approaches. While it is conceivable that better algorithms and improved programming could speed these up considerably, it is very unlikely that they will ever have comparable speed to the CML methods or the Chatterjee et al. (2016) approach. A general overview of the computational and implementation details for all the methods described in this paper are given in Supplementary Materials Appendix G.

The efficiency gain in the expanded model of interest depends on the sample size used to construct the established model and the sample size used to estimate the expanded model of interest. In our simulation studies the established models are based on large datasets with 10000 observations while the current datasets are very small. The relative efficiency gain in the regression coefficient of variables $\mathbf{X}$ by incorporating the external coefficient information is significant and the prediction power in the validation dataset is enhanced. However, when the sample size in the current dataset is large enough to estimate the expanded model, the constrained methods do not lead to much improvement in the predictive ability compared to direct regression, as was the case in the prostate cancer example. However, our numerical results suggest that improved precision of the coefficient estimates, as measured by standard errors, can be achieved even if the current dataset is not small.

A situation we did not consider in the simulation studies is when the event of interest is rare and the predicted probabilities are very low. While there may be other methods that exploit this assumption, we hypothesize that the relative performance of the methods we considered may be similar to those for the small samples sizes we did evaluate. This is because both situations have limited information in the data from which to estimate regression coefficients. However, this hypothesis would need to be investigated.

The approaches proposed in this paper are based on establishing a relationship between the parameters in the $\mathrm{Y}|\mathbf{X}$ model and the parameters in the $\mathrm{Y}|\mathbf{X},\mathrm{B}$ model. Depending on the form of B and the structure of the models these relationships need to be analytically derived and are approximations. The methods also require an explicit model for $\mathrm{B}|\mathbf{X}$. While it would be desirable to avoid having to specify a parametric model for $\mathrm{B}|\mathbf{X}$, we also note that the appropriateness of the model for $\mathrm{B}|\mathbf{X}$ can be checked to some degree from the small dataset. The differences in the distributions of $\mathbf{X}$ in the external and internal studies will not have much effect on the performance of our proposed constrained methods (simulation results not shown). This is because the coefficients' approximate relationship equations are constructed based on the conditional distributions $\mathrm{Y}|\mathbf{X},\mathrm{B}$ and $\mathrm{B}|\mathbf{X}$. As long as these two conditional distributions are correctly specified in the internal study, the approximate relationship equations will hold regardless of the differences in the distributions of $\mathbf{X}$ in these two studies. The approaches also do have the feature that they can directly incorporate the uncertainty in the parameters of the $\mathrm{Y}|\mathbf{X}$ model. An alternative method of using $\bar{p}$ as a covariate is appealing because of its simplicity and broad applicability, although it does appear to have slightly worse properties than the more sophisticated approaches. The approach of Chatterjee et al. (2016) is appealing because it is applicable for any form of B and it does not require an explicit model for $\mathrm{B}|\mathbf{X}$, however it does require the same distribution of $\mathbf{X}$ in the two populations and it does not incorporate the uncertainty in the parameters of the $\mathrm{Y}|\mathbf{X},\mathrm{B}$ model. We also found that it was sometimes numerically unstable for small sample sizes.

One point of future consideration is the distribution of the new biomarker B. We develop the approximate relationship equation for the scenarios that B is Gaussian and binary. When B is multivariate Gaussian, based on the generalization of equation (5), assuming $\mathbf{B}|\mathbf{X}$ is multivariate normal with L dimensions, mean $\mathbf{X}\boldsymbol{\theta}$ and covariance matrix $\mathbf{V}_{L\times L}$, the approximate relationship between $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ is:

$$\beta_j \approx (\gamma_j + \sum_{l=1}^{L} \gamma_{p+l}\theta_{lj})/(1+\boldsymbol{\gamma}_B^T\mathbf{V}\boldsymbol{\gamma}_B/1.7^2)^{\frac{1}{2}}, j=0,\ldots,p \tag{17}$$

Then the strategies to incorporate the external coefficient information described in Section 3 can be easily extended in this case. However, if additional biomarkers follow other distributions, these approximate relationship equations will fail. Therefore, further investigations are needed for the generalization of our proposed constrained solutions to flexibly adapt to other possible distributions of the new biomarker.

Acknowledgments

This research was supported by the NSF grant DMS 1406712 and NIH grant CA 129102.

References

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.

Cheng, W., Taylor, J. M. G., Vokonas, P. S., Park, S. K. and Mukherjee", B. (2018) Improving estimation and prediction in linear regression incorporating external information from an established reduced model. *Statistics in Medicine*, **37**, 1515–1530.

D'Agostino, R. B., Grundy, S., Sullivan, L. M., Wilson, P. and for the CHD Risk Prediction Group (2001) Validation of the framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *The Journal of the American Medical Association*, **286**, 180–187.

Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C. and Mulvihill, J. J. (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, **81**, 1879–1886.

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., Ankerst, D. P., Gail, M. H., Chatterjee, N. and Pfeiffer, R. M. (2017) Comparison of approaches for incorporating new information into existing risk prediction models. *Statistics in Medicine*, **36**, 1134–1156.

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.

Gunn, L. H. and Dunson, D. B. (2005) A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics*, **6**, 434–449.

Heinze, G. and Schemper, M. (2002) A solution to the problem of separation in logistic regression. *Statistics in Medicine*, **21**, 2409–2419.

Imbens, G. W. and Lancaster, T. (1994) Combining micro and macro data in microeconometric models. *The Review of Economic Studies*, **61**, 655–680.

Mealiffe, M. E., Stokowski, R. P., Rhees, B. K., Prentice, R. L., Pettinger, M. and Hinds, D. A. (2010) Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *Journal of the National Cancer Institute*, **102**, 1618–1627.

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.

Newcombe, P. J., Reck, B. H., Sun, J., Platek, G. T., Verzilli, C., Kader, A. K., Kim, S.-T., Hsu, F.-C., Zhang, Z., Zheng, S. L., Mooser, V. E., Condreay, L. D., Spraggs, C. F., Whittaker, J. C., Rittmaster, R. S. and Xu, J. (2012) A comparison of Bayesian and frequentist approaches to incorporating external information for the prediction of prostate cancer risk. *Genetic Epidemiology*, **36**, 71–83.

Qin, J. (2000) Combining parametric and empirical likelihoods. *Biometrika*, **87**, 484–490.

Qin, J., Zhang, H., Li, P., Albanes, D. and Yu, K. (2015) Using covariate-specific disease prevalence information to increase the power of case-control studies. *Biometrika*, **102**, 169–180.

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.

Steyerberg, E. W., Eijkemans, M. J. C., Van Houwelingen, J. C., Lee, K. L. and Habbema, J. D. F. (2000) Prognostic models based on literature and individual patient data in logistic regression analysis. *Statistics in Medicine*, **19**, 141–160.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J. and Kattan, M. W. (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, **21**, 128–138.

Thompson, I. M., Ankerst, D. P., Chi, C., Goodman, P. J., Tangen, C. M., Lucia, M. S., Feng, Z., Parnes, H. L. and Coltman, C. A. (2006) Assessing prostate cancer risk: results from the prostate cancer prevention trial. *Journal of the National Cancer Institute*, **98**, 529–534.

Tomlins, S. A., Day, J. R., Lonigro, R. J., Hovelson, D. H., Siddiqui, J., Kunju, L. P., Dunn, R. L., Meyer, S., Hodge, P., Groskopf, J., Wei, J. T. and Chinnaiyan, A. M. (2015) Urine TMPRSS2:ERG plus PCA3 for individualized prostate cancer risk assessment. *European Urology*, **70**, 45–53.

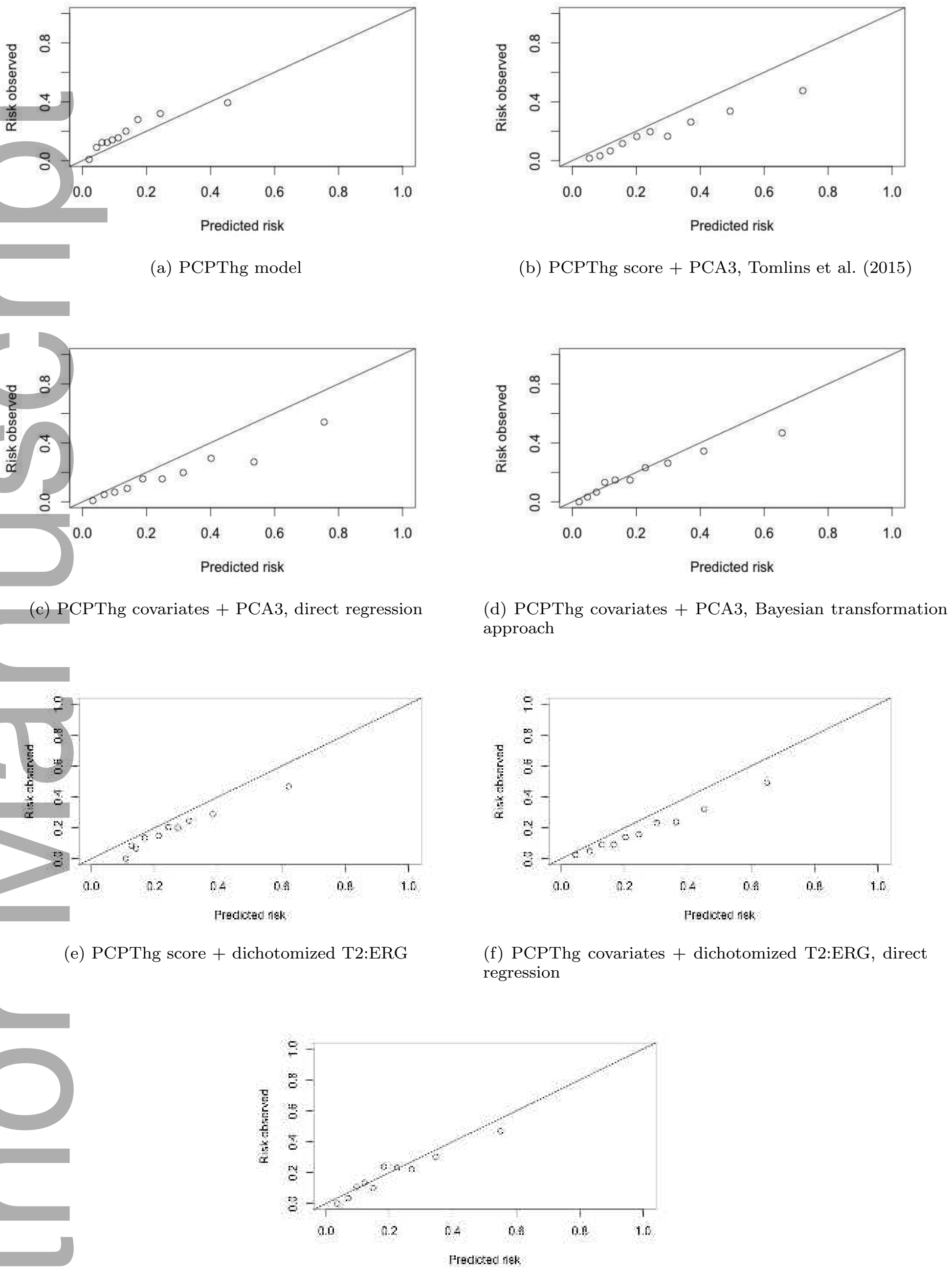
Truong, M., Yang, B. and Jarrard, D. F. (2013) Toward the detection of prostate cancer in urine: a critical analysis. *The Journal of Urology*, **189**, 422 – 429.

Table 1. Simulation results for Gaussian 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	0.605	0.864	0.49 (0.333)	-
Established model using known $\bar{\beta}$	-	-	-	0.796	0.761	0.51 (0.239)	-
Direct regression	3.37 (1)	3.40 (1)	2.35 (1)	0.661	0.852	0.49 (0.344)	1.3
MSE	3.36	3.48	0.96				
Direct regression + Firth	2.89 (1.55)	2.92 (1.52)	1.99(1.69)	0.651	0.852	0.49 (0.324)	2.4
MSE	2.10	2.18	0.49				
Non-informative Bayes	2.72 (1.79)	2.75 (1.78)	2.04 (1.77)	0.647	0.852	0.49 (0.307)	3.6
MSE	1.88	1.93	0.47				
Constrained ML	3.08 (3.60)	3.17 (3.91)	2.30 (1.11)	0.628	0.857	0.49 (0.313)	44.9
MSE	0.90	0.88	0.84				
Constrained ML + Firth	2.88 (6.01)	2.97 (6.32)	1.96 (1.85)	0.622	0.857	0.49 (0.303)	78.2
MSE	0.55	0.53	0.45				
Informative full Bayes	2.87 (4.93)	2.98 (5.15)	2.30 (1.33)	0.624	0.857	0.49 (0.301)	9097.6
MSE	0.64	0.67	0.72				
Transformation	2.90 (6.80)	3.00 (6.94)	1.93 (1.75)	0.622	0.857	0.49 (0.298)	888.2
MSE	0.48	0.48	0.48				
Chatterjee et al	3.17 (2.91)	3.29 (3.06)	2.34 (1.01)	0.631	0.859	0.49 (0.342)	43.2
MSE	1.13	1.17	0.94				
Simple logistic $(\bar{p}, B)$	3.27 (1.64)	3.34 (1.62)	2.28 (1.11)	0.644	0.858	0.49 (0.339)	0.9
MSE	2.04	2.16	0.84				

Table 2. Expanded PCPThg model: for each method, point estimate (standard error) from the training dataset, and the Brier score, the AUC and the mean and SD of predicted probabilities from the validation dataset. The sample size of the training dataset is 679. The sample size of the validation dataset is 1218.

Model	PSA	Age	DRE	Prior biopsy	Race		Scaled Brier Score	AUC	$\hat{p}$ mean(SD)
			findings	history					
Original PCPThg	1.29 (0.09)	0.031 (0.012)	1.00 (0.17)	-0.36 (0.18)	0.96 (0.27)	–	0.933	0.707	0.14 (0.132)
Estimated PCPThg	1.06 (0.18)	0.033 (0.012)	1.15 (0.26)	-1.44 (0.27)	0.44 (0.29)	–	0.975	0.716	0.27 (0.174)
Expanded model						PCA3			
with PCA3 score									
PCPThg score+PCA3	–	–	–	–	–	–	0.950	0.752	0.27 (0.201)
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)	0.950	0.767	0.28 (0.221)
Direct regression + Firth	0.97 (0.19)	0.009 (0.013)	1.06 (0.27)	-1.27 (0.27)	0.05 (0.31)	0.56 (0.08)	0.953	0.767	0.28 (0.219)
Non-informative Bayes	0.98 (0.18)	0.009 (0.013)	1.05 (0.27)	-1.27 (0.27)	0.04 (0.30)	0.56 (0.08)	0.950	0.767	0.28 (0.218)
Constrained ML	1.20 (0.09)	0.010 (0.007)	1.08 (0.14)	-0.55 (0.13)	0.30 (0.19)	0.59 (0.08)	0.948	0.766	0.27 (0.225)
Constrained ML + Firth	1.19 (0.09)	0.012 (0.006)	1.08 (0.14)	-0.54 (0.13)	0.47 (0.11)	0.53 (0.07)	0.947	0.764	0.27 (0.218)
Informative full Bayes	1.23 (0.10)	0.009 (0.008)	0.99 (0.17)	-0.73 (0.17)	0.26 (0.22)	0.60 (0.08)	0.946	0.767	0.27 (0.222)
Transformation	1.23 (0.07)	0.008 (0.009)	0.96 (0.14)	-0.50 (0.13)	0.41 (0.19)	0.55 (0.08)	0.883	0.765	0.22 (0.191)
Chatterjee et al	1.22 (0.08)	0.007 (0.005)	0.86 (0.10)	-0.20 (0.08)	0.58 (0.11)	0.56 (0.10)	0.888	0.759	0.15 (0.168)
Simple logistic ($\bar{p}, B$)	0.82 (0.18)	0.023 (0.000)	0.64 (0.11)	-0.23 (0.01)	0.61 (0.10)	0.55 (0.08)	0.940	0.759	0.27 (0.204)
Expanded model						T2:ERG			
with binary T2:ERG									
PCPThg score + T2:ERG	–	–	–	–	–	–	0.932	0.732	0.26 (0.153)
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)	0.929	0.745	0.26 (0.179)
Direct regression + Firth	0.98 (0.18)	0.032 (0.012)	1.02 (0.26)	-1.41 (0.27)	0.57 (0.29)	0.76 (0.20)	0.930	0.744	0.27 (0.177)
Non-informative Bayes	0.99 (0.18)	0.032 (0.012)	1.01 (0.26)	-1.40 (0.27)	0.55 (0.29)	0.76 (0.20)	0.926	0.745	0.27 (0.175)
Constrained ML	1.14 (0.07)	0.032 (0.004)	1.06(0.14)	-0.52 (0.11)	0.81 (0.18)	0.74 (0.21)	0.928	0.742	0.25 (0.176)
Constrained ML + Firth	1.14 (0.07)	0.032 (0.004)	1.06 (0.14)	-0.52 (0.11)	0.80 (0.17)	0.72 (0.20)	0.931	0.742	0.26 (0.176)
Informative full Bayes	1.14 (0.09)	0.033 (0.007)	0.95 (0.14)	-0.76 (0.16)	0.77 (0.21)	0.73 (0.19)	0.922	0.744	0.25 (0.175)
Transformation	1.17 (0.07)	0.030 (0.007)	0.94 (0.12)	-0.50 (0.11)	0.89 (0.16)	0.74 (0.14)	0.889	0.742	0.21 (0.152)
Chatterjee et al	1.25 (0.03)	0.029 (0.002)	0.85 (0.05)	-0.37 (0.04)	1.06 (0.05)	0.77 (0.27)	0.911	0.736	0.14 (0.129)
Simple logistic ($\bar{p}, B$)	0.98 (0.18)	0.023 (0.000)	0.76 (0.11)	-0.28 (0.01)	0.73 (0.10)	0.80 (0.19)	0.918	0.739	0.25 (0.155)



(a) PCPThg model

(b) PCPThg score + PCA3, Tomlins et al. (2015)

(c) PCPThg covariates + PCA3, direct regression

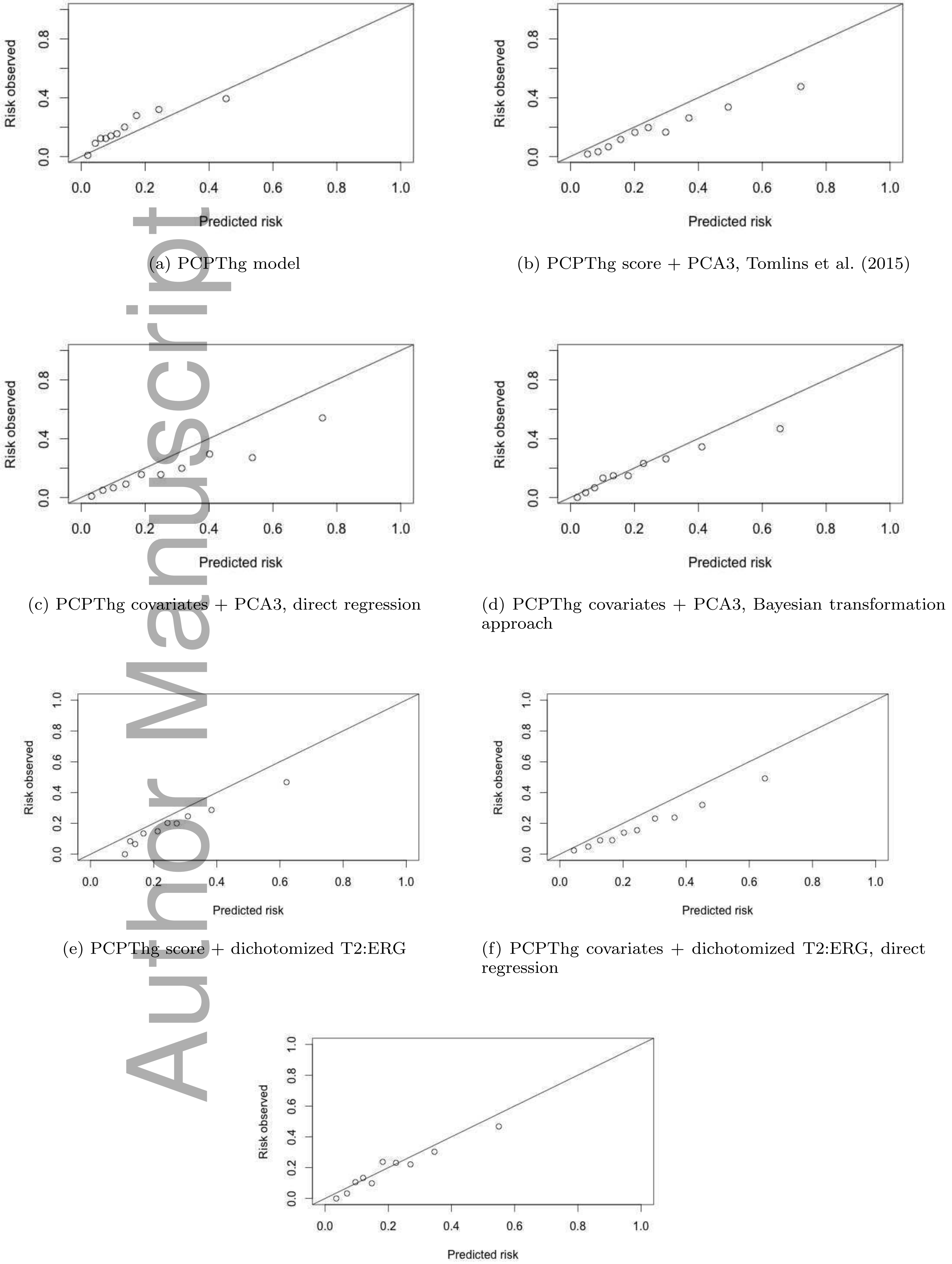
(d) PCPThg covariates + PCA3, Bayesian transformation approach

(e) PCPThg score + dichotomized T2:ERG

(f) PCPThg covariates + dichotomized T2:ERG, direct regression

(g) PCPThg covariates + dichotomized T2:ERG, Bayesian transformation approach

Fig. 1. Calibration plot of the original high-grade Prostate Cancer Prevention Trial risk calculator (PCPThg) and calibration plots of the expanded PCPThg model by incorporating PCA3 score and dichotomized T2:ERG



(a) PCPThg model

(b) PCPThg score + PCA3, Tomlins et al. (2015)

(c) PCPThg covariates + PCA3, direct regression

(d) PCPThg covariates + PCA3, Bayesian transformation approach

(e) PCPThg score + dichotomized T2:ERG

(f) PCPThg covariates + dichotomized T2:ERG, direct regression

(g) PCPThg covariates + dichotomized T2:ERG, Bayesian transformation approach