The Annals of Applied Statistics
2024, Vol. 18, No. 2, 1051–1063
https://doi.org/10.1214/23-AOAS1822

FUNCTIONAL PARTIAL LEAST SQUARES WITH CENSORED OUTCOMES: PREDICTION OF BREAST CANCER RISK WITH MAMMOGRAM IMAGES

BY SHU JIANG[1,a], JIGUO CAO[b] AND GRAHAM A. COLDITZ[c]

[1]*Division of Public Health Sciences, Washington University School of Medicine in St. Louis,* [a]*jiang.shu@wustl.edu,* [b]*jiguo_cao@sfu.ca,* [c]*colditzg@wustl.edu*

We consider the problem of predicting breast cancer risk using mammogram imaging data where the dimension of pixels greatly exceed the number of individuals in the cohort. The functional partial least squares (FPLS) is a popular dimensional reduction method in constructing latent explanatory components using linear combinations of the original predictor variables. While FPLS with scalar responses has been studied in the literature, the presence of right censoring under the survival framework poses challenges in modeling and estimation. Given several different representations for PLS with Cox regression in the literature, we unify and extend three formulations to deal with right censoring, that is, reweighing, mean imputation, and deviance residuals to the functional setting in this paper. We empirically investigate and compare the performance of the three proposed FPLS frameworks in the context of imaging predictor via intensive simulation studies. The proposed methods are applied to the Joanne Knight Breast Health Cohort where we show increased model discriminatory performance under the FPLS framework compared to competing models.

1. Introduction. Breast cancer prevention has evaded being a central focus in cancer control, in large part because it has many risk factors that reflect the normal biology and physiology for women (menarche, childbirth, menopause) that are regarded as nonmodifiable. However, the global burden continues to rise with economic development, improved childhood nutrition, and smaller family sizes as well as later age at first birth that accompanies higher education attainment for women. Now, one in four cancers diagnosed each year among women worldwide is breast cancer. Recent meta-analysis shows that adding mammographic breast density to models improves their performance (Vilmun et al. (2020)). While accepted as an intermediate marker for breast cancer risk (Boyd et al. (2009)), the mammogram image includes many additional texture features that add to prediction beyond density alone (Anandarajah et al. (2022), Jiang and Colditz (2023)). Shifting the framework from summary features to the digital image is thus required.

While mammography was developed for diagnosis and treatment of early stage breast cancer to improve outcomes and reduce mortality (Tabar et al. (1985)), the breast image can also contribute to long-term risk prediction identifying higher risk women who might benefit from lifestyle interventions or chemoprevention as reflected in current guidelines (Visvanathan et al. (2019)). It is also possible that low-risk women may be identified for whom a less intensive screening schedule may better balance the risks and benefits. The literature includes numerous studies evaluating approaches to such a more precision based approach (Pashayan et al. (2018)). To achieve this, one can envision making more use of the 13 million pixels in each image instead of using a summary measure such as breast density. (Anandarajah et al. (2023)) The richness of data in the full field digital breast image thus requires efficient dimension reduction to facilitate analysis.

Received February 2023; revised August 2023.
Key words and phrases. Functional partial least squares, survival analysis, image analysis, risk prediction.

Principal component analysis (PCA) is a popular method to reduce the number of predictors by extracting a limited number of principal components. The principal components are extracted such that the they can explain the maximum amount of variance among the predictors, without reference to the outcome variable. The partial least squares (PLS), on the other hand, is a widely adopted alternative in constructing new explanatory components using linear combinations of the original predictors. This iterative procedure maximizes the covariance between the outcome and the newly constructed components (Wold (1966), Wold (1975a)). PLS has been successfully employed in regression modeling with high-dimensional predictors, with majority of publications dealing with response variables that belong to the exponential family.

Due to the presence of right-censoring in the survival outcome, there have been several different formulations of the Cox regression model to enable the PLS estimation. For instance, Park, Tian and Kohane (2002) reformulated the survival data into a Poisson regression with appropriate scaling. Li and Gui (2004) proposed a partial Cox regression method. Datta, Le-Rademacher and Datta (2007) considered modeling the log-transformed failure times to enable PLS estimation. Nygård et al. (2008) modified the Poisson regression representation such that the baseline hazards and the predictors are obtained in separate steps. Bastien et al. (2015) considered using the normalized martingale residuals, a proxy of the excess of death, to overcome the right-censoring issue.

While PLS has been successfully employed in regression modeling with high-dimensional predictors, most of its application is in the microarray, genomic, and proteomics context. Our goal in this paper is to develop the functional partial least squares (FPLS) to accommodate image predictors and, at the same time, to select components highly associated with the survival outcome while controlling for the smoothness of these components. FPLS for response variables that belong to an exponential family have been studied previously in the literature; see Reiss and Ogden (2007), for example. Because there are different PLS formulations under the survival framework, we unify several widely adopted formulations here in this paper, extend to the functional framework with image predictors, and empirically investigate their prediction performance.

We summarize three main contributions of this article as follows. First, we propose a novel FPLS method for survival outcomes that: (a) may be right-censored and (b) involves predictors that are in the form of high-dimensional images. Second, due to various formulations in Cox regression with PLS, we study and unify three types of extensions, including reweighing, mean imputation, and the deviance residual based method, to the functional version, and empirically investigate and compare their prediction performance. Finally, we apply the proposed methods to our motivating dataset from the Joanne Knight Breast Health Cohort and leverage new insights relating features extracted from mammogram imaging data to breast cancer risk.

In Section 2 we introduce the notation used in this article and the underlying model setup. We first discuss in Section 2.1 the three modeling schemes, reweighing, mean imputation, and deviance residual based method to deal with censored observations. The roughness penalty as well as the computation algorithm are discussed in detail in Section 2.2. In Section 3 we illustrate results under a number of simulation studies to investigate and compare the empirical performance of the three proposals. The proposed methods are then applied to the motivating dataset from the Joanne Knight Breast Health cohort at Siteman Cancer Center in Section 4. We end this paper with a discussion in Section 5.

2. Model and estimation method.

Suppose there are n independent individuals in the cohort. For an individual i, we use the pair (T_i, δ_i) to represent the observed survival outcome. Here T_i stands for the minimum of failure time and censoring time, while δ_i serves as the censoring indicator. Specifically, $\delta_i = 1$ indicates that the observed T_i corresponds to the failure time. We let $\mathbb{S}$ be a two-dimensional bounded domain and $\boldsymbol{s} = (s_1, s_2)$ be a point in $\mathbb{S}$. We then define $Z_i = \{Z_i(\boldsymbol{s}), \forall \boldsymbol{s} \in \mathbb{S}\}$ to be the image data for individual i, $i = 1, \dots, n$.

The most widely used model for right-censored survival data is the Cox proportional hazards model. Our aim is to build the following hazard function:

$$h_i(t) = h_0(t) \exp(\beta_1 \xi_{i1} + \beta_2 \xi_{i2} + \cdots + \beta_K \xi_{iK}), \tag{1}$$

where $h_0(t)$ is the nonparametric baseline hazard function, and the kth latent component ξ_{ik} is the projection of the ith image $Z_i(\mathbf{s})$ onto a latent space, defined by the FPLS weight functions $\phi_k(\mathbf{s})$,

$$\xi_{ik} = \int_{\mathbb{S}} Z_i(\mathbf{s}) \phi_k(\mathbf{s}) \, d\mathbf{s}, \tag{2}$$

$k = 1, \dots, K$. The kth FPLS weight function $\phi_k(\mathbf{s})$ can be estimated as a linear combination of two-dimensional basis functions,

$$\phi_k(\mathbf{s}) = \sum_{m=1}^{M} w_{km} B_m(\mathbf{s}), \tag{3}$$

where $B_m(\mathbf{s})$ denotes the mth two-dimensional basis function, and w_{km} is the basis coefficient, $k = 1, \dots, K; m = 1, \dots, M$. The form of the basis function $B_m(\mathbf{s})$ is chosen to be tensor product of one-dimensional basis functions in this article, but this can be flexibly extended to other forms of two-dimensional basis functions, such as Bernstein polynomials defined over triangulations (Jiang et al. (2023a)). By substituing (3) into (2), the kth latent component can be written as

$$\xi_{ik} = \sum_{m=1}^{M} w_{km} \int_{\mathbb{S}} Z_i(\mathbf{s}) B_m(\mathbf{s}) \, d\mathbf{s},$$

where it is clear that our goal is to estimate the basis coefficients in order to determine the latent components. After the components $\xi_{i1}, \dots, \xi_{iK}$ are determined, model (1) can be used to estimate the hazards function by the standard partial likelihood approach under the Cox proportional hazards model.

As discussed in Section 1, the literature offers different variants of Cox-PLS as well as estimation techniques to extract the latent components In what follows, we focus on the reweighing, mean imputation, and residual deviance approach and extend these concepts to the functional partial least squares (FPLS) framework with the presence of right-censoring.

2.1. *Formulations under right-censoring.* REWEIGHING. Under the reweighing method, we consider estimating the latent components using some function of the failure time $f(T)$ as the outcome variable such that $f : [0, \infty) \to \mathrm{IR}$. An example of such function is the log transformation. Under the survival framework, however, T is not observed for all individuals due to the presence of right-censoring. Ignoring the censored observations would give rise to biased estimates. A potential solution is to reweigh the observed T_i by

$$\tilde{Y}_i = \frac{\delta_i f(T_i)}{\hat{S}^c(T_i^-)}, \tag{4}$$

where $\hat{S}^c(\cdot)$ denotes the Kaplan–Meier survival function of the censoring random variable C, and T^- denotes the left limit of T.

MEAN IMPUTATION. Another way to accommodate right-censoring is to use the mean imputation method. Specifically, $\tilde{Y}_i$ is set to $f(T_i)$ if failure is observed for individual i ($\delta_i =$

1). The unobserved failure times ($\delta_i = 0$), on the other hand, are replaced by their expected values, given that the failure time is larger than the censored time C_i,

$$\tilde{Y}_i = \frac{\sum_{\tau(j)>C_i} f(\tau_j)\triangle\hat{S}(\tau_{(j)})}{\hat{S}(C_i)}, \tag{5}$$

where $\tau_{(1)} < \tau_{(2)} < \cdots < \tau_{(J)}$ denote the J ordered distinct failure times, $\hat{S}(\cdot)$ is the Kaplan–Meier survival function of T, and $\triangle\hat{S}(\tau_{(j)})$ denotes the jump size of $\hat{S}(\cdot)$ at time $\tau_{(j)}$. Under this setup the largest observation will be treated as the true failure, amounting to making $\tau_{(J)}$ the largest mass point of the estimated survival function of T (Datta (2005)).

Deviance Residual. This last strategy involves substituting suitably chosen residuals for the survival endpoint, enabling inheritance of simple algorithms applicable to continuous outcomes, and bypassing difficulties deriving from censoring (Segal (2006)). Under the setting with no time-varying covariates, at most one event is observed per patient, and each patient is under observation from time 0; the martingale residual for the ith patient can be defined as

$$G_i = \delta_i - H_0(t)\exp(\beta_1\xi_{i1} + \beta_2\xi_{i2} + \cdots + \beta_K\xi_{iK}),$$

where $H_0(t) = \int_0^t h_0(t)\,dt$ is the cumulative hazard function at time t. The martingale residual can be interpreted as the difference between observed and expected number of events. Because the martingale residuals are highly skewed (i.e., takes value between $-\infty$ and 1), the deviance residual is usually adopted. Consistent with previous notation, we use $\tilde{Y}_i$ to denote the deviance residuals that serve as the outcome for the FPLS method,

$$\tilde{Y}_i = \operatorname{sign}(\hat{G}_i)\bigl[2\{-\hat{M}_i - \delta_i \log(\delta_i - \hat{G}_i)\}\bigr]^{1/2}. \tag{6}$$

The deviance residuals $\tilde{Y}_i$ can be viewed as a normalized version of the martingale residuals such that the log transformation expands G_i on a real line taking values from $-\infty$ to ∞. As FPLS is an iteratively updated algorithm, we will not have the estimates β on the first iteration, as shown in the subsequent section. The initial input value for the FPLS algorithm under the deviance residual method thus represents the null deviance residual without the inclusion of covariates, that is, regressing outcome against 1.

2.2. *Penalized Cox-FPLS.*

Here we provide some intuition on the estimation for the vector of basis coefficients, $\mathbf{w}_k = (w_{k1}, \ldots, w_{kM})^T$, to the kth FPLS weight function $\phi_k(\mathbf{s})$. Specifically, we aim to find the vector of basis coefficients $\mathbf{w}_1$ for the first FPLS weight function $\phi_1(\mathbf{s})$ by maximizing

$$\operatorname{cov}^2(\boldsymbol{\xi}_1, \tilde{\mathbf{Y}}) = \boldsymbol{\xi}_1^T\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T\boldsymbol{\xi}_1, \tag{7}$$

with the constraint that $\mathbf{w}_1^T\mathbf{w}_1 = 1$, where $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{n1})^T$, and $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)^T$. The vector of basis coefficients $\mathbf{w}_k, k = 2, \ldots, K; K \le M$, to the subsequent FPLS weight functions are also chosen to maximize the covariance function subject to the constraint that $\mathbf{w}_k^T\mathbf{w}_k = 1$ and $\mathbf{w}_k^T\mathbf{w}_j = 0$ if $j \neq k$.

Under the functional framework, we also add a roughness penalty function to control the roughness of the FPLS weight functions. Using the first FPLS weight function as an example, we estimate the vector of the basis coefficients $\mathbf{w}_1$ by maximizing $\operatorname{cov}^2(\boldsymbol{\xi}_1, \tilde{\mathbf{Y}})$ with the constraint that

$$\mathbf{w}_1^T\mathbf{w}_1 + \lambda\int_{\mathbb{S}}\biggl[\frac{\partial^2\phi_k(\mathbf{s})}{\partial s_1^2}\biggr]^2 + \biggl[\frac{\partial^2\phi_k(\mathbf{s})}{\partial s_2^2}\biggr]^2 ds_1\,ds_2 = \mathbf{w}_1^T(\mathbf{I} + \lambda\mathbf{P})\mathbf{w}_1 = 1,$$

Algorithm 2.1: Functional Partial Least Square Algorithm

```
Input: $\mathbf{ZB}, \tilde{\mathbf{Y}}^{(1)}, K, \lambda, \mathbf{P}$
Initialize: $\mathbf{z}_1 = \mathbf{ZB}$
for $k = 1$ to $K$ do
    Set $\boldsymbol{\xi}_k = 0, \mathbf{u} = \tilde{\mathbf{Y}}^{(k)}, tol = 1e^{-5}, true$
    while $true$ do
        $\mathbf{M} = (\mathbf{I} + \lambda \mathbf{P})^{-1}$
        $\mathbf{w}_k = \mathbf{M}\mathbf{z}_k^T \mathbf{u}$
        $\mathbf{w}_k = \mathbf{w}_k / \|\mathbf{w}_k\|$
        $\boldsymbol{\xi}_0 = \mathbf{z}_k \mathbf{w}_k$
        $rss = \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_0\|$
        $\boldsymbol{\xi}_k = \boldsymbol{\xi}_0$
        $\mathbf{p} = \mathbf{z}_k^T \boldsymbol{\xi}_k / \boldsymbol{\xi}_k^T \boldsymbol{\xi}_k$
        $q = \mathbf{u}^T \boldsymbol{\xi}_k / \boldsymbol{\xi}_k^T \boldsymbol{\xi}_k$
        $\mathbf{u} = \mathbf{u} / \|\mathbf{u}\|$
        if $rss \le tol$ then
            $false$
        end
    end
    $\mathbf{z}_k = \mathbf{z}_k - \boldsymbol{\xi}_k \mathbf{p}^T$
    $\tilde{\mathbf{Y}}^{(k)} = \tilde{\mathbf{Y}}^{(k)} - q\boldsymbol{\xi}_k$
end
return: $\mathbf{w}_1, \ldots, \mathbf{w}_K; \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$
```

where λ is the smoothing parameter, $\mathbf{P} = \int_{\mathbb{S}} [\frac{\partial^2 \mathbf{B}(\mathbf{s})}{\partial s_1^2}][\frac{\partial^2 \mathbf{B}^T(\mathbf{s})}{\partial s_1^2}] + [\frac{\partial^2 \mathbf{B}(\mathbf{s})}{\partial s_2^2}][\frac{\partial^2 \mathbf{B}^T(\mathbf{s})}{\partial s_2^2}]\, ds_1\, ds_2$ is the $M \times M$ matrix, and $\mathbf{B}(\mathbf{s}) = (B_1(\mathbf{s}, \ldots, B_M(\mathbf{s}))^T$. The other $K-1$ components, $\mathbf{w}_2, \ldots, \mathbf{w}_K$, follow where maximization is subject to the orthogonality constraint. The smoothing parameter λ can be chosen by cross-validation (Ramsay and Silverman (2005), Reiss and Ogden (2007)). For instance, we can choose λ for a given K such that $\mathbf{w}(\lambda(K))$ maximizes the prespecified prediction accuracy measure within the cross-validation.

We provide a detailed Algorithm 2.1 in carrying out the estimation of the latent components as a function of the weight vectors. The estimation procedure is in the same spirit as the well-developed NIPALS and SIMPLS algorithm (Wold (1975b), Martens and Næs (1989)). We note here that for a univariate response, both the NIPALS and SIMPLS estimates are equivalent (Martens and Næs (1989)). The term, $\mathbf{ZB}$, is defined as the $n \times M$ matrix with the imth element $\int_{\mathbb{S}} Z_i(\mathbf{s}) B_m(\mathbf{s})\, d\mathbf{s}$.

3. Simulation study. We aim to investigate and compare the finite sample performances of the three proposed Cox-FPLS methods. Additional comparison with the unsupervised functional principal component regression (FPCR) approach with survival outcome (Kong et al. (2018)) is also presented in this simulation study. Specifically, we simulate the imaging data from $Z_i(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{m=1}^{3} a_{im} B_m(\mathbf{s})$, where the basis functions $B_m, m = 1, 2, 3$, are formed by tensor products of Fourier basis functions on $[-\pi, \pi] \times [-\pi, \pi]$. Without loss of generality, we assume that the image is de-meaned, that is, $\mu(\mathbf{s}) = 0$. We consider the resolution of 32×32 which leads to 1024 equidistance pixels within each image. The coefficients, a_{i1}, a_{i2}, a_{i3}, are uncorrelated random variables simulated from a normal distribution with the mean $\mathbf{0}$ and decreasing variance $10, 8, 4$, respectively.

We consider the following hazard function $h_i(t) = h_0(t)\exp\{\int_{\mathbf{s}\in\mathbb{S}} c(\mathbf{s})Z_i(\mathbf{s})\,d\mathbf{s}\}$, where $c(\mathbf{s})$ represent the coefficient function. We aim to investigate two scenarios for $c(\mathbf{s})$: (1) $c(\mathbf{s}) = B_3(\mathbf{s})$, and (2) $c(\mathbf{s}) = 0.25B_1(\mathbf{s}) + 0.5B_2(\mathbf{s}) + B_3(\mathbf{s})$. Note that under scenario 1, only the third basis function with the least variability is associated with the hazard function. Under scenario 2 a linear combination of all three basis functions are associated with the hazard function but with decreasing magnitude. To investigate whether the proposed algorithm can efficiently identify the latent components that are associated with risk, we constrain the algorithm to choose $K = 1$ and $K = 2$ components only. Here we set the maximum time of the study to be 15 years. We simulate the baseline hazard function from a Weibull distribution $h_0(t) = \kappa\rho(\rho t)^{\kappa-1}$ with increasing risk over time, where $\kappa = 2$ and $\rho = 0.158$. The failure time for individual i is generated from $H_i^{-1}(u)$, $u \sim \text{unif}(0, 1)$, where $H(t)$ is the cumulative hazard function. We assume independent censoring where the censoring time $C_i \sim (0, C_{\max})$, with $C_{\max}$ set at a value such that the percentage of being censored by the end of the study is approximately 30%.

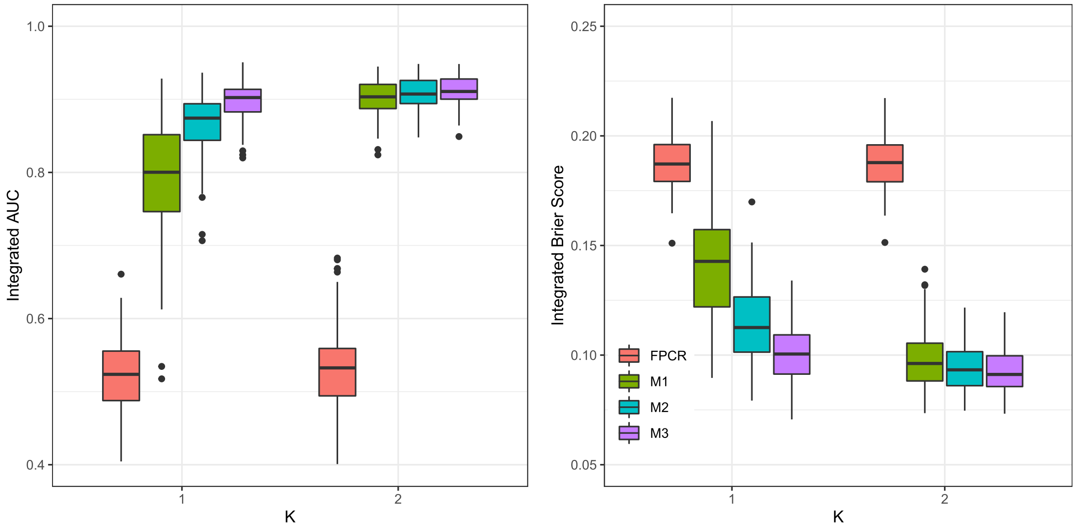
We simulated 400 individuals per dataset and randomly chose 300 to be in the training set and the remaining 100 in the validation set to avoid over-optimism on the model prediction performance. Within the training set, we conducted a five-fold cross-validation to select the smoothing parameter λ from the grid of 0, 1, 10, 100, and 1000. A log transformation is used as the function f to convert the observed times on only the positive half of the real line to values spanning the entire real line. The above simulation process is replicated 100 times in this simulation study.

Figure 1 illustrates the model performance under Scenario 1 where the coefficient function is $c(\mathbf{s}) = B_3(\mathbf{s})$. We measure the prediction performance in terms of model discrimination and calibration as a function of integrated area under the receiver operating characteristics (ROC) curve (AUC) (Uno et al. (2007)) and integrated Brier scores (Gerds, Cai and Schumacher (2008), Graf et al. (1999)). Figure 1 shows that all three variations of the proposed FPLS methods retain superior performance in comparison with the FPCR method. We see no improvements in the unsupervised FPCR method when the number of components K is increased from one to two. In this particular scenario, because B_3 explains the least amount of variation, the FPCR approach simply ignores this component and barely has any predictive power when we constrain the number of components $K = 1$ and 2, as can be seen with an integrated AUC scattered around 0.5. These results are in accordance with our expectations, because the FPCR approach extracts imaging components that explain most of the variation among image predictors, which are independent of the outcome variable.

To confirm that the proposed methods indeed captured the latent component that is most associated with the hazard function, Figure 2 displays the estimated $\mathbf{B}^T\hat{\mathbf{w}}$ when $K = 1$, from the three variations of FPLS (panels a–c) and FPCR (panel d) in comparison with the truth (panel e). As we can see, the top three panels in Figure 2 mimic the truth fairly well, whereas panel d retains a form that is similar to B_1, which explains most of the variation.

We further show a similar set of results under scenario 1 without fixing K. The K and λ are then selected automatically with a five-fold inner cross-validation. For a fair comparison, K under the FPCR method was chosen to explain at least 80% of the variability in the data. As shown in Figure 3, similar results can be drawn under this setting where all three FPLS retain superior performance in comparison to the FPCR.

In addition to the simulation results discussed above, simulation results under scenario 2 as well as under a plasmode simulation study mimicking the mammography data are included in the Supplementary Material (Jiang, Cao and Colditz (2024)). Similar conclusions can be drawn under these settings where the proposed methods retain superior prediction performance.

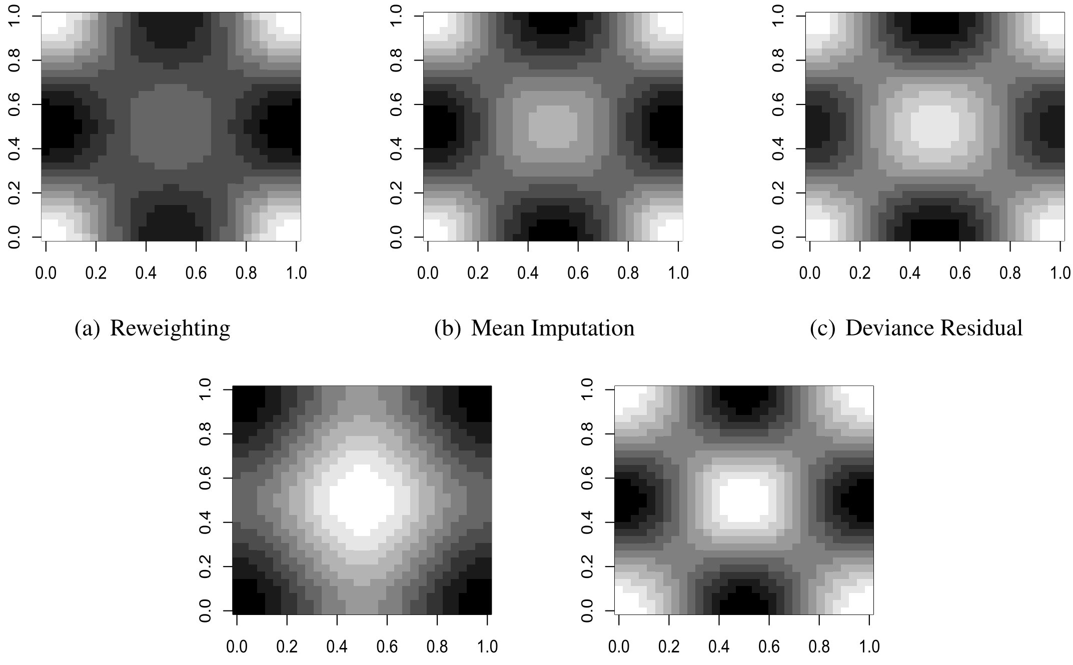


(a) AUC

(b) BS

FIG. 1. *Boxplots for the estimated integrated area under the receiver operating characteristic* (*ROC*) *curve* (*AUC*) *and integrated Brier scores* (*BS*) *with FPCR and three types of FPLS*: *M*1 (*reweighing*); *M*2 (*mean imputation*); *M*3 (*deviance residuals*) *under simulation scenario* 1 *with fixed* K *components and cross-validation selected* λ.

4. The Joanne Knight Breast Health Cohort. The Joanne Knight Breast Health Cohort (JKBH) was established to link breast cancer risk factors, mammographic breast density, and blood markers in a diverse population of women undergoing routine mammographic



(a) Reweighting

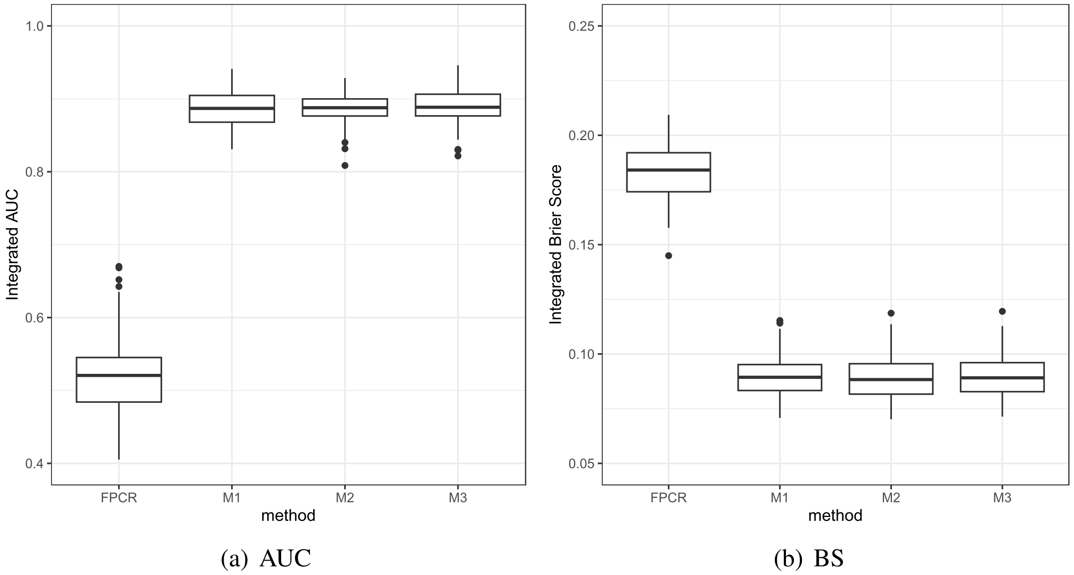
(b) Mean Imputation

(c) Deviance Residual

(d) FPCR

(e) Truth

FIG. 2. *Estimated* $\mathbf{B}^T\hat{\mathbf{w}}$ *for* $K = 1$, *from the three variations of FPLS* (*panels a–c*) *and FPCR* (*panel d*) *in comparison to the truth* (*panel e*) *for one randomly chosen simulation run under scenario* 1.



(a) AUC (b) BS

FIG. 3. *Boxplots for the estimated integrated area under the receiver operating characteristic* (*ROC*) *curve* (*AUC*) *and integrated Brier scores* (*BS*) *with FPCR and three types of FPLS*: $M1$ (*reweighing*); $M2$ (*mean imputation*); $M3$ (*deviance residuals*) *under simulation scenario* 1 *with cross-validation selected* K *and* λ.

screening. Women were recruited in St Louis, MO, from November 2008 to April 2012. Follow-up through 2018 was over 80% complete, and women continued to receive screening mammograms and follow-up diagnostics that included confirming pathology samples of any breast lesions and cancers. We used the nested case-control cohort within the JKBH cohort in this analysis, which consists of 947 women who had a full field digital craniocaudal view mammogram available at the baseline (Chen et al. (2023)). These women were matched by their baseline age and year of entry to the cohort (Colditz et al. (2022)). The mean age at entry was 56.7 years within the cohort and 289 women were diagnosed with breast cancer prior to the end of the follow-up (excluding those diagnosed within the first six months since entering the cohort). Mammograms at baseline were prealigned and averaged between the two breasts within a woman prior to analysis; see Jiang et al. (2023b) for more details.

For the breast cancer application, we use baseline age, BMI, history of benign breast biopsy (HBB), family history of breast cancer (FH), menopausal status (meno), parous (yes/no), age at menarche (age_m), breast density categorized in BI-RADS levels (A/B/C/D) with A being the reference category, and date of entry (ED) to the cohort as demographic variables. We project the mammogram imaging data onto a 12 × 12 tensor product of cubic B-spline, **B**, in estimating the basis coefficients, **w**, using the three methods proposed in this paper: reweighing, mean imputation, and deviance residual based FPLS. We have assessed three choices of tensor product dimensions with sensitivity analyses, and all three performed similarly in this application; see Supplementary Material for more details (Jiang, Cao and Colditz (2024)). Specifically, we consider the following proportional hazards model:

$$\begin{aligned} h_i(t) = h_0(t)\exp\{&\alpha_1\mathrm{age}_i + \alpha_2\mathrm{BMI}_i + \alpha_3\mathrm{HBB}_i + \alpha_4\mathrm{FH}_i + \alpha_5\mathrm{age\_m}_i + \alpha_6\mathrm{meno}_i \\ &+ \alpha_7\mathrm{parous}_i + \alpha_8\mathrm{ED}_i + \alpha_9\mathrm{biradB}_i + \alpha_{10}\mathrm{biradC}_i + \alpha_{11}\mathrm{biradD}_i + \beta^T\hat{\xi}_i\}, \end{aligned} \tag{8}$$

where $\beta = (\beta_1, \dots, \beta_K)^T$ denote the coefficients for K latent components. Here the proportional hazards assumption was deemed reasonable upon formally inspecting the Schoenfeld residual plot for each baseline covariate.

To avoid over-optimism on the prediction performance, we conduct a 10-fold internal cross-validation. For each of the training splits within the 10-fold, a nested five-fold cross-validation is used to find the number of latent components K and the smoothing parameter λ via a two-dimensional grid search (see Supplementary Material Section S2 for more details) (Jiang, Cao and Colditz (2024)). The value for the smoothing parameter is searched on the grid $(0, 1, 10, 10^2, 10^3, 10^6)$, and the number of latent components is searched on the grid of 1 to 30 with an equal increment of 1. In this study we report both the short-term risk prediction (one-year interval) and the long-term risk prediction (five-year interval) in terms of AUC. The standard deviation for the AUC in each fold has been estimated with 1000 bootstraps (LeDell, Petersen and van der Laan (2015)).

Table 1 illustrates the results under different model setups for short term (one-year) and long-term (five-year) risk prediction. The proposed FPLS formulations are compared with: 1) benchmark model with demographic variables only, that is, without using mammogram images, 2) Gail model which is a well-accepted breast cancer model across the globe (Gail et al. (1989)), 3) convolutional neural network (CNN), and 4) FPCR. We see that the benchmark and Gail model gave the lowest AUC. Although the AUC under these two models seem low, this is in accordance with the breast cancer literature; some of the widely used models in breast cancer include the Gail model and Tyrer–Cuzick model that results in a five-year AUC of approximately 0.58–0.61 (Maas et al. (2016), Brentnall et al. (2015)).

We then include the mammogram in the subsequent models in Table 1. We see that the CNN is much improved over the previous two models, suggesting that the mammogram image is informative in future risk. However, the CNN retained the worse performance for the five-year AUC in comparison to the functional models. This performance may be due to the moderate sample size in our dataset, as CNN usually requires massive training data to guarantee good performance (Keshari et al. (2018), Wagner et al. (2013)). Details on implementation along with tuning parameter selection for the CNN are embedded in the Supplementary Material (Jiang, Cao and Colditz (2024)). Next, we utilized features extracted from the mammogram images using FPCR and the three FPLS methods. As we can see, all three FPLS methods have outperformed the FPCR. To assess whether the latent components extracted using best performing mean imputation method from the mammogram images are significant, compared to the benchmark model, we have conducted a likelihood ratio test and have obtained a significant result, that is, p-value < 0.001 (Vickers, Cronin and Begg (2011), Demler, Pencina and D'Agostino (2012)).

The AUC itself does not provide much information in terms of the gain in clinical risk stratification for improvement in the low-end and high-end in the population, as suggested by one of the reviewers. Therefore, we have added the expected 10-year breast cancer risk by case-control status to assess the risk recalibration for the proposed model. The risk stratification is calibrated in the SEER population breast cancer database; see Rosner et al. (2021). After generating the individualized risk scores, we stratify women using the U.S. guideline categories, below average ($< 2\%$), average (2%–3%), above average (3%–5%), moderately increased (5%–8%), and high ($\geq 8\%$).

For clinical translational purposes, a new woman that comes into the clinic can now be categorized into one of the five risk categories listed above. For women that are in the very low-risk category, less frequent screening can be recommended, for example. For women that are in the very high-risk category, on the other hand, more frequent screening as well as preventive strategies can be recommended. Such gain in risk stratification using the proposed model (compare to the Gail model) can be seen in Table 2.

To further aid interpretation of the proposed method, we have added a regression with the corresponding coefficient surface estimated with mean imputation FPLS. In fact, the coefficient surface for the mammograms can be interpreted just as ordinary regression coefficients;

TABLE 1

One-year and five-year risk prediction performance using the Benchmark model with demographic covariates only, the Gail model, the convolutional neural network (CNN), the functional principal components regression (FPCR) model, and the three proposed functional partial least squares (FPLS) methods

	One-year AUC		Five-year AUC	
	Est.	SE	Est.	SE
Benchmark	0.648	0.05	0.606	0.05
Gail	0.588	0.06	0.569	0.06
CNN	0.773	0.06	0.626	0.06
FPCR	0.734	0.06	0.651	0.06
FPLS—Reweighting	0.807	0.07	0.703	0.06
FPLS—Mean Imputation	0.812	0.05	0.717	0.05
FPLS—Deviance Residual	0.810	0.05	0.702	0.05

see Figure 4 that outlines this relationship between the covariates X_i and the image $Z_i(s)$. For instance, we see that the positive coefficient (hot color) indicates a worse prognosis and a negative coefficient (cold color) indicates a protective effect of the pixels to breast cancer risk. The coefficient surface for the mammogram image is of the same dimension as the image, and thus each estimate on the surface has a one-to-one matching back to the pixels on the image. Therefore, we can interpret this as the weighted average between the pixel intensities and the coefficient surface that affects the breast cancer risk outcome.

5. Discussion. This paper presented three formulations of the FPLS method based on reweighing, mean imputation, and deviance residual based approach to extract latent components that are most associated with survival outcomes accompanied with right-censoring. While there exists several different methodologies on identifying the low-rank structure from the image predictors with survival outcome, these low-rank structures may or may not be associated with the survival outcome. The select latent components identified using the proposed FPLS method are ordered by their association with the survival outcome. This paper also investigates and compares the prediction performance of the proposed FPLS methods with existing breast cancer models and convolutional neural network in the presence of a high-dimensional imaging predictor. We have assessed the model performance via intensive simulation studies and demonstrated that all three formulations of the FPLS outperform the unsupervised FPCR. The three proposed formulations of the FPLS are then applied to the motivating dataset from the Joanne Knight Breast Health Cohort at Siteman Cancer Center where the goal lies in improving the risk prediction performance by fully utilizing the mammogram imaging data that are routinely available. Our results suggest that latent features

TABLE 2

Risk stratification using the proposed FPLS method using mean imputation vs. the Gail model calibrated with the SEER database

	FPLS		Gail	
	Cases	Controls	Cases	Controls
below average ($< 2\%$)	15%	48%	9%	18%
average (2%–3%)	20%	25%	40%	47%
above average (3%–5%)	34%	16%	33%	31%
moderately increased (5%–8%)	18%	9%	11%	3%
high ($\geq 8\%$)	13%	2%	7%	1%

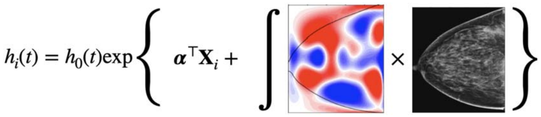

$h_i(t) = h_0(t)\exp\Big\{ \boldsymbol{\alpha}^\top \mathbf{X}_i + \int$
×
$\Big\}$

FIG. 4. *Visual representation of Cox regression with the corresponding coefficient surface estimated with mean imputation FPLS.*

extracted from the baseline mammogram images are informative in predicting breast cancer risk, as demonstrated with increased discriminatory performance. For our application all three FPLS methods performed similarly in terms of one-year and five-year AUC; the mean imputation method performed slightly better among all three.

In general, the appropriateness the functional data analysis hinges on the assumption that the observed $Z_i(\boldsymbol{s})$ are realizations of a stochastic process $\{Z(\boldsymbol{s}), \forall \boldsymbol{s} \in \mathbb{S}\}$ in a square integrable rectangle in IR^2: $L^2(\mathbb{S}) = \{f : \mathbb{S} \to \mathrm{IR} || \int_{\mathbf{s}\in\mathbb{S}} f(\mathbf{s})^2 \, d\mathbf{s}| < \infty\}$. In the current breast cancer application, all women who enter the cohort are constraint to be cancer free and tumor-free within at least six months since baseline (average year from entry to diagnosis of breast cancer was 5.19 years). Thus, at entry to the cohort, we are modeling the breast tissue distribution across women and not the diagnostic locations of the potential tumor. It has been well documented that patterns of breast parenchymal complexity are formed by the x-ray attenuation of fatty, fibroglandular, and stromal tissues. This is assumed to be comparable across healthy breasts, because each region within the breast has equal chance to develop future tumor. Individual-specific deviation from the underlying distribution in the population can then be used to identify very low- and very high-risk individuals. While this assumption is deemed reasonable in this particular scenario, it may be violated when tumors are present on the set of images. As pointed out by one of the reviewers, when tumors have developed in different regions of the breast across women, this assumption of a common underlying distribution may not be suitable.

The proposed model in this paper has room for improvement in prediction performance by incorporating other risk factors, such as the questionnaire scores and genetic risk factors (Rosner et al. (2021)). If 3D images are available, the proposed method to the 3D coordinate system encompasses a trivial extension by replacing $B_m(s)$ by a 3D basis function. As these remain part of future exploration, the current proposed modeling framework sets a foundation for incorporating the mammogram and other types of images for risk prediction to aid long-term prevention. Our approach can also be used to assess the added value of image data in other contexts, for example, in predicting the risk of Alzheimer's Disease with brain images.

Acknowledgments. The authors wish to thank the Editor and anonymous referees of this article for their valuable and thoughtful input. These suggestions have significantly enhanced the quality of the manuscript.

Funding. This research is partially supported by the NIH (R37 CA256810), Breast Cancer Research Foundation (BCRF 20-028), and the NSERC Discovery grant (RGPIN-2018-06008).

SUPPLEMENTARY MATERIAL

Supplement for "Functional partial least squares with censored outcomes: Prediction of breast cancer risk with mammogram images" (DOI: 10.1214/23-AOAS1822SUPP; .pdf). Tables, figures, and additional simulation and application results.

REFERENCES

ANANDARAJAH, A., CHEN, Y., COLDITZ, G. A., HARDI, A., STOLL, C. and JIANG, S. (2022). Studies of parenchymal texture added to mammographic breast density and risk of breast cancer: A systematic review of the methods used in the literature. *Breast Cancer Res*. **24** 1–18.

ANANDARAJAH, A., CHEN, Y., STOLL, C., HARDI, A., JIANG, S. and COLDITZ, G. A. (2023). Repeated measures of mammographic density and texture to evaluate prediction and risk of breast cancer: A systematic review of the methods used in the literature. *Cancer Causes Control* 1–10.

BASTIEN, P., BERTRAND, F., MEYER, N. and MAUMY-BERTRAND, M. (2015). Deviance residuals-based sparse PLS and sparse kernel PLS regression for censored data. *Bioinformatics* **31** 397–404.

BOYD, N. F., MARTIN, L. J., ROMMENS, J. M., PATERSON, A. D., MINKIN, S., YAFFE, M. J., STONE, J. and HOPPER, J. L. (2009). Mammographic density: A heritable risk factor for breast cancer. In *Cancer Epidemiology* 343–360.

BRENTNALL, A. R., HARKNESS, E. F., ASTLEY, S. M., DONNELLY, L. S., STAVRINOS, P., SAMPSON, S., FOX, L., SERGEANT, J. C., HARVIE, M. N. et al. (2015). Mammographic density adds accuracy to both the Tyrer–Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. **17** 1–10.

CHEN, S., TAMIMI, R. M., COLDITZ, G. A. and JIANG, S. (2023). Association and prediction utilizing craniocaudal and mediolateral oblique view digital mammography and long-term breast cancer risk. *Cancer Prev. Res*. OF1–OF8.

COLDITZ, G. A., BENNETT, D. L., TAPPENDEN, J., BEERS, C., ACKERMANN, N., WU, N., LUO, J., HUMBLE, S., LINNENBRINGER, E. et al. (2022). Joanne Knight breast health cohort at Siteman Cancer Center. *Cancer Causes Control* **33** 623–629.

DATTA, S. (2005). Estimating the mean life time using right censored data. *Stat. Methodol*. **2** 65–69.

DATTA, S., LE-RADEMACHER, J. and DATTA, S. (2007). Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO. *Biometrics* **63** 259–271.

DEMLER, O. V., PENCINA, M. J. and D'AGOSTINO, R. B. SR. (2012). Misuse of DeLong test to compare AUCs for nested models. *Stat. Med*. **31** 2577–2587.

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. *J. Natl. Cancer Inst*. **81** 1879–1886. https://doi.org/10.1093/jnci/81.24.1879

GERDS, T. A., CAI, T. and SCHUMACHER, M. (2008). The performance of risk prediction models. *Biom. J*. **50** 457–479.

GRAF, E., SCHMOOR, C., SAUERBREI, W. and SCHUMACHER, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Stat. Med*. **18** 2529–2545. https://doi.org/10.1002/1097-0258

JIANG, S., CAO, J. and COLDITZ, G. A (2024). Supplement to "Functional partial least squares with censored outcomes: Prediction of breast cancer risk with mammogram images." https://doi.org/10.1214/23-AOAS1822SUPP

JIANG, S., CAO, J., COLDITZ, G. A. and ROSNER, B. (2023a). Predicting the onset of breast cancer using mammogram imaging data with irregular boundary. *Biostatistics* **24** 358–371.

JIANG, S., CAO, J., ROSNER, B. and COLDITZ, G. A. (2023b). Supervised two-dimensional functional principal component analysis with time-to-event outcomes and mammogram imaging data. *Biometrics* **79** 1359–1369.

JIANG, S. and COLDITZ, G. A. (2023). Causal mediation analysis using high-dimensional image mediator bounded in irregular domain with an application to breast cancer. *Biometrics*. https://doi.org/10.1111/biom.13847

KESHARI, R., VATSA, M., SINGH, R. and NOORE, A. (2018). Learning structure and strength of CNN filters for small sample size training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 9349–9358.

KONG, D., IBRAHIM, J. G., LEE, E. and ZHU, H. (2018). FLCRM: Functional linear Cox regression model. *Biometrics* **74** 109–117.

LEDELL, E., PETERSEN, M. and VAN DER LAAN, M. (2015). Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron. J. Stat*. **9** 1583.

LI, H. and GUI, J. (2004). Partial Cox regression analysis for high-dimensional microarray gene expression data. *Bioinformatics* **20** i208–i215.

MAAS, P., BARRDAHL, M., JOSHI, A. D., AUER, P. L., GAUDET, M. M., MILNE, R. L., SCHUMACHER, F. R., ANDERSON, W. F., CHECK, D. et al. (2016). Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol*. **2** 1295–1302.

MARTENS, H. and NÆS, T. (1992). *Multivariate Calibration*. Wiley, Chichester.

NYGÅRD, S., BORGAN, Ø., LINGJÆRDE, O. C. and STØRVOLD, H. L. (2008). Partial least squares Cox regression for genome-wide data. *Lifetime Data Anal*. **14** 179–195.

PARK, P. J., TIAN, L. and KOHANE, I. S. (2002). Linking gene expression data with patient survival times using partial least squares. *Bioinformatics* **18** S120–S127.

PASHAYAN, N., MORRIS, S., GILBERT, F. J. and PHAROAH, P. D. P. (2018). Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: A life-table model. *JAMA Oncol.* **4** 1504–1510.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York.

REISS, P. T. and OGDEN, R. T. (2007). Functional principal component regression and functional partial least squares. *J. Amer. Statist. Assoc.* **102** 984–996.

ROSNER, B., TAMIMI, R. M., KRAFT, P., GAO, C., MU, Y., SCOTT, C., WINHAM, S. J., VACHON, C. M. and COLDITZ, G. A. (2021). Simplified breast risk tool integrating questionnaire risk factors, mammographic density, and polygenic risk score: Development and validation. *Cancer Epidemiol. Biomark. Prev.* **30** 600–607.

SEGAL, M. R. (2006). Microarray gene expression data with linked survival phenotypes: Diffuse large-B-cell lymphoma revisited. *Biostatistics* **7** 268–285.

TABAR, L., GAD, A., HOLMBERG, L., LJUNGQUIST, U., FAGERBERG, C., BALDETORP, L., GRÖNTOFT, O., LUNDSTRÖM, B., MÅNSON, J. et al. (1985). Reduction in mortality from breast cancer after mass screening with mammography: Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* **325** 829–832.

UNO, H., CAI, T., TIAN, L. and WEI, L. J. (2007). Evaluating prediction rules for t-year survivors with censored regression models. *J. Amer. Statist. Assoc.* **102** 527–537.

VICKERS, A. J., CRONIN, A. M. and BEGG, C. B. (2011). One statistical test is sufficient for assessing new predictive markers. *BMC Med. Res. Methodol.* **11** 1–7.

VILMUN, B. M., VEJBORG, I., LYNGE, E., LILLHOLM, M., NIELSEN, M., NIELSEN, M. B. and CARLSEN, J. F. (2020). Impact of adding breast density to breast cancer risk models: A systematic review. *Eur. J. Radiol.* **127** 109019.

VISVANATHAN, K., FABIAN, C. J., BANTUG, E., BREWSTER, A. M., DAVIDSON, N. E., DECENSI, A., FLOYD, J. D., GARBER, J. E., HOFSTATTER, E. W. et al. (2019). Use of endocrine therapy for breast cancer risk reduction: ASCO clinical practice guideline update. *J. Clin. Oncol.* **37** 3152–3165.

WAGNER, R., THOM, M., SCHWEIGER, R., PALM, G. and ROTHERMEL, A. (2013). Learning convolutional neural networks from few samples. In *The* 2013 *International Joint Conference on Neural Networks* (*IJCNN*) 1–7. IEEE Press, New York.

WOLD, H. (1966). Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis* (*Proc. Internat. Sympos.*, *Dayton*, *Ohio*, 1965) 391–420. Academic Press, New York.

WOLD, H. (1975a). Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. *J. Appl. Probab.* **12** 117–142.

WOLD, H. (1975b). Path models with latent variables: The NIPALS approach. In *Quantitative Sociology* 307–357.