

UCLA

UCLA Previously Published Works

Title

Scalable Algorithms for Large Competing Risks Data

Permalink

<https://escholarship.org/uc/item/2p8647b1>

Journal

Journal of Computational and Graphical Statistics, 30(3)

ISSN

1061-8600

Authors

Kawaguchi, Eric S
Shen, Jenny I
Suchard, Marc A
[et al.](#)

Publication Date

2021-07-03

DOI

10.1080/10618600.2020.1841650

Peer reviewed



Published in final edited form as:

J Comput Graph Stat. 2021 ; 30(3): 685–693. doi:10.1080/10618600.2020.1841650.

Scalable Algorithms for Large Competing Risks Data

Eric S. Kawaguchi¹, Jenny I. Shen², Marc A. Suchard^{3,4,5}, Gang Li^{3,4,*}

¹Department of Preventive Medicine, University of Southern California

²Division of Nephrology and Hypertension Los Angeles Biomedical Institute at Harbor-UCLA Medical Center

³Department of Biostatistics, University of California, Los Angeles

⁴Department of Computational Medicine, University of California, Los Angeles

⁵Department of Human Genetics, University of California Los Angeles

Abstract

This paper develops two orthogonal contributions to scalable sparse regression for competing risks time-to-event data. First, we study and accelerate the broken adaptive ridge method (BAR), a surrogate ℓ_1 -based iteratively reweighted ℓ_2 -penalization algorithm that achieves sparsity in its limit, in the context of the Fine-Gray (1999) proportional subdistributional hazards (PSH) model. In particular, we derive a new algorithm for BAR regression, named cycBAR, that performs cyclic update of each coordinate using an explicit thresholding formula. The new cycBAR algorithm effectively avoids fitting multiple reweighted ℓ_2 -penalizations and thus yields impressive speedups over the original BAR algorithm. Second, we address a pivotal computational issue related to fitting the PSH model. Specifically, the computation costs of the log-pseudo likelihood and its derivatives for PSH model grow at the rate of $\mathcal{O}(n^2)$ with the sample size n in current implementations. We propose a novel forward-backward scan algorithm that reduces the computation costs to $\mathcal{O}(n)$. The proposed method applies to both unpenalized and penalized estimation for the PSH model and has exhibited drastic speedups over current implementations. Finally, combining the two algorithms can yields $> 1,000$ fold speedups over the original BAR algorithm. Illustrations of the impressive scalability of our proposed algorithm for large competing risks data are given using both simulations and a United States Renal Data System data. Supplementary materials for this article are available online.

Keywords

Broken Adaptive Ridge; Fine-Gray model; ℓ_1 -regularization; Massive Sample Size; Model Selection/Variable selection; Oracle property; Subdistribution hazard

* vli@ucla.edu .

Supplementary Materials

Supplementary materials are available online and include:

kawaguchi2020supp: PDF file that contains proofs, supplemental figures, and supplemental tables referenced within the main text. (kawaguchi2020supp.pdf)

kawaguchi2020sim: Zip file that contains R code to reproduce simulation results described in the main text and in kawaguchi2020supp.pdf. A README file is provided. (kawaguchi2020sim.zip)

1 Introduction

Advancing informatics tools make large-scale data such as electronic health record (EHR) data and genomic data routinely accessible to researchers. This data deluge offers unprecedented opportunities for new and innovative approaches to improve research and learning (Schuemie et al., 2018). However, it also presents new computational challenges and barriers for quantitative researchers as many current statistical methodologies and computational tools may grind to a halt as the sample size (n) and/or number of covariates (p_n) grow large. Such challenges are particularly common in time-to-event data analysis where the likelihood function (such as the partial likelihood for the Cox model with data) and its derivatives typically require $O(n^2)$ number of operations, which will explode quickly as n increases. The computational burden can be further aggravated as the number of covariates (p_n) increases. Statistical methods coupled with high-performance algorithms are critically needed for large-scale time-to-event data analysis.

This paper aims to develop high-performance computational methods for large-scale competing risks time-to-event data analysis by addressing two orthogonal computational challenges due to large p_n and large n respectively. First, we develop a scalable surrogate ℓ_1 -based method for simultaneous variable selection and parameter estimation for the large p_n problem. It is well known that ℓ_1 -penalized regression is natural for variable selection (Breiman, 1996; Shen et al., 2012), but is computationally NP hard and not scalable to even moderate p_n . As a scalable approximation to ℓ_1 -penalized regression, the broken adaptive ridge BAR estimator, defined as the limit of an ℓ_1 -based iteratively reweighted ℓ_2 -penalization algorithm, has been recently studied for simultaneous variable selection and parameter estimation and shown to possess some desirable selection, estimation, and grouping properties under various model settings (see, e.g., Zhao et al. (2018), Dai et al. (2018), Zhao et al. (2019b), Zhao et al. (2019a), and Kawaguchi et al. (2020)). Since BAR requires fitting multiple reweighted ℓ_2 -penalized regressions until convergence, it is not computationally as efficient as other single step penalization methods such as SCAD and MCP, especially when a large number of iterations is needed for convergence. As demonstrated in Section 4 (Table 1), BAR can grind to a halt for large data, which calls for more efficient BAR algorithms. Second, we address a pivotal computational issue specifically related to fitting the PSH model when n is large. In Section 2.4, we will show how the computation of the log-pseudo likelihood and its derivatives for the PSH model involves $O(n^2)$ number of operations, and that commonly used efficient computational techniques for fitting the classical Cox (1972) model do not apply to the PSH model since the computations involve weighted sums over some risk sets where the weights are subject-specific and the risks sets are not monotone over time. To the best of our knowledge, no algorithm has been developed in the literature to reduce the computational cost for the PSH model from $O(n^2)$ to a lower order.

In addressing the aforementioned computational challenges for large data, the contribution of this paper is two fold:

1. We propose a novel cyclic coordinate-wise update algorithm for BAR, referred to as *cycBAR*, by deriving an explicit analytic coordinate-wise update for a

fixed-point problem whose unique solution approximates the BAR estimator. Because the cycBAR algorithm avoids carrying out iteratively reweighted ℓ_2 -penalizations, it can result in substantial gains in computational efficiency. We emphasize that the application of the cycBAR algorithm over the original BAR method is not limited to the PSH model and spans a variety of models and data settings such as generalized linear models and time-to-event models, as well as sparse signal reconstruction (Gorodnitsky and Rao, 1997) and compressive sensing Candes et al., 2008; Chartrand and Yin, 2008; Gasso et al., 2009; Daubechies et al., 2010; Wipf and Nagarajan, 2010 where the ℓ_2 -based iteratively reweighted ℓ_2 -penalization algorithm are popularly used. In our numerical studies (Section 3.3, Figure 1(b)), cycBAR shows marked reduction in runtime over the standard BAR implementation.

2. By exploiting the special structure of the risk set and the subject-specific weight functions associated with the Fine-Gray pseudo likelihood and its derivatives, we derive a novel forward-backward scan algorithm to reduce their computational costs from $O(n^2)$ to $O(n)$, allowing one to analyze competing risks data much quicker than current approaches. We have observed in empirical studies, e.g. Figure 1(c) in Section 3.3, that the forward-backward scan algorithm can yield dramatic speedups over standard implementations. We point out that our proposed forward-backward scan algorithm for the PSH model is not specific to the BAR method and can be applied to accelerate other penalized regression methods such as LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), adaptive LASSO (Zou, 2006), and MCP (Zhang, 2010) for the PSH model (Fu et al., 2017), and the unpenalized estimation method of Fine and Gray (1999), as well as hypothesis testing and cumulative incidence estimation for the PSH model.

The rest of this article is organized as follows. In Section 2.1, we review the mathematical formulation of competing risks data and the Fine and Gray (1999) proportional subdistribution hazards model. Section 2.2, introduces the BAR estimator for the PSH model and refers its asymptotic properties to the Online Supplementary Material. Section 2.3 derives the cyclic coordinate-wise BAR algorithm. The forward-backward scan method for the PSH model is described in Section 2.4. Section 3 presents some simulation studies to demonstrate the computational efficiency gains of both the cycBAR and forward-backward scan algorithms. We provide a proof-of-concept real data example for fitting large-scale competing risks data in Section 4 using a subset of the United States Renal Data System (USRDS). Lastly, we give concluding remarks in Section 5. The proposed method has been implemented in an R package, named **pshBAR**, which is available at <https://github.com/erickawaguchi/pshBAR>.

2 Methodology

2.1 Competing risks data, model, and parameter estimation

Competing risks time-to-event data arises frequently in clinical trials, reliability testing, social science, and many other fields (Prentice et al., 1978; Pintilie, 2006; Putter et al., 2007). Competing risks occur when individuals are susceptible to more than one types of

possibly correlated events or causes and the occurrence of one event precludes the others from happening. For example, one may wish to study time until first kidney transplant for kidney dialysis patients with end-stage renal disease. Then terminating events such as death, renal function recovery, or discontinuation of dialysis are competing risks as their occurrence will prevent subjects from receiving a transplant. For $i = 1, \dots, n$, let T_i , C_i , ϵ_i , and \mathbf{z}_i be the event time, possible right-censoring time, cause (event type), and a p_n -dimensional vector of time-independent covariates, respectively, for subject i . Without loss of generality assume there are two event types $\epsilon \in \{1, 2\}$ where $\epsilon=1$ is the event of interest and $\epsilon=2$ is the competing risk. With the presence of right censoring, we generally observe $X_i = T_i \wedge C_i$, $\delta_i = I(T_i < C_i)$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. Competing risks data consists of n independent and identically distributed quadruplets $\{(X_i, \delta_i, \epsilon_i, \mathbf{z}_i)\}_{i=1}^n$. Assume that there exists a τ such that (1) for some arbitrary time t , $t \in [0, \tau]$; (2) $\Pr(T_i > \tau) > 0$ and $\Pr(C_i > \tau) > 0$ for all $i=1, \dots, n$.

An important quantity for competing risks data is the cumulative incidence function (CIF), which describes the probability of failing from a certain cause of interest before the other causes. The CIF for cause 1 events conditional on the covariates is defined as $F_1(t; \mathbf{z}) = \Pr(T_{i, \epsilon=1} \leq t | \mathbf{z})$. To model $F_1(t; \mathbf{z})$, Fine and Gray (1999) introduced the now popular proportional subdistribution hazards (PSH) model:

$$h_1(t | \mathbf{z}) = h_{10}(t) \exp(\mathbf{z}'\boldsymbol{\beta}), \tag{1}$$

where

$$h_1(t | \mathbf{z}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T \leq t + \Delta t, \epsilon = 1 | T \geq t \cup (T \leq t \cap \epsilon \neq 1), \mathbf{z}\}}{\Delta t} = -\frac{d}{dt} \log\{1 - F_1(t; \mathbf{z})\}$$

is a subdistribution hazard (Gray, 1988), $h_{10}(t)$ is a completely unspecified baseline subdistribution hazard, and $\boldsymbol{\beta}$ is a $p_n \times 1$ vector of regression coefficients. As Fine and Gray (1999) mentioned, the risk set associated with $h_1(t; \mathbf{z})$ is somewhat counterfactual as it includes subjects who are still at risk ($T > t$) and those who have already observed the competing risk prior to time t ($T \leq t \cap \epsilon \neq 1$). However, this construction is useful for direct modeling of the CIF.

Inference for the PSH model based on the following log-pseudo likelihood (Fine and Gray, 1999):

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \left(\mathbf{z}_i' \boldsymbol{\beta} - \log \left\{ \sum_j \hat{w}_j(s) Y_j(s) \exp(\mathbf{z}_j' \boldsymbol{\beta}) \right\} \right) \times \hat{w}_i(s) dN_i(s), \tag{2}$$

where $N_i(t) = I(T_i \leq t, \epsilon_i = 1)$, $Y_i(t) = 1 - N_i(t-)$, $\hat{w}_i(t)$ is a time-dependent weight for subject i at time t defined as $\hat{w}_i(t) = I(C_i \geq T_i \wedge t) \hat{G}(t) / \hat{G}(X_i \wedge t)$, and $\hat{G}(t)$ is the Kaplan and Meier (1958) estimate for $G(t) = \Pr(C > t)$, the survival function of the censoring variable C . Note that, for any subject i and time t , $\hat{w}_i(t) Y_i(t) = 0$ if an individual is right censored or has

experienced the event of interest; and $\widehat{w}_i(t)Y_i(t) = 1$ if $t < X_i$, and $\widehat{w}_i(t)Y_i(t) = \widehat{G}(t)/\widehat{G}(X_i)$ for events due to the competing risk.

Commonly-used optimization routines to estimate the parameters of the PSH model typically require the calculation of the log-pseudo likelihood (2), the score function

$$i_{j}(\boldsymbol{\beta}) = \sum_{i=1}^n I(\delta_i \epsilon_i = 1) z_{ij} - \sum_{i=1}^n I(\delta_i \epsilon_i = 1) \frac{\sum_{k \in R_i} z_{kj} \widetilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \widetilde{w}_{ik} \exp(\eta_k)}, \quad (3)$$

and, in some cases, the Hessian diagonals

$$\ddot{i}_{jj}(\boldsymbol{\beta}) = \sum_{i=1}^n I(\delta_i \epsilon_i = 1) \left[\frac{\sum_{k \in R_i} z_{kj}^2 \widetilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \widetilde{w}_{ik} \exp(\eta_k)} - \left\{ \frac{\sum_{k \in R_i} z_{kj} \widetilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \widetilde{w}_{ik} \exp(\eta_k)} \right\}^2 \right], \quad (4)$$

where

$$\widetilde{w}_{ik} = \widehat{w}_k(X_i) = \widehat{G}(X_i) / \widehat{G}(X_i \wedge X_k), k \in R_i,$$

$R_i = \{y : (X_y = X_i) \cup (X_y = X_i \cap \epsilon_y = 2)\}$ and $\eta_k = \mathbf{z}_k' \boldsymbol{\beta}$. Direct calculations using the above formulas will need $O(n^2)$ operations due to the the double summations and is computationally taxing for large n . We will show how to calculate the double summation linearly in Section 2.4, allowing us to calculate these quantities in $O(n)$ time.

2.2 Broken adaptive ridge estimation for the proportional subdistribution hazards model

Penalized regression is useful for simultaneous variable selection and parameter estimation and has recently been introduced to the PSH model for competing risks data Ha et al., 2014; Fu et al., 2017; Ahn et al., 2018; Hou et al., 2018. Below we extend the broken adaptive ridge (BAR) estimator to the PSH model.

Let $l(\boldsymbol{\beta})$ be the log-pseudo likelihood defined by (2). The BAR estimator of $\boldsymbol{\beta}$ starts with an initial ξ_n -penalized (or ridge) estimator

$$\boldsymbol{\beta}^{(0)} = \arg \min_{\boldsymbol{\beta}} \{-2l(\boldsymbol{\beta}) + \xi_n \sum_{j=1}^p \beta_j^2\}, \quad (5)$$

which is updated iteratively by a reweighted ξ_n -penalized estimator

$$\boldsymbol{\beta}^{(k)} = \arg \min_{\boldsymbol{\beta}} \left\{ -2l(\boldsymbol{\beta}) + \lambda_n \sum_{j=1}^p \frac{\beta_j^2}{|\widehat{\beta}_j^{(k-1)}|^2} \right\}, k \geq 1, \quad (6)$$

where ξ_n and λ_n are non-negative penalization tuning parameters. The BAR estimator of $\boldsymbol{\beta}$ is defined as the limit of this iterative algorithm:

$$\boldsymbol{\beta} = \lim_{k \rightarrow \infty} \boldsymbol{\beta}^{(k)}, \quad (7)$$

which can be viewed as a surrogate to ℓ_2 -penalized regression.

Note that adaptively reweighting the penalty of a coefficient by the inverse of its squared estimate from the previous iteration allows each coefficient to be penalized differently. At each successive iteration, coefficients whose true values are zero will have larger penalties that will shrink the estimate further towards zero. We have shown in Section S1 of the Online Supplementary Material that the BAR estimator has an oracle property for selection and estimation and a grouping property for highly correlated covariates.

The BAR estimator can be implemented using the algorithm outlined in Section S2.1 Algorithm S1 of the Online Supplementary Material in which a cyclic coordinate decent (CCD) algorithm is employed for each reweighted ℓ_2 -penalized regression. Because the algorithm runs a sequence ($k = 0, 1, \dots$) of adaptively reweighted ridge regressions, it adds an extra layer of computational complexity as compared to other popular single-step penalization methods such as LASSO and can create a bottleneck when a large number of iterations is needed. Moreover, because ridge regression is not sparse and thus the limit is never achieved at any given step of the BAR algorithm, an arbitrarily small cutoff value ϵ^* has to be used to induce sparsity in Algorithm S1 (line 18), which is an unpleasant feature. Below we show that these issues can be avoided using a new cyclic BAR algorithm.

2.3 A cyclic coordinate-wise BAR algorithm

In this section, we derive a fast cyclic coordinate-wise BAR algorithm that will result in the elimination of performing multiple ridge regressions and avoid using a cutoff ϵ^* to introduce sparsity as required by the original BAR algorithm (Algorithm S1 in the Online Supplementary Material). For a consistent estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, consider the Cholesky decomposition $-\ddot{i}(\boldsymbol{\beta}) = \mathbf{X}'\mathbf{X}$ and define $\mathbf{y} = (\mathbf{X}')^{-1} \{-\ddot{i}(\boldsymbol{\beta})\boldsymbol{\beta} + \dot{i}(\boldsymbol{\beta})\}$ as the pseudo-response vector. Approximating the negative log-pseudo likelihood by $-l(\boldsymbol{\beta}) \approx (1/2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ using a second-order Taylor expansion in (6) leads to the following solution

$$\boldsymbol{\beta}^{(k)} = g(\boldsymbol{\beta}^{(k-1)}),$$

where $g(\boldsymbol{\beta}) = \{\mathbf{X}'\mathbf{X} + \lambda_n D(\boldsymbol{\beta})\}^{-1} \mathbf{X}'\mathbf{y}$, and $D(\boldsymbol{\beta}) = \text{diag}(\beta_1^{-2}, \dots, \beta_{p_n}^{-2})$. Hence, as $k \rightarrow \infty$, the limit of the sequence $\{\boldsymbol{\beta}^{(k)}\}$ is the fixed point of the function $g(\cdot)$ or the solution of $g(\boldsymbol{\beta}) = \boldsymbol{\beta}$.

Remark 2.1.—*Floating errors can arise when calculating $D(\boldsymbol{\beta})$ as it involves the inverse of β^2 . However, this can be avoided by rewriting $g(\boldsymbol{\beta})$ as*

$$\begin{aligned} g(\boldsymbol{\beta}) &= \{\mathbf{X}'\mathbf{X} + \lambda_n D(\boldsymbol{\beta})\}^{-1} \mathbf{X}'\mathbf{y} \\ &= D(\boldsymbol{\beta})^{-1/2} \{D(\boldsymbol{\beta})^{-1/2} \mathbf{X}'\mathbf{X} D(\boldsymbol{\beta})^{-1/2} + \lambda_n \mathbf{I}_{p_n}\}^{-1} D(\boldsymbol{\beta})^{-1/2} \mathbf{X}'\mathbf{y}, \end{aligned}$$

which involves only multiplication rather than division by β .

The next theorem shows that each component of the fixed-point solution of g can be expressed as a function of all other components. The proof is deferred to Section S1.5 of the Online Supplementary Material.

Theorem 1.—Let β be the fixed-point solution of $g(\cdot)$. Then, for each $j = 1, \dots, p_n$, the j th component of β can be expressed as follows

$$\hat{\beta}_j = g_j(\beta_{-j}) \equiv \begin{cases} 0, & \text{if } |b_j| < 2\sqrt{\lambda_n \mathbf{x}'_j \mathbf{x}_j}, \\ \frac{b_j + \text{sign}(b_j)\sqrt{(b_j)^2 - 4\lambda_n \mathbf{x}'_j \mathbf{x}_j}}{2\mathbf{x}'_j \mathbf{x}_j}, & \text{otherwise,} \end{cases} \quad (8)$$

where $b_j = \mathbf{x}'_j(\mathbf{y} - \sum_{i \neq j} \mathbf{x}_i \hat{\beta}_i)$ and $\beta_{-j} = (\hat{\beta}_1, \dots, \hat{\beta}_{j-1}, \hat{\beta}_{j+1}, \dots, \hat{\beta}_{p_n})'$.

The above result motivates our cyclic coordinate-wise broken adaptive ridge (cycBAR) algorithm which performs cyclic coordinate-wise updates for the fixed point of $g(\cdot)$ using equation (8) as outlined in Algorithm 1 below. In Algorithm 1, \mathbf{X} and \mathbf{y} are initially estimated using the initial ridge estimate $\beta^{(0)}$ and then subsequently updated at step s using the previous estimate $\beta^{(s-1)}$ for $s \geq 1$. Consequently, at step s , we have

$$b_j^{(s)} \equiv \mathbf{x}'_j \left\{ \mathbf{y} - \sum_{i \neq j} \mathbf{x}_i \beta_i^{(s-1)} \right\} = -\ddot{l}_{jj}(\beta^{(s-1)})\beta_j^{(s-1)} + \dot{l}_j(\beta^{(s-1)}), \text{ for } j = 1, \dots, p_n,$$

where $\dot{l}_j(\beta)$ is the j th element of $-\dot{l}(\beta)$ and $-\ddot{l}_{jj}(\beta)$ is the j th diagonal element of $\ddot{l}(\beta)$. Note that an unpenalized estimator at each iteration may also be used in place of $\beta^{(s-1)}$ to construct \mathbf{X} and \mathbf{y} , which could conceivably reduce estimation bias with some increased computational cost. This is corroborated by our limited simulation studies (not reported here), which also showed that the performance differences between the two methods become negligible as the sample size increases.

Algorithm 1:

The cycBAR algorithm

-
- 1 Set $\beta^{(0)} = \beta_{\text{ridge}}$;
 - 2 for $s = 1, 2, \dots$ do
 - 3 # Enter cyclic coordinate-wise BAR algorithm
 - 4 for $j = 1, \dots, p_n$ do
 - 5 Calculate $c_{1j} = -\dot{l}_j(\beta^{(s-1)})$, $c_{2j} = -\ddot{l}_{jj}(\beta^{(s-1)})$ and $b_j^{(s)} = c_{2j}\beta_j^{(s-1)} - c_{1j}$;
 - 6 if $|b_j^{(s)}| < 2\sqrt{c_{2j}\lambda_n}$ then
 - 7 $\beta_j^{(s)} = 0$;
 - 8 else

$$9 \beta_j^{(s)} = \frac{b_j^{(s)} + \text{sign}(b_j^{(s)}) \sqrt{(b_j^{(s)})^2 - 4c_{2j}\lambda_n}}{2c_{2j}};$$

10 end

11 end

12 if $\|\beta^{(s)} - \beta^{(s-1)}\| < \text{tol}$ then

13 $\beta_{BAR} = \beta^{(s)}$ and break;

14 end

15 end

Remark 2.2. (cycBAR versus BAR)—The cycBAR algorithm is derived by approximating the log-likelihood with a quadratic approximation, so it provides an approximation of the BAR estimator. Because the quadratic approximation is updated iteratively in the algorithm, the difference between them are expected to be mostly negligible, which has been corroborated by our empirical studies.

Remark 2.3. (Convergence of cycBAR)—The cycBAR algorithm resembles the well-known cyclic coordinate decent (CCD) algorithm that has been commonly used for some popular single-step penalized regression methods such as LASSO. However, its numerical convergence is guaranteed by a different mechanism since the cycBAR algorithm makes coordinate-wise updates for a fixed-point problem whereas CCD aims to decrease an objective function with each coordinate update. Some graphical illustrations of the convergence of the cycBAR algorithm for $p_n = 2$ are given in Section S2.2 Figures S1 and S2 of the Online Supplementary Material. A rigorous proof of the numerical convergence of the cycBAR algorithm is however not trivial and needs to be investigated in future research.

2.4 Scalable parameter estimation via forward-backward scan

Before proceeding further, we note that for the Cox proportional hazards model with no competing risks, $R_i = \{y: X_y \leq X_i\}$ and $\tilde{w}_{ik} \equiv 1$ for all i and k . Therefore the score function can be written as

$$\dot{l}_j(\beta) = \sum_{i=1}^n I(\delta_i = 1) z_{ij} - \sum_{i=1}^n I(\delta_i = 1) \frac{\sum_{k \in R_i} z_{kj} \exp(\eta_k)}{\sum_{k \in R_i} \exp(\eta_k)}, \quad (9)$$

$j = 1, \dots, p_n$. Again, if done directly, calculating $\dot{l}_j(\beta)$ will require $\mathcal{O}(n^2)$ calculations.

Suchard et al. (2013), Mittal et al. (2014), Kawaguchi et al. (2020), among others, have implemented the following technique to calculate (9) in $\mathcal{O}(n)$ calculations. Assume, for now, that event times are unique. If the event times are arranged in decreasing order, both $\sum_{k \in R_i} z_{kj} \exp(\eta_k)$ and $\sum_{k \in R_i} \exp(\eta_k)$ are a series of cumulative sums. For example, let X_j and $X_{j'}$ be two consecutive event times such that $X_j > X_{j'}$. Then, the set $R_{j'}$ consists of the observations from R_j and the set of observations $\{y: X_y \in [X_{j'}, X_j]\}$. Therefore $\sum_{k \in R_{j'}} z_{kj} \exp(\eta_k) = \sum_{k \in R_j} z_{kj} \exp(\eta_k) + \sum_{k \in \{y: X_y \in [X_{j'}, X_j]\}} z_{kj} \exp(\eta_k)$ and calculating

both $\sum_{k \in R_i} z_{kj} \exp(\eta_k)$ and $\sum_{k \in R_i} \exp(\eta_k)$, and consequently its ratio, for all $i=1, \dots, n$ will only require $O(n)$ calculations in total. Furthermore, the outer summation of subjects who observe the event of interest is also a cumulative sum since, provided that $X_i > X_{i'}$ and both $\delta_j = 1$ and $\delta_{j'} = 1$,

$$\begin{aligned} & \sum_{l=1}^i I(\delta_l = 1) \frac{\sum_{k \in R_l} z_{kj} \exp(\eta_k)}{\sum_{k \in R_l} \exp(\eta_k)} \\ &= \sum_{l=1}^{i'} I(\delta_l = 1) \frac{\sum_{k \in R_l} z_{kj} \exp(\eta_k)}{\sum_{k \in R_l} \exp(\eta_k)} + \sum_{m \in (i', i]} I(\delta_m = 1) \frac{\sum_{k \in R_m} z_{kj} \exp(\eta_k)}{\sum_{k \in R_m} \exp(\eta_k)} \quad (10) \\ &= \sum_{l=1}^{i'} I(\delta_l = 1) \frac{\sum_{k \in R_l} z_{kj} \exp(\eta_k)}{\sum_{k \in R_l} \exp(\eta_k)} + I(\delta_i = 1) \frac{\sum_{k \in R_i} z_{kj} \exp(\eta_k)}{\sum_{k \in R_i} \exp(\eta_k)}, \end{aligned}$$

where the last equality holds since X_i and $X_{i'}$ are consecutive event times and $\delta_m = 0$ for all $m \in (i', i)$. Clearly, (10) will only require $O(n)$ calculations since the ratio can be precomputed in $O(n)$ calculations. The diagonal elements of the Hessian also follow a similar derivation and can also be calculated in $O(n)$ calculations.

For the PSH model, however, $\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_j)$, $i = 1, \dots, n$, are not a series of simple cumulative sums because 1) the risk sets R_i are not monotone over time, and 2) for each i , a different set of weights $\tilde{w}_{ik} = \hat{G}(X_i) / \hat{G}(X_i \wedge X_k)$, $k \in R_i$ are required. To overcome this problem, we show in Lemma 1 below that $\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_j)$ can be decomposed into a forward cumulative sum and a backward cumulative sum over two disjoint monotone sets. A simple proof is provided in Section S1.6 of the Online Supplementary Material.

Lemma 1.—*Assume that no ties are present. Then, for any $1 \leq r \leq p$, $1 \leq s \leq p$, and $u, v = 0, 1$, we have*

$$\sum_{k \in R_i} z_{kr}^u z_{ks}^v \tilde{w}_{ik} \exp(\eta_k) = \sum_{k \in R_i(1)} z_{kr}^u z_{ks}^v \exp(\eta_k) + \hat{G}(X_i) \sum_{k \in R_i(2)} z_{kr}^u z_{ks}^v \exp(\eta_k) / \hat{G}(X_k), \quad (11)$$

where $R_i(1) = \{y : (X_y = X_i)\}$ and $R_i(2) = \{y : (X_y < X_i \wedge \epsilon_y = 2)\}$ are distinct partitions of R_i . Furthermore, $R_i(1)$ is monotonically decreasing over time and $R_i(2)$ is monotonically increasing over time.

Because $R_i(1)$ grows cumulatively as the event times decrease from the largest to the smallest, whereas $R_i(2)$ grows cumulatively as the observed event times increase from the smallest to the largest since it only involves subjects who observed a competing risk and had an observed event time smaller than subject i . Thus, similar to the Cox model, the ratio of summations for the score and diagonal Hessian values can be calculated in linear time via a forward-backward scan where one scan goes in one direction to calculate the cumulative sums associated with $R_i(1)$ and the other scan goes in the opposite direction to calculate the cumulative sum associated with $R_i(2)$. Therefore, we can effectively reduce the number of operations from $O(n^2)$ to $O(n)$.

Remark 2.4. (Fitting the proportional cause-specific hazards model)—It is well known that for a given cause, fitting the proportional cause-specific hazards (PCSH) model is identical to fitting the standard Cox proportional hazards model for right censored data by treating the competing events as right censored. We further observe that the first term of (11) in Lemma 1 corresponds to the unweighted sum involved in fitting the standard Cox proportional hazards model for right censored data and that the second term of (11) in Lemma 1 will disappear for right censored data (in the absence of competing risks). Therefore, as a by-product, our developed algorithms for the PSH model can be directly applied to fit a PCSH model by treating the competing events as right censored.

3 Simulation study

3.1 Simulation setup

We simulate datasets under various sample sizes and parameter dimensions. The design matrix, \mathbf{Z} was generated from a p_n -dimensional standard normal distribution with mean zero and pairwise correlation $\text{corr}(z_i, z_j) = \rho^{|i-j|}$, where $\rho = 0.5$ simulates moderate correlation. The vector of regression parameters for cause 1, the cause of interest, is $\boldsymbol{\beta}_1 = (0.40, 0.45, 0, 0.50, 0, 0.60, 0.75, 0, 0, 0.80, \mathbf{0}_{p-10})$. The data generation scheme follows a similar design to that of Fine and Gray (1999) and Fu et al. (2017). The CIF for cause 1 is $F_1(t; \mathbf{z}_i) = \Pr(T_i \leq t, \epsilon_i = 1 | \mathbf{z}_i) = 1 - [1 - \pi \{1 - \exp(-t)\}]^{\exp(\mathbf{z}_i' \boldsymbol{\beta}_1)}$, which is a unit exponential mixture with mass $1 - \pi$ at ∞ when $\mathbf{z}_i = \mathbf{0}$. Unless otherwise noted, the value of π is set to 0.5, which corresponds to a cause 1 event rate of approximately 41%. The CIF for cause 2 is obtained by setting $\Pr(\epsilon_j = 2 | \mathbf{z}_j) = 1 - \Pr(\epsilon_j = 1 | \mathbf{z}_j)$ and then using an exponential distribution with rate $\exp(\mathbf{z}_j' \boldsymbol{\beta}_2)$ for the conditional CIF $\Pr(T_j \leq t | \epsilon_j = 2, \mathbf{z}_j)$ with $\boldsymbol{\beta}_2 = -\boldsymbol{\beta}_1$. Censoring times are independently generated from a uniform distribution $U(0, u_{max})$ where u_{max} controls the censoring percentage. The average censoring percentage for our simulations vary between 30 – 35%.

3.2 Finite-sample properties of BAR

In this section, we briefly summarize the results for comparing the operating characteristics of BAR along with LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), adaptive LASSO (Zou, 2006, ALASSO), and MCP (Zhang, 2010) which are implemented in the **crpp** package (Fu et al., 2017). Our simulations illustrate that 1) the BAR estimator is insensitive over the choice of ξ_n over a large interval and 2) BAR performs as well as other oracle-based procedures in terms of estimation and variable selection. This has been observed consistently over several different combinations of model dimension, event rates, signal values, sample sizes, and model sparsity. Due to the page limitation, we refer readers to Section S3 of the Online Supplementary Material for a more detailed explanation of the conclusions from the study.

3.3 Computational savings via cycBAR and forward-backward scan

In this simulation we illustrate the impressive computational savings obtained from cycBAR and the forward-backward scan described in Sections 2.3 and 2.4. We compare three implementations of BAR for the PSH model: the original BAR without the forward-

backward scan, cycBAR without the forward-backward scan, and cycBAR with the forward-backward scan. We let n vary from 600 to 2000, $p_n = 100$, and $\rho = 0.5$ and compute the runtime of each method averaged over 100 simulations. We report the runtime on a system with an Intel Xeon processor at 2.60 GHz and 64GB of memory.

Figure 1(a) displays box plots of runtime (in seconds) for each method as the sample size increases, which shows that the runtime of the original BAR (right box plot of each triple) increases quickly while the runtime of BAR implementing both cycBAR and forward-backward scan (left box plot of each triple) grows at a much slower rate. Figure S6 in the Online Supplementary Material is a magnified version of Figure 1(a) to focus on the differences between both cycBAR implementations. Panels (b) and (c) further demonstrate the separate contributions of cycBAR and the forward-backward scan method, respectively, using fold change, defined as the ratio of runtime between implementations. Panel (b) shows a 15–20 fold decrease in runtime between cycBAR and the original BAR. Panel (c) shows the benefit of linearized estimation, with a 50–225 fold decrease in runtime between cycBAR with and without the forward-backward scan. Additionally, we perform both SCAD and MCP penalizations both with and without the forward-backward scan implementation and observe similar gains in computation efficiency and the results are presented in Table S6 and Figure S7 of the Online Supplementary Material. Panel (d) illustrates that using both cycBAR and the forward-backward scan results in a multiplicative gain, yielding an impressive 1,000–2,000 fold speedup in runtime. Our simulation studies in Figures 1 and S7 strongly suggest that parameter estimation without linearization is computationally infeasible for even moderately large n . As a comparison, Figure S8 of the Online Supplementary Material illustrates that for even much larger sample sizes ($n = 10,000$ to $500,000$), cycBAR, SCAD, and MCP using the forward-backward scan can be performed within minutes.

4 End-stage renal disease

The United States Renal Data System (USRDS) is a national data system that collects information about end-stage renal disease in the United States. Patients with end-stage renal disease are known to have a shorter life expectancy compared to their disease-free peers (USRDS Annual Report 2017) and kidney transplantation has been shown to provide better health outcomes for patients with end-stage renal disease (Wolfe et al., 1999; Purnell et al., 2016). As an illustration of the scalability of various methods for large data, we run penalized regressions for a PSH model with 63 demographic and clinical variables using a subset of $n = 225,000$ patients from the USRDS that spans a 10-year study time between January 2005 to June 2015. The event of interest was first kidney transplant for patients who were currently on dialysis. Death, renal function recovery, and discontinuation of dialysis are competing risks. Subjects who are lost to follow up or had no event by the end of study period are considered as right censored. We randomly split the data into a training set ($n = 125,000$) and test set ($n = 100,000$). Table S7 in the Online Supplementary Material shows that the proportions of each type of event are similar across the training and test sets.

The BAR method along with SCAD and MCP penalizations are used to fit the PSH model using the training set. As with Section 3.3, we consider four implementations of BAR: 1)

without both `cycBAR` and the forward-backward scan; 2) without `cycBAR` and with the forward-backward scan; 3) with `cycBAR` and without the forward-backward scan; and 4) with both `cycBAR` and the forward-backward scan. BIC score minimization, implemented with a 25-value grid search, is used to find the optimal value for the tuning parameter for all three methods. We fix $\xi_n = \log(p_n)$ for the BAR method. SCAD and MCP were performed using the `crpp` @R package (Fu et al., 2017) where its generalized cross validation estimation component is removed to allow a fair comparison of their runtime to BAR with respect to parameter estimation. Additionally, we run SCAD and MCP penalizations using our forward-backward scan to compare the computational performance of our new implementation to the current state of the art. The BIC score based on the training data is used to compare selection performance between models and predictive performance is measured by the concordance index (c-index) proposed by Wolbers et al. (2009) based on the test data. Table 1 summarizes the computational time (in seconds), the BIC score, the c-index, and the number of selected variables for each method.

We observe from Table 1 that `cycBAR`, without the forward-backward scan, took 46 hours to finish, a marked reduction in runtime over the original BAR implementation which did not finish after 96 hours and was terminated. More impressively, adding the forward-backward scan resulted in an enormous boost in speeding up the computation, performing the same task in 40 seconds. We observe similar trends in both SCAD and MCP implementations as well. Our forward-backward scan algorithm results in significant reduction in runtime, over-thousand fold, for BAR, SCAD, and MCP, allowing us to perform variable selection for large-scale competing risks data within seconds rather than days. Moreover, since the `cycBAR` algorithm resembles cyclic coordinate descent (see Remark 2.3) it is computationally on par with SCAD and MCP.

The predictive and selection performances of all methods are comparable with similar BIC scores, c-index values and model sizes (number of selected variables), that we attribute to the massive sample size of both the training and test set. As expected, BAR (and `cycBAR`) selects a sparser model compared to both SCAD and MCP. This is due to BAR being an ℓ_1 -based approach as opposed to an ℓ_2 -based approach. The variables selected by BAR are also a subset of the variables selected by both SCAD and MCP. The magnitude and sign for the variables selected are consistent between methods and with some previous findings in the literature. For example, smoking has a negative effect on the subdistribution hazard revealed by all four proposed methods (BAR: -0.59 , `cycBAR`: -0.60 , SCAD: -0.61 , MCP: -0.62) for kidney transplantation and is consistent with the results of Stack et al. (2016). Other variables such as racial differences Kasiske et al., 1991; Purnell et al., 2016, 2018, insurance type Keith et al., 2008; Schold et al., 2011, and neighborhood poverty (Patzner et al., 2009) have also been previously reported to have an impact on kidney transplantation.

We also fitted penalized proportional cause-specific hazards (CSH) models using our methods as discussed in Remark 2.5. BAR, SCAD, and MCP-penalized proportional CSH regression models selected 41, 46, and 47 nonzero variables, respectively. While the CSH and PSH models estimate covariate effects on two distinct quantities of interest for competing risks data, all three penalized proportional CSH regression models yield similar

inferential conclusions to their PSH counterparts in terms of variables selected, effect sizes, and sign.

5 Discussion

In extending the surrogate ℓ_1 -based BAR methodology to the Fine and Gray (1999) PSH model for competing risks data, we have developed a novel coordinate-wise update (cycBAR) algorithm to avoid carrying out multiple ridge regressions in the original BAR implementation. Furthermore, we introduce a forward-backward scan algorithm to reduce the computational cost of the log-pseudo likelihood and its derivatives for the PSH model from the order of $O(n^2)$ to $O(n)$. While showing comparable selection and estimation performance, the BAR method for the PSH model using the two new algorithms can produce greater than 1,000 fold speedups over some current penalization methods for the PSH model in numerical studies.

While our methodology enables scalable penalized PSH regression, data storage continues to be a challenge for high-dimensional and massive sample size (HDMSS) data in our modern era. To this end, it is helpful to distinguish between HDMSS data with sparsely represented covariates and those with densely represented covariates. Sparse HDMSS data arises when only a small portion of covariates are nonzero for a given subject. This is often the case for massive electronic health record (EHR) databases such as the Observational Health Data Sciences and Informatics (OHDSI) program (Hripcsak et al., 2015) (<https://ohdsi.org/>) and the U.S. FDA's Sentinel Initiative (<https://www.fda.gov/safety/fdas-sentinel-initiative>). In this domain of applications, an effective strategy is to store the data in a sparse format by exploiting the sparsity in the data matrix. This approach has been implemented for generalized linear models Genkin et al., 2007; Friedman et al., 2010; Suchard et al., 2013 and for the standard Cox model (Mittal et al., 2014). More recently, Kawaguchi et al. (2020) has implemented the standard BAR algorithm for sparse HDMSS right-censored data. We are currently working on implementing our developed algorithms using the sparse format for massive PSH model with sparse HDMSS competing risks data, which will enable one to efficiently fit a massive PSH model using our forward-backward scan method. However, when covariates are densely represented, loading the entire HDMSS data may often be infeasible since it will exceed a computer's storage limit. In such a scenario, our fast algorithms can be coupled with distributed computing/learning methods such as divide-and-conquer (Wang et al., 2019) to improve the scalability of existing algorithms for massive size competing risks data. Distributed computing/learning methods for the PSH model remain open and warrants further investigations.

Currently, our forward-backward scan algorithm require covariates to be fixed. When covariate values are time varying, we can no longer accumulate the risk set contributions using a simple forward (or backward) scan. Efficiently estimating time-dependent covariate effects in linear time remains an open area of research for both right-censored and competing risks data.

Finally, we emphasize that the developed cycBAR method in Section 2.3 and the forward-backward scan method of Lemma 1 in Section 2.4 are of interest on their own. The cycBAR

method can be applied directly to other models and data settings. It is also straightforward to apply the forward-backward scan method to accelerate other estimation methods for the PSH model. Using this approach, we are currently developing a stand-alone package for R that includes the unpenalized estimation method of Fine and Gray (1999) and other popular penalization methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank the editor, the associate editor, and the three anonymous reviewers for their helpful comments that improved the presentation of the article. The manuscript was reviewed and approved for publication by an officer of the National Institute of Diabetes and Digestive and Kidney Diseases. Data reported herein were supplied by the USRDS. Interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as official policy or interpretation of the US government. Jenny I. Shen's work is partly supported through the National Institutes of Health grant K23DK103972. Marc A. Suchard's work is partially supported through the National Institute of Health grant U19AI135995. The research of Gang Li was partly supported by National Institute of Health Grants P30CA-16042, UL1TR000124-02, and P50CA211015.

References

- Ahn KW, Banerjee A, Sahr N, and Kim S (2018), "Group and within-group variable selection for competing risks data," *Lifetime Data Analysis*, 24, 407–424. [PubMed: 28779228]
- Breiman L (1996), "Heuristics of instability and stabilization in model selection," *Ann Stat*, 24, 2350–2383.
- Candes EJ, Wakin MB, and Boyd SP (2008), "Enhancing sparsity by reweighted L1 minimization," *Journal of Fourier Analysis and Applications*, 14, 877–905.
- Chartrand R and Yin W (2008), "Iterative Reweighted Algorithms for Compressive Sensing," in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing.
- Cox DR (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34, 187–220.
- Dai L, Chen K, Sun Z, Liu Z, and Li G (2018), "Broken adaptive ridge regression and its asymptotic properties," *Journal of Multivariate Analysis*, 168, 334–351. [PubMed: 30911202]
- Daubechies I, DeVore R, Fornasier M, and Güntürk CS (2010), "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, 63, 1–38.
- Fan J and Li R (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.
- Fine JP and Gray RJ (1999), "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, 94, 496–509.
- Friedman J, Hastie T, and Tibshirani R (2010), "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, 33, 1. [PubMed: 20808728]
- Fu Z, Parikh CR, and Zhou B (2017), "Penalized variable selection in competing risks regression," *Lifetime Data Analysis*, 23, 353–376. [PubMed: 27016934]
- Gasso G, Rakotomamonjy A, and Canu S (2009), "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Transactions on Signal Processing*, 57, 4686–4698.
- Genkin A, Lewis DD, and Madigan D (2007), "Large-scale Bayesian logistic regression for text categorization," *Technometrics*, 49, 291–304.
- Gorodnitsky IF and Rao BD (1997), "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, 45, 600–616.

- Gray RJ (1988), “A class of K-sample tests for comparing the cumulative incidence of a competing risk,” *The Annals of Statistics*, 16, 1141–1154.
- Ha ID, Lee M, Oh S, Jeong J-H, Sylvester R, and Lee Y (2014), “Variable selection in subdistribution hazard frailty models with competing risks data,” *Statistics in Medicine*, 33, 4590–4604. [PubMed: 25042872]
- Hou J, Paravati A, Hou J, Xu R, and Murphy J (2018), “High-dimensional variable selection and prediction under competing risks with application to SEER-Medicare linked data,” *Statistics in Medicine*, 37, 3486–3502. [PubMed: 29845637]
- Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, et al. (2015), “Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers,” *Studies in Health Technology and Informatics*, 216, 574–578. [PubMed: 26262116]
- Kaplan EL and Meier P (1958), “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, 53, 457–481.
- Kasiske BL, Neylan JF III, Riggio RR, Danovitch GM, Kahana L, Alexander SR, and White MG (1991), “The effect of race on access and outcome in transplantation,” *New England Journal of Medicine*, 324, 302–307. [PubMed: 1898431]
- Kawaguchi ES, Suchard MA, Liu Z, and Li G (2020), “A surrogate ℓ_0 sparse Cox’s regression with applications to sparse high-dimensional massive sample size time-to-event data,” *Statistics in Medicine*, 39, 675–686. [PubMed: 31814146]
- Keith D, Ashby VB, Port FK, and Leichtman AB (2008), “Insurance type and minority status associated with large disparities in prelisting dialysis among candidates for kidney transplantation,” *Clinical Journal of the American Society of Nephrology*, 3, 463–470. [PubMed: 18199847]
- Mittal S, Madigan D, Burd RS, and Suchard MA (2014), “High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis,” *Biostatistics*, 15, 207–221. [PubMed: 24096388]
- Patzner RE, Amaral S, Wasse H, Volkova N, Kleinbaum D, and McClellan WM (2009), “Neighborhood poverty and racial disparities in kidney transplant waitlisting,” *Journal of the American Society of Nephrology*, 20, 1333–1340. [PubMed: 19339381]
- Pintilie M (2006), *Competing Risks: A Practical Perspective*, John Wiley & Sons.
- Prentice R, Kalbfleisch J, Peterson A Jr, Flournoy N, Farewell V, and Breslow N (1978), “The analysis of failure times in the presence of competing risks,” *Biometrics*, 34, 541–554. [PubMed: 373811]
- Purnell TS, Luo X, Cooper LA, Massie AB, Kucirka LM, Henderson ML, Gordon EJ, Crews DC, Boulware LE, and Segev DL (2018), “Association of race and ethnicity with live donor kidney transplantation in the United States from 1995 to 2014,” *Jama*, 319, 49–61. [PubMed: 29297077]
- Purnell TS, Luo X, Kucirka LM, Cooper LA, Crews DC, Massie AB, Boulware LE, and Segev DL (2016), “Reduced racial disparity in kidney transplant outcomes in the United States from 1990 to 2012,” *Journal of the American Society of Nephrology*, 27, 2511–2518. [PubMed: 26848153]
- Putter H, Fiocco M, and Geskus R (2007), “Tutorial in biostatistics: competing risks and multi-state models,” *Statistics in Medicine*, 26, 2389–2430. [PubMed: 17031868]
- Schold JD, Gregg JA, Harman JS, Hall AG, Patton PR, and Meier-Kriesche H-U (2011), “Barriers to Evaluation and Wait Listing for Kidney Transplantation,” *Clinical Journal of the American Society of Nephrology*, 6, 1760–1767. [PubMed: 21597030]
- Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, and Suchard MA (2018), “Improving reproducibility by using high-throughput observational studies with empirical calibration,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 20170356.
- Shen X, Pan W, and Zhu Y (2012), “Likelihood-based selection and sharp parameter estimation,” *Journal of the American Statistical Association*, 107, 223–232. [PubMed: 22736876]
- Stack AG, Yermak D, Roche DG, Ferguson JP, Elsayed M, Mohammed W, Casserly LF, Walsh SR, and Cronin CJ (2016), “Differential impact of smoking on mortality and kidney transplantation among adult Men and Women undergoing dialysis,” *BMC Nephrology*, 17, 1–12. [PubMed: 26727891]

- Suchard MA, Simpson SE, Zorych I, Ryan P, and Madigan D (2013), “Massive parallelization of serial inference algorithms for a complex generalized linear model,” *ACM Transactions on Modeling and Computer Simulation*, 23.
- Tibshirani R (1996), “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58, 267–288.
- Wang Y, Hong C, Palmer N, Di Q, Schwartz J, Kohane I, and Cai T (2019), “A fast divide-and-conquer sparse Cox regression,” *Biostatistics*.
- Wipf D and Nagarajan S (2010), “Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions,” *IEEE Journal of Selected Topics in Signal Processing*, 4, 317–329.
- Wolbers M, Koller MT, Witteman JC, and Steyerberg EW (2009), “Prognostic models with competing risks: methods and application to coronary risk prediction,” *Epidemiology*, 20, 555–561. [PubMed: 19367167]
- Wolfe RA, Ashby VB, Milford EL, Ojo AO, Ettenger RE, Agodoa LY, Held PJ, and Port FK (1999), “Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant,” *New England Journal of Medicine*, 341, 1725–1730. [PubMed: 10580071]
- Zhang C-H (2010), “Nearly unbiased variable selection under minimax concave penalty,” *The Annals of Statistics*, 38, 894–942.
- Zhao H, Sun D, Li G, and Sun J (2018), “Variable selection for recurrent event data with broken adaptive ridge regression,” *Canadian Journal of Statistics*, 46, 416–428.
- (2019a), “Simultaneous estimation and variable selection for incomplete event history studies,” *Journal of Multivariate Analysis*, 171, 359–361.
- Zhao H, Wu Q, Li G, and Sun J (2019b), “Simultaneous Estimation and Variable Selection for Interval-Censored Data With Broken Adaptive Ridge Regression,” *Journal of the American Statistical Association*, 115, 204–216, URL 10.1080/01621459.2018.1537922. [PubMed: 32742044]
- Zou H (2006), “The adaptive lasso and its oracle properties,” *Journal of the American Statistical Association*, 101, 1418–1429.

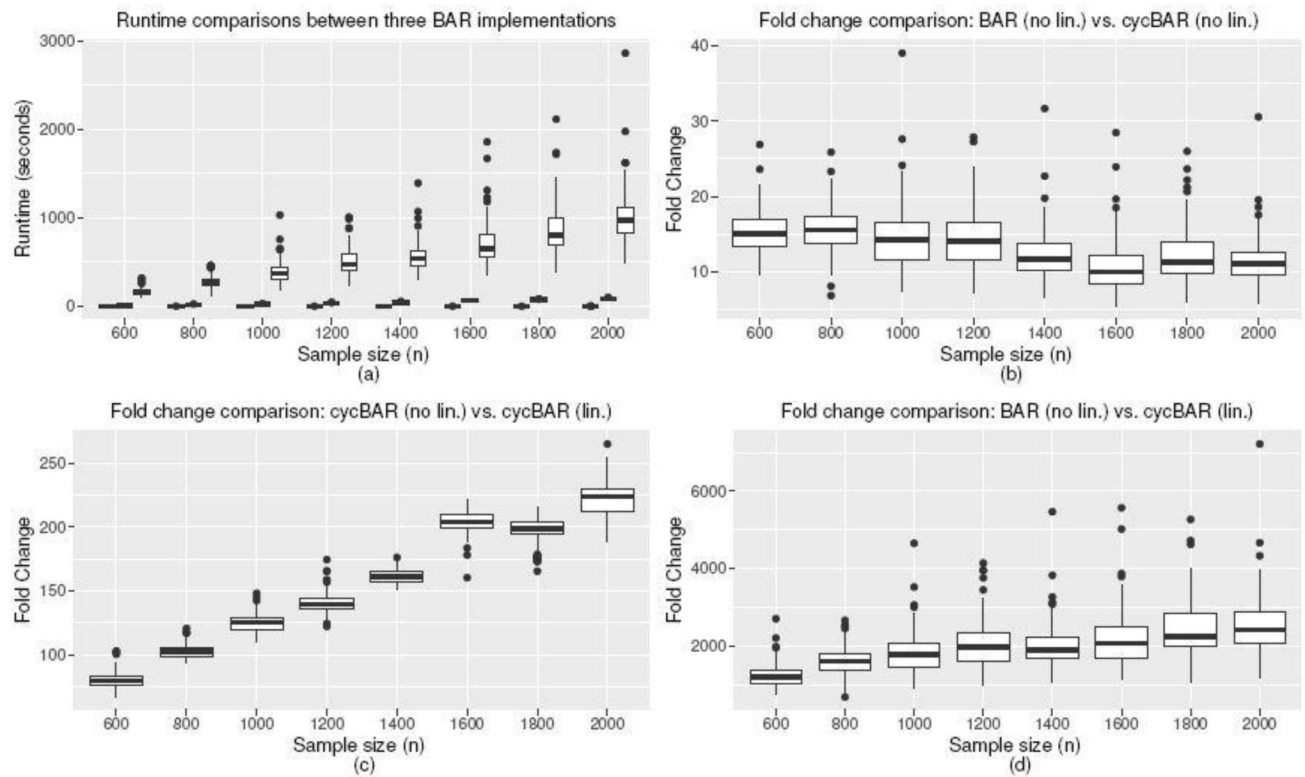


Fig. 1. Runtime comparison between three BAR (λ_n) implementations (cycBAR = cycBAR described in Section 2.3; lin. = forward-backward scan described in Section 2.4). For each triple of box plots in Figure 1(a): Left - cycBAR w/forward-backward scan; Middle - cycBAR w/o forward-backward scan; Right - BAR w/o forward-backward scan. Fold change is calculated as the ratio of runtime between two implementations.

Table 1

Analysis results of a USRDS data using BAR and CYCBAR along with MCP and SCAD. (BAR/CYCBAR: $\xi_n = \log(p_n)$ and λ_n selected through a grid search; BIC was used to select tuning parameters for all methods; Seconds: Runtime in seconds without the forward-backward scan (no scan) and with (scan); BIC score: BIC score based on the training data; c -index: c -index based on the test data; Model size: Number of nonzero parameters

	BAR	CYCBAR	SCAD	MCP
Seconds (no scan)	345,600+ *	167,020	92,571	102,565
Seconds (scan)	1,401	40	37	35
BIC score	251873.7	251867.6	251929.9	251895.3
c -index	0.85	0.85	0.85	0.85
Model size	43	42	48	49

* The original BAR without CYCBAR and forward-backward scan did not finish after 96 hours.)