Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2016, Vol. 25(6) 2685–2703

DOI: 10.1177/0962280214530608
smm.sagepub.com
SAGE

Penalized count data regression with application to hospital stay after pediatric cardiac surgery

Zhu Wang,[1] Shuangge Ma,[2] Michael Zappitelli,[3] Chirag Parikh,[4] Ching-Yun Wang[5] and Prasad Devarajan[6]

Abstract

Pediatric cardiac surgery may lead to poor outcomes such as acute kidney injury (AKI) and prolonged hospital length of stay (LOS). Plasma and urine biomarkers may help with early identification and prediction of these adverse clinical outcomes. In a recent multi-center study, 311 children undergoing cardiac surgery were enrolled to evaluate multiple biomarkers for diagnosis and prognosis of AKI and other clinical outcomes. LOS is often analyzed as count data, thus Poisson regression and negative binomial (NB) regression are common choices for developing predictive models. With many correlated prognostic factors and biomarkers, variable selection is an important step. The present paper proposes new variable selection methods for Poisson and NB regression. We evaluated regularized regression through penalized likelihood function. We first extend the elastic net (Enet) Poisson to two penalized Poisson regression: Mnet, a combination of minimax concave and ridge penalties; and Snet, a combination of smoothly clipped absolute deviation (SCAD) and ridge penalties. Furthermore, we extend the above methods to the penalized NB regression. For the Enet, Mnet, and Snet penalties (EMSnet), we develop a unified algorithm to estimate the parameters and conduct variable selection simultaneously. Simulation studies show that the proposed methods have advantages with highly correlated predictors, against some of the competing methods. Applying the proposed methods to the aforementioned data, it is discovered that early postoperative urine biomarkers including NGAL, IL18, and KIM-1 independently predict LOS, after adjusting for risk and biomarker variables.

Keywords

Poisson regression, negative binomial regression, variable selection, Enet, Mnet, Snet

[1]Department of Research, Connecticut Children's Medical Center, Hartford, CT, USA
[2]Department of Biostatistics, Yale University, New Haven, CT, USA
[3]Division of Nephrology, Department of Pediatrics, Montreal Children's Hospital, McGill University Health Centre, Montreal, Quebec, Canada
[4]Section of Nephrology and Program of Applied Translational Research, Yale University School of Medicine, New Haven, CT, USA
[5]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[6]Department of Nephrology and Hypertension, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Corresponding author:
Zhu Wang, Department of Research, Connecticut Children's Medical Center, Hartford, CT, USA.
Email: zwang@connecticutchildrens.org

1 Introduction

Despite the decrease of peri-operative mortality rate, pediatric cardiac surgery for congenital cardiac anomalies can generate complications which may contribute to patient morbidity and resource utilization. Acute kidney injury (AKI) is a frequent complication of pediatric cardiac surgery and is associated with prolonged hospital length of stay.[1] There is active research to evaluate the role of new biomarkers to facilitate with prevention or treatment of AKI. The Transnational Research Investigating Biomarker Endpoints in AKI (TRIBE-AKI) study is a multi-center prospectively study, which enrolled 311 children undergoing cardiac surgery. Half of the pediatric patients were younger than two years old and 91% surgeries were elective. The average hospital LOS was 8.5 (SD, 10.8) days. The study was designed to evaluate the role of early postoperative biomarkers in plasma and urine to predict AKI and other adverse patient outcomes such as prolonged LOS. Thus, in addition to the clinical risk factors, the TRIBE-AKI study collected preoperative and early postoperative biomarkers of AKI. The 0–6 h postoperative biomarkers were collected at a median of 0.5 h after arrival in the intensive care unit. The 0–6 and 6–12 h postoperative biomarkers included serum creatinine (SER_CRE), serum Cystatin C (SER_CYSC), urine IL18 (UR_IL18), urine neutrophil gelatinase-associated lipocalin (UR_NGAL), urine creatinine (UR_CRE), urine kidney injury molecule 1 (UR_KIM-1), urine liver-type fatty acid binding protein (LFABP), urine Cystatin C (UR_CYSC), urine albumin to creatinine ratio (UR_AC), and urine microalbumin (UR_MALB). The data also included preoperative estimated glomerular filtration rate or renal function (eGFR). In this paper, we develop new statistical techniques to study the association between multiple biomarkers and LOS. Patients with prolonged LOS after cardiac surgery have high mortality risk, and contribute to increased resource utilization. Accurate prediction of LOS before and after surgery can help determine indications for surgery, provide patient counseling, allocate resources, and monitor quality of surgeons and institutions.

LOS often shows a highly skewed distribution as some of the patients with peri-operative complications have longer LOS. Data from the TRIBE-AKI study also demonstrate that LOS is highly skewed. Thus, a linear regression model may not adequately fit this skewed data. An alternative approach is the generalized linear model including Poisson regression and negative binomial (NB) regression.[2] Another feature of the TRIBE-AKI data is that predictor variables are often correlated. Some correlations are as high as 0.93, stemming from different postoperative biomarkers and different time points for the same biomarker. In addition, age and weight are highly correlated for young children. With highly correlated predictors as seen in the TRIBE-AKI data, the traditional Poisson regression and NB regression can be challenged. In real data analysis, it is often required to select a subset of predictors from a large pool of variables. A parsimonious model has better interpretation and often leads to more accurate prediction. For the TRIBE-AKI data, an optimal goal is to accurately predict an outcome such as LOS with meaningful biomarkers. In this paper, we aim to provide new methodology to identify a subset of biomarkers which can more accurately predict LOS. These biomarkers increase the specificity of capturing AKI diagnosis, and the use of AKI biomarkers is at the forefront of research evaluating the effect of AKI on outcomes. Despite recent progress on variable selection, there is a limited option for highly correlated variables with Poisson regression. Most recent variable selection methods are embedded in the estimation process by optimizing penalized objective functions. The penalized methods include the least absolute shrinkage and selection operator[3] (LASSO), the minimum concave penalty[4] (MCP) and the smoothly clipped absolute deviation[5] (SCAD). The LASSO tends to select only one predictor among a group of highly correlated predictors. The elastic net (Enet) is a mixture of LASSO and ridge penalty to avoid the aforementioned drawback of the LASSO and can improve the variable selection.[6]

Following the same idea, the mixture of MCP and ridge penalty (Mnet) has been developed for linear regression.[7] With count data, some variable selection methods have been proposed. The Enet has been incorporated in the generalized linear model (GLM) including Poisson regression.[8] Fan and Lv developed the methodologies of the MCP and SCAD for the GLM.[9] Also, the penalized logistic regression with the MCP and SCAD[10] can be extended to Poisson regression. To our best knowledge, however, the Mnet and the mixture of SCAD and ridge penalty (Snet) have not been studied in Poisson regression. Furthermore, there is a lack of methodology for the penalized NB regression. To advance the knowledge in this area, the present paper proposes new algorithms and also applies them to the analysis of LOS with highly correlated biomarkers. The rest of the article is organized as follows. In Section 2, we describe penalized count regression for Poisson and NB models. Section 3 presents the algorithm for model parameter estimation, namely, the iteratively reweighted least squares (IRLS) and the coordinate descent algorithm. In Section 4, a simulation study is conducted to evaluate and compare the proposed algorithms and alternative methods. In Section 5, the proposed methods are applied to the TRIBE-AKI hospital length of stay. Finally, Section 6 concludes with discussions.

2 Penalized generalized linear regression

Let random variable Y follow a distribution in the exponential family with mean $\mu = E(Y)$ and variance $V = var(Y)$. For the n pairs of observations (x_{ij}, y_i), $i = 1, \ldots, n$, $j = 0, 1, \ldots, p$, each predictor variable $x_i = (x_{i0}, \ldots, x_{ip})^{\mathrm{T}}$ is length of $p+1$ and $x_{i0} = 1$, for $i = 1, \ldots, n$. Without loss of generality, we consider standardized variables $\sum_{i=1}^{n} x_{ij} = 0$, $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$, for $j = 1, \ldots, p$. The GLM contains Poisson distribution which has a probability function

$$f(y; \mu) = \frac{\mu^y}{y!} \mathrm{e}^{-\mu}$$

where $E(Y) = var(Y) = \mu$. The real data, however, often show that the variance is larger than the mean, which is called overdispersion. The two-parameter negative binomial distribution is more flexible and can model overdispersed count data

$$f(y; \theta, \mu) = \frac{\Gamma(\theta + y)}{\Gamma(\theta)\Gamma(1 + y)} \left(\frac{\mu}{\theta + \mu}\right)^y \left(\frac{\theta}{\theta + \mu}\right)^\theta \tag{1}$$

where $E(Y) = \mu$ and $var(Y) = \mu + \mu^2/\theta$. The NB distribution is a generalization of Poisson distribution and the former approaches the latter when $\theta \to \infty$. For a known dispersion parameter θ, the NB distribution belongs to the exponential family, thus the NB regression model can be conveniently estimated like other GLMs in the framework of maximum likelihood. Consider a link function g such that $\eta_i = g(\mu_i) = x_i^T \beta$, where the coefficient vector $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and β_0 is the intercept. Throughout the paper, the logarithmic link function $\eta = \log(\mu)$ is employed.

For the purpose of variable selection, we define a penalized log-likelihood function

$$p\ell(\beta; y) = \ell(\beta; y) - n \sum_{j=1}^{p} p\left(\lambda; |\beta_j|\right) \tag{2}$$

where $\ell(\beta; y)$ is the log-likelihood function of Poisson or NB distribution (θ is suppressed for ease of notation) and $p(\lambda; |\beta_j|)$ is the penalty function with tuning parameter λ which may represent a vector (λ_1, λ_2) depending on the type of penalty function. Maximizing the penalized log-likelihood function (2) can estimate parameters and shrink small coefficients to zero. We focus on three penalty functions

(i) The Enet penalty[6]

$$p(\lambda_1, \lambda_2; |\beta|) = \lambda_1|\beta| + \frac{1}{2}\lambda_2\beta^2 \tag{3}$$

for $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$.

(ii) The Mnet penalty[7]

$$\begin{aligned} p(\lambda_1, \lambda_2; |\beta|) &= p(\lambda_1; |\beta|) + \frac{1}{2}\lambda_2\beta^2, \\ p(\lambda_1; \beta)' &= (\lambda_1 - \beta/\gamma)I(\beta \leq \gamma\lambda_1) \end{aligned} \tag{4}$$

for $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\gamma > 1$.

(iii) The combination of SCAD and ridge penalty (Snet)[5]

$$\begin{aligned} p(\lambda_1, \lambda_2; |\beta|) &= p(\lambda_1; |\beta|) + \frac{1}{2}\lambda_2\beta^2, \\ p(\lambda_1, \beta)' &= \lambda_1\left\{I(\beta < \lambda_1) + \frac{(\gamma\lambda_1 - \beta)_+}{(\gamma - 1)\lambda_1}I(\beta \geq \lambda_1)\right\} \end{aligned} \tag{5}$$

for $\lambda_1 \geq 0, \lambda_2 \geq 0$ and $\gamma > 2$, where t_+ denotes the positive part of t.

The ridge penalty component $\frac{1}{2}\lambda_2\beta^2$ is designed to select a group of variables when the correlations are high. The ridge penalty may be dropped if we let $\lambda_2 = 0$, corresponding to the LASSO, MCP, and SCAD penalty, respectively. However, when a group of variables have high correlations, the LASSO, MCP, and SCAD may select only one of the variables, leading to possibly inferior prediction performance than the ridge regression.[6,7] In medical research, individual biomarker may have useful but limited power to predict clinical outcomes. For instance, the accuracy of urine IL-18 and urine NGAL for diagnosis of severe AKI is moderate.[1] To improve prediction accuracy, therefore, it is interesting to combine the biomarkers in a meaningful way. By adding the ridge penalty, the EMSnet (short for Enet, Mnet, and Snet) has the advantage for selecting highly correlated variables. This will be illustrated in the simulation study. Since the EMSnet is a combination of ridge penalty and the second penalty component (namely LASSO, MCP, or SCAD), we briefly describe the difference among the second penalty component. The LASSO is a convex penalty which has the computing advantage. The MCP and SCAD are not convex penalties but penalize large coefficients less than LASSO, which may reduce estimation bias. In some settings, the MCP and SCAD enjoy the oracle properties[4,5,9] which implies that the variables can be selected as if we had known the true effective predictors in advance.

3 Parameter estimation

To estimate the parameters for the penalized count regression, we utilize a combination of the IRLS algorithm and coordinate descent algorithm. The IRLS algorithm is commonly applied when fitting GLMs and the coordinate descent algorithm is integrated within the IRLS algorithm for variable selection.

3.1 Iteratively reweighted least squares algorithm

We first present the estimation algorithm for penalized Poisson regression and NB regression with known θ. A strategy for unknown θ will be proposed afterwards. Maximization of the penalized log-likelihood (2) can be accomplished by coupling the IRLS algorithm with a suitable variable selection algorithm[8–9]. Starting with initial values, the IRLS algorithm maximizes the penalized log-likelihood iteratively in the following steps

(1) Compute observation weights for $i = 1, \ldots, n$

$$w_i = 1/\left(V_i g_i'^2\right). \tag{6}$$

Thus, $w_i = \mu_i$ and $w_i = \frac{\mu_i}{1+\mu_i/\theta}$ for Poisson and NB regressions, respectively.

(2) Compute working responses for $i = 1, \ldots, n$

$$z_i = \eta_i + (y_i - \mu_i)\, g_i' = \eta_i + \frac{y_i - \mu_i}{\mu_i}$$

for Poisson and NB regressions.

(3) Minimize the penalized weighted least squares

$$\beta = \arg\ \min \frac{1}{2n} \sum_{i=1}^{n} w_i \left(z_i - x_i^{\mathrm{T}} \beta\right)^2 + \sum_{j=1}^{p} p\left(\lambda; |\beta_j|\right). \tag{7}$$

(4) Compute the linear predictor based on the regression estimates

$$\eta_i = x_i^{\mathrm{T}} \beta.$$

(5) Compute μ_i or $E(y_i)$

$$\mu_i = g^{-1}(\eta_i).$$

If θ has to be estimated for the NB regression, as is often the case in practice, we adopt an iterative procedure. Initially we fit a non-penalized and intercept-only NB regression model, i.e. we only estimate θ and β_0, while letting $\beta_j = 0, \quad j = 1, \ldots, p$. There are existing algorithms for non-penalized NB regression. We then optimize the penalized log-likelihood function by alternating

between estimating β (with the current estimate θ fixed) and estimating θ (with the current estimate β fixed). The alternation is repeated until certain convergence criteria are reached.

3.2 Coordinate descent algorithm

Variable selection is achieved when minimizing the penalized weighted least squares (equation (7)). Here we use the coordinate descent algorithm.[8,10,11] This algorithm can handle all the aforementioned penalties in a unified fashion. With fast speed, the algorithm can estimate coefficients along a regularized path. Iteratively, the coordinate descent algorithm minimizes (7) in the following three steps:

(1) Calculate $r_i = (y_i - \mu_i)\, g_i'$, where r_i is the current residual for observation i.
(2) For $j = 1, \ldots, p$, calculate

$$q_j = \frac{1}{n}\sum_{i=1}^{n} w_i x_{ij} r_i + \nu_j \widehat{\beta}_j$$

and $\nu_j = \frac{1}{n}\sum_{i=1}^{n} w_i x_{ij}^2$.

(3) Update $\widehat{\beta}_j$ based on the penalty function.
   (a) For the Enet penalty

$$\widehat{\beta}_j = \frac{S(q_j, \lambda_1)}{\nu_j + \lambda_2} \tag{8}$$

   (b) For the Mnet penalty

$$\widehat{\beta}_j = \begin{cases} \dfrac{S(q_j, \lambda_1)}{\nu_j + \lambda_2 - 1/\gamma}, & \text{if } |q| \le \nu_j \gamma \lambda_1 (1 + \lambda_2) \\ \dfrac{q_j}{\nu_j + \lambda_2}, & \text{otherwise} \end{cases} \tag{9}$$

for $\gamma(1 + \lambda_2) > 1 + 1/\nu_j$.

   (c) For the Snet penalty

$$\widehat{\beta}_j = \begin{cases} \dfrac{S(q_j, \lambda_1)}{\nu_j + \lambda_2}, & \text{if } |q_j| \le \lambda_1 + \lambda_1(\nu_j + \lambda_2) \\ \dfrac{S(q_j, \gamma\lambda_1/(\gamma - 1))}{\nu_j - 1/(\gamma - 1) + \lambda_2}, & \text{if } \lambda_1 + \lambda_1(\nu_j + \lambda_2) < |q_j| \le \nu_j \gamma \lambda_1 (1 + \lambda_2) \\ \dfrac{q_j}{\nu_j + \lambda_2}, & \text{if } |q_j| > \nu_j \gamma \lambda_1 (1 + \lambda_2) \end{cases} \tag{10}$$

for $\gamma(1 + \lambda_2) > 1/\nu_j$.

The soft-thresholding operator $S(q,t)$ is defined by

$$S(q,t) = \text{sign}(q)(|q| - t)_+ = \begin{cases} q - t, & \text{if } q > 0 \text{ and } t < |q| \\ q + t, & \text{if } q < 0 \text{ and } t < |q| \\ 0, & \text{if } t \geq |q| \end{cases}. \quad (11)$$

3.3 Pathwise coordinate descent

We compute the solutions $\hat{\beta}$ along a regularized path. The penalty function $p(\beta)$ has two tuning parameters (λ_1, λ_2). Additionally, the Mnet and Snet have the third tuning parameter γ. A reparameterization is often convenient for implementation: $\phi = \lambda_1 + \lambda_2$ and $\alpha = \lambda_1/\phi$.[7,8] Therefore, for a fixed α value, we can compute solutions for a decreasing sequence of ϕ. Specifically, for the candidate values of ϕ, we consider a decreasing sequence of $\phi_k, k = 1, \ldots, K$ from $\phi_{\max}$ to $\phi_{\min}$ on the log scale with $\phi_{\min} = \epsilon\phi_{\max}$. The starting value $\phi_{\max}$ can be chosen such that the entire coefficients $\hat{\beta}_j = 0, \quad j = 1, \ldots, p$.[8,10] We let $K = 100$ and $\epsilon = 0.001$.

3.4 Convexity

The objective function in (2) may not be convex if the penalty function $p(\lambda, \beta)$ is not convex, such as the Mnet or Snet. For nonconvex objective functions, the proposed estimating algorithm may not converge to the global optimal point and the estimated $\hat{\beta}$ may become discontinuous along the solution path. Under some conditions, however, results on convexity have been established for the linear regression and GLM with MCP and SCAD penalties.[4,10] The following result is a simple extension for the Mnet and Snet penalties.

Proposition 1: *Let $e(\beta)$ denote the minimum eigenvalue of $\frac{1}{n}X^{\mathrm{T}}WX$, where X is the predictor variable matrix and W is a diagonal matrix with elements defined in (6) and evaluated at current estimate β. Then the objective function (2) is a convex function of β on the region where $e(\beta) > 1/\gamma - \lambda_2$ for Mnet, and where $e(\beta) > \frac{1}{\gamma - 1} - \lambda_2$ for Snet.*

The proposition can be easily validated following the proof of proposition 2 in Breheny and Huang.[10]

4 Simulation study

The simulation study was to demonstrate that for the highly correlated predictors, the EMSnet-penalized count regression is a better choice than the corresponding penalized regression without the ridge penalty component. The proposed methods were also compared with the oracle estimator in which the true effective predictors were known in advance. For each model described below, we generated training, validation, and testing data. The training data were used for model fitting; the tuning parameters were selected based on the log-likelihood value from the validation data; the testing data were used to evaluate the prediction accuracy. For the simulated Poisson, we fit the penalized Poisson, including Enet, Mnet, Snet, LASSO, MCP, and SCAD penalties. This process was repeated for the NB regression. The common tuning parameter is ϕ in these models. For EMSnet, the tuning parameters also include α. Except for the Enet and LASSO, we also had to choose γ. For the EMSnet, we first picked a small grid of values for $\alpha \in (1, 0.8, 0.6, 0.4, 0.2)$, then, for each α, the log-likelihood values were evaluated from the validation data over a sequence of

candidate values for ϕ. The optimal tuning parameters were chosen based on the maximum log-likelihood value. Additionally, the tuning parameter γ was chosen from 2.5, 3.7, 6 which worked well in our simulation and data analysis. We computed the prediction error (PE) $E(Y - g(X^{\mathrm{T}}\hat{\beta}))^2$ from the testing data. Other performance metrics follow. Denote the mean squared error $\mathrm{MSE} = \frac{1}{p}\sum_{j=1}^{p}(\hat{\beta}_j - \beta_j)^2$. A smaller MSE indicates a better estimator. The deviance is a measure of discrepancy between the fitted model and the observations. The true positive rate (sensitivity) and the true negative rate (specificity) are used to evaluate variable selection performance

$$\text{sensitivity} = \frac{\text{number of correctly selected predictors}}{\text{number of effective predictors}} \tag{12}$$

and

$$\text{specificity} = \frac{\text{number of correctly unselected predictors}}{\text{number of ineffective predictors}} \tag{13}$$

The sensitivity and specificity are expected to reach 1 for a good estimator.

4.1 Data generation

The simulated data contained four models for Poisson regression and NB regression, respectively. We first investigate low-dimensional problems with $p = 20$.

Scenario 1: The predictor variables X were randomly drawn from multivariate normal distributions $N_{20}(\mathbf{0}, \Sigma)$, where Σ has elements $\rho^{|i-j|}(i, j = 1, \ldots, 20)$. The correlation among predictor variables was controlled by ρ with $\rho = 0.5$ and 0.8, respectively. The parameter was set as

$$\beta = \left(0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, \ldots, 0}_{5}, 0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, \ldots, 0}_{5}\right).$$

Scenario 2: Similarly to scenario 1, we generated multivariate predictor variables X but with uniform correlations $\rho = 0.5$ and 0.8, respectively, and parameter

$$\beta = \left(1.25, -0.95, 0.9, -1.1, 0.6, \underbrace{0, \ldots, 0}_{15}\right).$$

Scenario 3: We considered uniform correlations ρ as in scenario 2, and parameter

$$\beta = \left(\underbrace{0, \ldots, 0}_{5}, \underbrace{0.1, \ldots, 0.1}_{5}, \underbrace{0, \ldots, 0}_{5}, \underbrace{0.3, \ldots, 0.3}_{5}\right).$$

Scenario 4: In this setting, we considered the grouping effect. There are $p = 20$ predictors. The grouped predictors were generated as follows

$$\begin{aligned} x_i &= Z_1 + \epsilon_i, Z_1 \sim N(0,1), \quad i = 1, \ldots, 3, \\ x_i &= Z_2 + \epsilon_i, Z_2 \sim N(0,1), \quad i = 4, \ldots, 6, \\ x_i &= Z_3 + \epsilon_i, Z_3 \sim N(0,1), \quad i = 7, \ldots, 9 \end{aligned}$$

where $\epsilon_i \sim N(0, 0.01), \quad i = 1, \ldots, 9$ and $x_i \sim N(0,1), \quad i = 11, \ldots, 20$. The parameter was set as

$$\beta = \left(\underbrace{0.3 \ldots, 0.3}_{9}, \underbrace{0, \ldots, 0}_{11} \right).$$

For each scenario, we also considered high-dimensional problems, for example, $p = 100$ or $p = 200$. There are additional $p - 20$ predictor variables with no effect (i.e., $\beta_j = 0, \quad j = 21, \ldots, p$), which were generated based on the setting for $p = 20$. The parameters in the simulations were chosen to obtain simulated response Y in a range comparable to the LOS in the TRIBE-AKI data. The response vector Y was generated from Poisson or NB distribution with conditional mean vector $\exp(X^{\mathrm{T}}\beta)$. To generate the NB distribution, $\theta = 2$ in (1) was chosen. For each model, we generated observation numbers of the training/validation/testing data 100/100/400 in scenarios 1–3 and 200/200/400 in scenario 4. The proposed methods were fitted to the simulated data and performance measures were computed. We repeated 100 times.

4.2 Simulation results

4.2.1 Poisson regression

The performance of various methods for Poisson regression is summarized in Tables 1–4. In each setting, the bold font indicates the best model on MSE, PE and deviance, respectively. The investigated methods typically have better results when the predictor variables are low-dimensional than high-dimensional, with smaller prediction error (PE) and more accurate variable selection. Notice the specificity is not directly comparable when p is different. The EMSnet has more accurate or comparable results than their corresponding counterparts without the ridge penalty component. The EMSnet typically generated smaller or comparable median MSE and prediction error. For variable selection, the EMSnet usually selected more effective predictors than their counterparts, which resulted in increased sensitivity. This is most obvious with the grouped predictors in scenario 4. Recall that there are three groups and each group has three similar predictors. The MCP and SCAD, however, only select one of three effective predictors in each of the three groups. Among the EMSnet estimators, the Enet incorrectly selects more predictors than the Mnet and Snet, leading to smaller specificity. In addition, the Enet typically has a larger prediction error than the MCP and SCAD. When correlations increase from $\rho = 0.5$ to 0.8, the MSEs increase as well. No method performs universally better than others on deviance. The oracle estimator may not generate the best MSE or prediction error for the highly correlated predictors, for instance, in scenario 3 with $\rho = 0.8$ and scenario 4. This is due to the fact that the oracle estimator is not the oracle ridge estimator, thus the larger MSE or prediction error should not be surprising for the highly correlated predictors.

Table 1. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, and specificity.

Method	MSE	PE	Deviance	Sensitivity	Specificity
		$p = 20$, $\rho = 0.5$			
Enet	0.0121 (0.006)	3.009 (0.9)	83.738 (10.42)	1 (0)	0.2 (0.15)
Mnet	0.0072 (0.005)	2.598 (0.86)	86.607 (10.6)	1 (0)	0.8 (0.15)
Snet	**0.0062** (0.005)	**2.596** (0.84)	86.809 (9.99)	1 (0)	0.7 (0.15)
LASSO	0.0121 (0.007)	3.015 (0.92)	**83.349** (10.63)	1 (0)	0.2 (0.15)
MCP	0.0074 (0.005)	2.77 (0.92)	86.59 (9.32)	1 (0)	0.8 (0.15)
SCAD	0.0072 (0.005)	2.771 (0.85)	86.545 (10.44)	1 (0)	0.8 (0.15)
ORACLE	0.0052 (0.004)	2.516 (0.77)	89.427 (11.17)	1 (0)	1 (0)
		$p = 20$, $\rho = 0.8$			
Enet	0.0352 (0.017)	1.914 (0.36)	94.034 (12.11)	1 (0)	0.1 (0.15)
Mnet	**0.0319** (0.019)	**1.859** (0.35)	94.875 (12.34)	1 (0)	0.7 (0.3)
Snet	0.0323 (0.018)	1.862 (0.33)	93.856 (12.9)	1 (0)	0.5 (0.3)
LASSO	0.0357 (0.017)	1.957 (0.4)	**92.821** (12.92)	1 (0)	0.2 (0.3)
MCP	0.0407 (0.019)	2.024 (0.36)	93.668 (12.81)	1 (0)	0.7 (0.3)
SCAD	0.0405 (0.021)	2.048 (0.41)	93.595 (13.28)	1 (0)	0.6 (0.3)
ORACLE	0.0166 (0.01)	1.705 (0.31)	97.083 (12.12)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.5$			
Enet	0.0103 (0.003)	6.174 (1.93)	89.718 (51.64)	0.7 (0.3)	0.83 (0.09)
Mnet	0.0075 (0.005)	**5.645** (2.07)	78.997 (31.37)	0.7 (0.3)	0.95 (0.04)
Snet	**0.0073** (0.005)	5.658 (2.24)	**77.751** (31.13)	0.7 (0.3)	0.90 (0.05)
LASSO	0.0099 (0.003)	6.174 (1.86)	88.821 (50.7)	0.7 (0.3)	0.84 (0.08)
MCP	0.0086 (0.005)	6.064 (2.07)	93.367 (43.1)	0.6 (0.3)	0.96 (0.03)
SCAD	0.0082 (0.004)	6.141 (2.01)	90.603 (38.04)	0.7 (0.3)	0.92 (0.04)
ORACLE	6e-04 (3e–04)	2.423 (0.67)	88.011 (12.2)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.8$			
Enet	0.0124 (6e–04)	**2.39** (0.33)	121.469 (26.11)	0.35 (0.07)	0.94 (0.05)
Mnet	0.0124 (8e–04)	2.44 (0.35)	121.093 (27.7)	0.3 (0.15)	0.96 (0.05)
Snet	**0.0123** (8e–04)	2.441 (0.38)	**116.42** (21.76)	0.4 (0.15)	0.92 (0.06)
LASSO	0.0124 (7e–04)	2.393 (0.37)	124.061 (25.25)	0.3 (0.15)	0.96 (0.03)
MCP	0.0125 (0.001)	2.428 (0.38)	125.907 (27.04)	0.2 (0.15)	0.98 (0.02)
SCAD	0.0124 (9e–04)	2.413 (0.42)	123.299 (25)	0.3 (0.15)	0.96 (0.03)
ORACLE	0.0016 (0.0011)	1.674 (0.25)	95.334 (11.17)	1 (0)	1 (0)

Note: Methods in Poisson regression with scenario 1.

4.2.2 *Negative binomial regression*

The simulation results for the NB regression are displayed in Tables 5–8. In addition to the performance measures as for Poisson regression, the estimation results for θ are also summarized. Different from the above Poisson regression in scenario 4, the results are reported with $\alpha = 0.2$ fixed rather than selected from the data. In fact, the tuning parameter strategy may produce a large α value, leading to a failure of selecting grouped variables for the Mnet and Snet. This phenomenon also occurs in the penalized Poisson regression. The tuning parameter α is important; however, the likelihood-based tuning parameter strategy cannot distinguish grouped variables. In this situation, one may adopt some scientific guidance and consider α as a hyper-parameter.[8] Here the results are

Table 2. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, and specificity.

Method	MSE	PE	Deviance	Sensitivity	Specificity
		$p = 20$, $\rho = 0.5$			
Enet	0.0055 (0.003)	11.893 (7.45)	**76.814** (12.8)	1 (0)	0.5 (0.22)
Mnet	0.0013 (0.001)	5.831 (2.73)	85.39 (12.53)	1 (0)	1 (0)
Snet	0.0013 (0.001)	**5.821** (2.72)	85.39 (13.28)	1 (0)	1 (0)
LASSO	0.0054 (0.003)	11.893 (7.69)	77.29 (12.55)	1 (0)	0.5 (0.22)
MCP	0.0013 (0.001)	5.831 (2.64)	85.139 (12.5)	1 (0)	1 (0)
SCAD	**0.0012** (0.001)	**5.821** (2.72)	84.972 (12.78)	1 (0)	1 (0)
ORACLE	0.001 (0.001)	5.718 (2.44)	85.39 (11.57)	1 (0)	1 (0)
		$p = 20$, $\rho = 0.8$			
Enet	0.0186 (0.01)	3.292 (1.15)	**89.147** (12.3)	1 (0)	0.575 (0.26)
Mnet	0.0079 (0.008)	2.781 (1.12)	92.361 (9.83)	1 (0)	0.975 (0.04)
Snet	0.0067 (0.007)	**2.627** (0.96)	92.399 (11.02)	1 (0)	0.95 (0.07)
LASSO	0.0183 (0.01)	3.244 (1.13)	**89.147** (12.18)	1 (0)	0.6 (0.22)
MCP	0.0067 (0.006)	2.83 (1.14)	91.277 (9.31)	1 (0)	1 (0)
SCAD	**0.0064** (0.006)	2.678 (0.93)	91.252 (10.37)	1 (0)	0.95 (0.07)
ORACLE	0.0048 (0.003)	2.401 (0.63)	95.22 (11.08)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.5$			
Enet	0.0028 (0.0012)	38.479 (33.7)	**80.497** (15.53)	1 (0)	0.915 (0.03)
Mnet	**2e–04** (1e–04)	6.595 (3.36)	84.467 (12.65)	1 (0)	1 (0)
Snet	**2e–04** (1e–04)	6.835 (3.32)	83.262 (11.06)	1 (0)	0.995 (0.01)
LASSO	0.0028 (0.0012)	38.479 (33.7)	**80.497** (15.53)	1 (0)	0.915 (0.03)
MCP	2e–04 (1e–04)	**6.458** (3.15)	85.473 (12.01)	1 (0)	1 (0)
SCAD	2e–04 (1e–04)	6.695 (3.21)	83.45 (11.06)	1 (0)	0.995 (0.01)
ORACLE	2e–04 (1e–04)	6.343 (3.11)	88.008 (11.22)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.8$			
Enet	0.0062 (0.003)	5.083 (2.03)	**85.332** (15.09)	1 (0)	0.913 (0.04)
Mnet	0.0027 (0.003)	3.409 (1.61)	93.023 (13.24)	1 (0)	0.995 (0.01)
Snet	**0.0025** (0.003)	3.455 (1.54)	92.929 (13.1)	1 (0)	0.99 (0.02)
LASSO	0.0062 (0.003)	5.083 (2.03)	85.67 (15.28)	1 (0)	0.913 (0.04)
MCP	0.0027 (0.003)	**3.309** (1.61)	92.441 (12.53)	1 (0)	0.995 (0.01)
SCAD	0.0026 (0.003)	3.415 (1.65)	93.009 (11.77)	1 (0)	0.99 (0.01)
ORACLE	5e–04 (4e–04)	2.316 (0.57)	94.138 (12.85)	1 (0)	1 (0)

Note: Methods in Poisson regression with scenario 2.

based on fixed $\alpha = 0.2$, which presumably is not the optimal value. Nevertheless, we are able to show that the EMSnet can select grouped variables by adding the ridge penalty, while the LASSO, MCP, and SCAD cannot. The EMSnet identified all nine true positives while LASSO/MCP/SCAD only identified three true positives. As a result, the EMSnet produced much smaller MSE than their counterparts.

The simulation results lead to similar conclusions as for Poisson regression. In particular, the EMSnet methods provide more accurate parameter estimation and prediction error, and select more predictors. In scenario 4, very large MSEs are generated by the oracle estimator while the deviance and prediction errors maintain the similar level compared to other methods. This is not

Table 3. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, and specificity.

Method	MSE	PE	Deviance	Sensitivity	Specificity
		$p = 20$, $\rho = 0.5$			
Enet	**0.0057** (0.003)	**6.383** (3.2)	81.549 (12.59)	0.9 (0.15)	0.4 (0.15)
Mnet	0.0058 (0.003)	6.525 (3.63)	84.901 (12.63)	0.8 (0.15)	0.7 (0.3)
Snet	0.0058 (0.002)	6.666 (3.65)	84.613 (10.98)	0.9 (0.15)	0.6 (0.3)
LASSO	0.0063 (0.003)	6.711 (3.3)	**80.006** (12)	0.9 (0.15)	0.5 (0.22)
MCP	0.0097 (0.005)	8.586 (5.27)	81.952 (14.64)	0.6 (0.15)	0.8 (0.15)
SCAD	0.0097 (0.005)	8.663 (5.64)	81.08 (14.8)	0.7 (0.15)	0.8 (0.15)
ORACLE	0.0046 (0.002)	6.139 (3.06)	82.161 (10.92)	1 (0)	1 (0)
		$p = 20$, $\rho = 0.8$			
Enet	**0.0082** (0.003)	13.604 (10.18)	79.263 (14.07)	0.9 (0.15)	0.35 (0.22)
Mnet	0.0096 (0.003)	**12.629** (8.05)	83.358 (16.24)	0.9 (0.15)	0.5 (0.3)
Snet	0.0095 (0.003)	12.923 (7.82)	83.071 (15.06)	0.9 (0.15)	0.4 (0.44)
LASSO	0.010 (0.006)	14.543 (11.21)	**76.805** (13.55)	0.8 (0.15)	0.6 (0.15)
MCP	0.0191 (0.011)	20.393 (19.07)	77.843 (14.42)	0.6 (0.15)	0.7 (0.3)
SCAD	0.0199 (0.011)	20.163 (17.4)	78.021 (14.98)	0.6 (0.15)	0.7 (0.3)
ORACLE	0.0087 (0.005)	15.023 (10.93)	79.544 (13.01)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.5$			
Enet	**0.0016** (4e–04)	**11.192** (6.64)	73.817 (10.46)	0.7 (0.15)	0.88 (0.05)
Mnet	0.0018 (6e–04)	11.762 (7.18)	77.487 (12.72)	0.5 (0.15)	0.95 (0.05)
Snet	**0.0016** (5e–04)	11.473 (7.29)	78.231 (11.39)	0.7 (0.15)	0.87 (0.12)
LASSO	**0.0016** (5e–04)	11.643 (6.35)	**71.968** (11.73)	0.6 (0.15)	0.91 (0.02)
MCP	0.0032 (0.0013)	17.144 (10.21)	85.899 (17.98)	0.3 (0.15)	0.98 (0.01)
SCAD	0.0032 (0.0017)	18.038 (12.06)	86.981 (17.01)	0.3 (0.15)	0.97 (0.02)
ORACLE	4e–04 (3e–04)	5.951 (3.56)	84.226 (10.48)	1 (0)	1 (0)
		$p = 200$, $\rho = 0.8$			
Enet	**0.0022** (2e–04)	31.232 (26.5)	80.271 (13.79)	0.6 (0.15)	0.78 (0.16)
Mnet	0.0023 (2e–04)	30.063 (23.2)	83.624 (15.14)	0.65 (0.37)	0.77 (0.29)
Snet	**0.0022** (2e–04)	**29.949** (23.15)	87.688 (18.96)	0.8 (0.15)	0.5 (0.34)
LASSO	0.0025 (6e–04)	33.08 (28.89)	**77.644** (12.57)	0.4 (0.15)	0.92 (0.02)
MCP	0.007 (0.0018)	56.54 (40.61)	104.46 (26.18)	0.1 (0.15)	0.98 (0.01)
SCAD	0.0072 (0.0017)	67.32 (49.87)	104.373 (29.03)	0.1 (0.15)	0.98 (0.01)
ORACLE	8e–04 (5e–04)	15.47 (11.21)	80.875 (12.86)	1 (0)	1 (0)

Note: Methods in Poisson regression with scenario 3.

unique to this particular $\alpha = 0.2$. While it seems not intuitive, carefully evaluating the estimated coefficients provides the reasonable explanation. Again, the oracle estimator is not the oracle ridge estimator, thus it suffers the inherit drawback with highly correlated predictors. An oracle estimator may have generated large estimates in magnitude; however, the summation of these estimates is comparable to that of the true coefficients. For the simulated data in scenario 4, it is the summation of estimates that determines the accuracy of model fitting. Therefore, the deviance and prediction error are not affected by the large magnitude of the estimated coefficients. With multicollinearity as such, different models may fit the data equally well, but they provide quite different model structures.[7,12]

Table 4. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, and specificity.

Method	MSE	PE	Deviance	Sensitivity	Specificity
		$p = 20$			
Enet	**0.007** (0.007)	5.149 (2.6)	176.057 (20.02)	1 (0)	0.46 (0.27)
Mnet	0.0078 (0.011)	**4.780** (2.15)	180.402 (23.11)	0.89 (0.16)	0.68 (0.2)
Snet	0.0131 (0.018)	4.840 (2.17)	178.506 (23.61)	0.78 (0.33)	0.64 (0.27)
LASSO	0.0344 (0.012)	5.509 (2.46)	**175.001** (20.83)	0.78 (0.16)	0.55 (0.27)
MCP	0.0753 (0.005)	5.136 (2.13)	186.87 (25.02)	0.33 (0)	0.91 (0.13)
SCAD	0.0746 (0.005)	5.698 (2.74)	183.495 (21.73)	0.33 (0)	0.82 (0.27)
ORACLE	0.0624 (0.038)	4.464 (1.54)	180.304 (21.73)	1 (0)	1 (0)
		$p = 200$			
Enet	**0.0056** (0.0011)	7.359 (4.84)	**167.143** (23.06)	0.44 (0.16)	0.91 (0.05)
Mnet	0.0078 (5e–04)	3.849 (1.24)	181.971 (12.74)	0.33 (0)	1 (0)
Snet	0.0078 (4e–04)	3.777 (1.14)	180.463 (14.45)	0.33 (0)	1 (0)
LASSO	0.0057 (0.0011)	7.359 (4.84)	**167.143** (23.06)	0.44 (0.16)	0.91 (0.05)
MCP	0.0079 (4e–04)	**3.651** (0.93)	182.384 (13.23)	0.33 (0)	1 (0)
SCAD	0.0079 (4e–04)	3.729 (1.04)	181.178 (14.11)	0.33 (0)	1 (0)
ORACLE	0.4978 (0.3709)	4.683 (1.94)	179.469 (14.99)	1 (0)	1 (0)

Note: Methods in Poisson regression with scenario 4.

5 Analysis of the LOS

The TRIBE-AKI study aimed to evaluate the impact of biomarkers on the clinical outcomes after cardiac surgery. We applied the penalized count data regression to the hospital LOS. For Poisson and NB regression, it was convenient to subtract 2 from the observed LOS such that the minimum LOS value was zero. The distribution of LOS was then illustrated in Figure 1. The skewed distribution of LOS suggests that the ordinary linear regression will not be valid. The complete data had a sample size of 250 and the biomarkers were logarithmically transformed. Estimating methods include the EMSnet-penalized Poisson and EMSnet-penalized NB regression. The tuning parameters were selected via 10-fold cross-validation. For the selected parameters, the standard errors were estimated using the nonparametric bootstrap. We randomly sampled subjects with replacement and computed the bootstrap estimators with the same tuning parameters. The bootstrap procedure was repeated 500 times and the standard deviation of bootstrap estimators was calculated. The analysis results are summarized in Table 9. It is clear that the NB regression fit the data much better than Poisson regression, with substantial improvements on cross-validated log-likelihood, log-likelihood with the full data, AIC and BIC. Therefore, we focus on the analysis results from the NB regression. The models fitted by the EMSnet-penalty have similar log-likelihood values and dispersion parameter $\hat{\theta}$. The Mnet and Snet, however, outperform the Enet penalty with less predictors and smaller AIC and BIC. The selected clinical risk factors can be interpreted as follows. Since patients were recruited from study sites in the USA and Canada, apparently there existed some practical differences in the hospital management. The Risk Adjustment for Congenital Heart Surgery 1 (RACHS-1) consensus-based scoring system has been commonly utilized as a predictor of post-operative mortality.[13] Intraoperative variables include cardiopulmonary bypass (CPB) time (>120 min) and cross clamp time (>60 min). These prolonged times were negatively associated with hospital LOS. The use of an ACE inhibitor

Table 5. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, specificity, and $\hat{\theta}$.

Method	MSE	PE	Deviance	Sensitivity	Specificity	$\hat{\theta}$
			$p = 20,\ \rho = 0.5$			
Enet	0.0372 (0.026)	8.1578 (2.69)	98.992 (5.3)	1 (0)	0.3 (0.3)	2.41 (1.29)
Mnet	**0.0225** (0.015)	**7.587** (2.33)	98.151 (5.93)	1 (0)	0.7 (0.15)	3 (1.42)
Snet	0.0246 (0.018)	7.9424 (2.76)	98.354 (6.86)	1 (0)	0.6 (0.15)	2.76 (1.44)
LASSO	0.0366 (0.025)	8.084 (2.53)	98.879 (5.01)	1 (0)	0.5 (0.3)	2.34 (1.48)
MCP	0.0333 (0.029)	8.6329 (3.23)	98.142 (5.5)	0.9 (0.15)	0.8 (0.15)	2.4 (1.8)
SCAD	0.0286 (0.023)	8.6964 (3.57)	**97.828** (5.95)	1 (0)	0.7 (0.15)	2.65 (2)
ORACLE	0.0115 (0.006)	7.499 (2.86)	99.187 (5.68)	1 (0)	1 (0)	3.3 (1.87)
			$p = 20,\ \rho = 0.8$			
Enet	0.1018 (0.038)	3.495 (0.87)	99.701 (4.56)	0.65 (0.52)	0.5 (0.3)	1.83 (0.93)
Mnet	**0.0781** (0.049)	3.505 (0.92)	98.291 (6.41)	0.6 (0.44)	0.7 (0.3)	1.96 (0.97)
Snet	0.0871 (0.053)	**3.494** (0.92)	99.272 (5.18)	0.65 (0.37)	0.7 (0.22)	1.86 (0.88)
LASSO	0.1025 (0.043)	3.551 (0.86)	99.873 (4.97)	0.5 (0.44)	0.65 (0.37)	1.8 (0.9)
MCP	0.1087 (0.03)	3.755 (1.06)	**97.975** (5.34)	0.4 (0.3)	0.9 (0.15)	1.5 (0.63)
SCAD	0.1147 (0.022)	3.774 (1.09)	98.806 (5.27)	0.4 (0.3)	0.8 (0.15)	1.46 (0.66)
ORACLE	0.0314 (0.019)	3.609 (1.01)	100.603 (6.28)	1 (0)	1 (0)	2.74 (1.26)
			$p = 100,\ \rho = 0.5$			
Enet	0.026 (0.002)	10.635 (4.14)	95.679 (5.42)	0.2 (0.3)	0.97 (0.04)	0.67 (0.26)
Mnet	**0.025** (0.0028)	10.527 (4.01)	94.938 (6.83)	0.2 (0.3)	0.98 (0.03)	0.76 (0.32)
Snet	0.0255 (0.003)	**10.466** (4.12)	**94.799** (6.72)	0.2 (0.3)	0.97 (0.04)	0.74 (0.29)
LASSO	0.0266 (9e–04)	10.649 (3.98)	96.565 (5.28)	0.1 (0.15)	0.99 (0.02)	0.64 (0.26)
MCP	0.0257 (0.002)	10.529 (4.04)	95.4 (6.24)	0.1 (0.15)	0.99 (0.02)	0.69 (0.29)
SCAD	0.0262 (0.002)	10.671 (4.05)	96.197 (5.49)	0.1 (0.15)	0.99 (0.02)	0.67 (0.3)
ORACLE	0.0026 (0.001)	7.521 (2.99)	99.147 (6.24)	1 (0)	1 (0)	2.96 (1.2)
			$p = 100,\ \rho = 0.8$			
Enet	**0.0256** (0.001)	**3.781** (0.88)	99.745 (3.56)	0.2 (0.15)	0.94 (0.05)	1.2 (0.42)
Mnet	0.0257 (0.001)	3.795 (0.84)	99.537 (3.34)	0.2 (0.15)	0.97 (0.03)	1.24 (0.4)
Snet	0.0258 (0.002)	3.824 (0.86)	**99.382** (3.76)	0.2 (0.15)	0.94 (0.05)	1.21 (0.45)
LASSO	0.0259 (0.001)	**3.781** (0.88)	99.605 (3.52)	0.2 (0.15)	0.97 (0.03)	1.2 (0.42)
MCP	0.0258 (0.002)	3.818 (0.92)	99.701 (3.82)	0.1 (0.15)	0.98 (0.02)	1.26 (0.38)
SCAD	0.026 (0.001)	3.821 (0.93)	99.423 (3.74)	0.2 (0.15)	0.97 (0.03)	1.21 (0.44)
ORACLE	0.0063 (0.004)	3.364 (0.94)	101.27 (5.56)	1 (0)	1 (0)	3.04 (1.17)

Note: Methods in negative binomial regression with scenario 1.

(or angiotensin-converting-enzyme inhibitor) was an indication of severity of disease. Increased early postoperative urine biomarkers NGAL, IL18, and urine KIM-1 were associated with prolonged LOS. Also, reduced urine creatinine was associated with prolonged LOS. Our conclusions are similar to the previous analysis;[1] for instance, the early postoperative biomarkers urine NGAL, IL18, and creatinine predict hospital LOS. However, it is worth emphasizing that in the previous analysis these biomarkers were analyzed separately with linear regression and it did not answer the question whether these biomarkers independently predict LOS adjusting for other risk factors and biomarkers. The analysis in this paper fit the biomarkers jointly and the predictive power can be increased with multiple biomarkers. It also suggests that these biomarkers have different and

Table 6. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, specificity and $\hat{\theta}$.

Method	MSE	PE	Deviance	Sensitivity	Specificity	$\hat{\theta}$
			$p = 20$, $\rho = 0.5$			
Enet	0.021 (0.013)	76.121 (61.03)	94.255 (6.95)	1 (0)	0.7 (0.15)	2.47 (1.04)
Mnet	**0.0071** (0.006)	56.34 (38.34)	**90.303** (8.35)	1 (0)	0.95 (0.07)	2.29 (0.86)
Snet	0.0074 (0.006)	**55.016** (37.79)	91.235 (8.74)	1 (0)	0.9 (0.07)	2.29 (0.87)
LASSO	0.0188 (0.011)	68.984 (57.17)	94.16 (7.07)	1 (0)	0.7 (0.15)	2.53 (1.06)
MCP	0.0075 (0.007)	62.329 (47.92)	91.448 (9.99)	1 (0)	0.95 (0.07)	2.39 (1.03)
SCAD	0.0075 (0.007)	60.889 (44.03)	92.652 (9.17)	1 (0)	0.9 (0.07)	2.47 (1.13)
ORACLE	0.0049 (0.004)	59.678 (44.62)	91.914 (8.98)	1 (0)	1 (0)	2.46 (0.97)
			$p = 20$, $\rho = 0.8$			
Enet	0.0455 (0.022)	10.913 (4.28)	96.632 (5.05)	1 (0)	0.75 (0.22)	2.27 (0.88)
Mnet	**0.0213** (0.019)	**9.985** (3.93)	90.991 (8.41)	1 (0)	0.95 (0.07)	1.92 (0.64)
Snet	0.024 (0.02)	10.115 (3.62)	93.246 (5.87)	1 (0)	0.9 (0.07)	2.12 (0.68)
LASSO	0.0447 (0.021)	10.767 (4.11)	96.635 (5.15)	1 (0)	0.75 (0.22)	2.27 (0.89)
MCP	0.023 (0.02)	10.420 (4.49)	**88.797** (8.03)	1 (0)	1 (0)	1.84 (0.68)
SCAD	0.0282 (0.027)	10.506 (4.42)	94.384 (6.6)	1 (0)	0.95 (0.07)	2.09 (1.03)
ORACLE	0.0105 (0.008)	9.165 (3.92)	96.338 (6.38)	1 (0)	1 (0)	2.48 (1.07)
			$p = 200$, $\rho = 0.5$			
Enet	0.0091 (0.004)	85.508 (62.87)	91.331 (6.53)	1 (0)	0.96 (0.03)	1.04 (0.78)
Mnet	0.0028 (0.003)	63.604 (52.87)	85.253 (11.3)	1 (0)	0.98 (0.02)	2.19 (1.28)
Snet	**0.002** (0.0018)	**61.234** (45.44)	**84.568** (9.96)	1 (0)	0.97 (0.03)	2.01 (1.32)
LASSO	0.0097 (0.005)	86.875 (62.07)	92.244 (5.52)	0.8 (0.3)	0.97 (0.03)	1.03 (0.91)
MCP	0.0046 (0.005)	79.446 (58.12)	87.394 (10.13)	1 (0)	0.99 (0.01)	1.75 (1.83)
SCAD	0.0037 (0.004)	79.867 (66.98)	87.339 (9.3)	1 (0)	0.98 (0.02)	1.67 (1.69)
ORACLE	6e–04 (4e–04)	55.999 (42.52)	90.372 (6.05)	1 (0)	1 (0)	2.56 (0.89)
			$p = 200$, $\rho = 0.8$			
Enet	0.0147 (0.006)	12.051 (5.7)	96.769 (4.06)	0.6 (0.3)	0.97 (0.02)	1.34 (0.64)
Mnet	**0.010** (0.0047)	11.734 (6.21)	89.789 (10.94)	0.6 (0.3)	0.99 (0.01)	1.49 (0.69)
Snet	0.0109 (0.006)	**11.051** (5.18)	92.474 (6.37)	0.6 (0.3)	0.99 (0.01)	1.45 (0.68)
LASSO	0.0146 (0.006)	12.096 (5.75)	96.996 (4.27)	0.6 (0.3)	0.97 (0.02)	1.35 (0.67)
MCP	0.0113 (0.006)	12.024 (5.96)	**89.58** (10.85)	0.6 (0.3)	1 (0)	1.24 (0.62)
SCAD	0.0119 (0.007)	11.743 (5.23)	93.21 (5.16)	0.6 (0.3)	0.99 (0.01)	1.22 (0.58)
ORACLE	0.0013 (0.001)	8.057 (3.39)	97.203 (6.32)	1 (0)	1 (0)	2.47 (1.13)

Note: Methods in negative binomial regression with scenario 2.

strong predictive effect on postoperative LOS. For instance, even adjusting for early postoperative urine NGAL and IL18, KIM-1 still independently predicts LOS.

6 Discussion

In clinical and translational research, investigators are often interested in how to combine variables in a statistical model to optimize prediction power. With a small number of variables, one may test all possible subsets of variables for inclusion in the model. With many variables, such an exhaustive search can be daunting due to the "combinatorial explosion". The task only becomes more

Table 7. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, specificity, and $\hat{\theta}$.

Method	MSE	PE	Deviance	Sensitivity	Specificity	$\hat{\theta}$
			$p = 20,\ \rho = 0.5$			
Enet	**0.0109** (0.003)	**29.595** (19.41)	93.67 (7.71)	0.8 (0.15)	0.4 (0.15)	2.65 (0.89)
Mnet	0.0132 (0.006)	31.943 (22.34)	93.169 (6.03)	0.7 (0.3)	0.7 (0.3)	2.64 (0.87)
Snet	0.0136 (0.006)	30.311 (20.64)	**92.871** (7.27)	0.8 (0.15)	0.6 (0.3)	2.62 (0.95)
LASSO	0.0132 (0.005)	32.698 (22.96)	94.13 (8.56)	0.7 (0.15)	0.7 (0.15)	2.91 (1.18)
MCP	0.0215 (0.011)	39.927 (26.3)	93.264 (7.58)	0.5 (0.15)	0.8 (0.15)	2.62 (1.04)
SCAD	0.0219 (0.009)	39.667 (24.93)	93.586 (7.34)	0.5 (0.15)	0.8 (0.15)	2.54 (0.97)
ORACLE	0.0138 (0.007)	40.094 (30.41)	93.404 (6.98)	1 (0)	1 (0)	3.01 (1.32)
			$p = 20,\ \rho = 0.8$			
Enet	**0.0141** (0.004)	**115.236** (94.34)	91.06 (4.83)	0.8 (0.15)	0.4 (0.3)	2.36 (0.66)
Mnet	0.0183 (0.01)	123.947 (99.93)	89.614 (5.91)	0.6 (0.3)	0.7 (0.3)	2.33 (0.77)
Snet	0.0168 (0.008)	125.675 (99.39)	90.548 (5.59)	0.7 (0.3)	0.6 (0.3)	2.45 (0.73)
LASSO	0.0223 (0.009)	127.008 (98.66)	90.801 (5.57)	0.6 (0.15)	0.65 (0.07)	2.69 (0.99)
MCP	0.0539 (0.027)	150.202 (131.33)	**88.317** (7.17)	0.2 (0.15)	0.9 (0.15)	2.15 (0.75)
SCAD	0.0485 (0.026)	157.418 (135.13)	90.131 (6.97)	0.3 (0.15)	0.9 (0.15)	2.31 (0.94)
ORACLE	0.0364 (0.022)	132.236 (112.03)	91.037 (5.84)	1 (0)	1 (0)	2.93 (1.09)
			$p = 200,\ \rho = 0.5$			
Enet	**0.004** (9e–04)	37.238 (24.75)	93.155 (7.58)	0.6 (0.15)	0.78 (0.1)	2.56 (1.05)
Mnet	0.0041 (0.001)	**36.936** (22.11)	93.325 (7.41)	0.5 (0.15)	0.86 (0.13)	2.53 (1.03)
Snet	0.0043 (0.001)	39.274 (24.59)	93.283 (8)	0.5 (0.22)	0.83 (0.12)	2.56 (0.93)
LASSO	0.0043 (0.001)	38.161 (23.8)	93.414 (7.32)	0.5 (0.15)	0.88 (0.03)	3.21 (1.63)
MCP	0.0067 (0.003)	46.445 (32.34)	**90.789** (8.7)	0.3 (0.15)	0.97 (0.03)	2.48 (1.09)
SCAD	0.0058 (0.002)	47.863 (32.39)	91.906 (8.1)	0.4 (0.15)	0.91 (0.03)	2.39 (1)
ORACLE	0.0027 (0.002)	37.206 (22.67)	95.405 (7.5)	1 (0)	1 (0)	2.97 (1.51)
			$p = 200,\ \rho = 0.8$			
Enet	**0.005** (7e–04)	**125.678** (98.24)	91.589 (5.84)	0.5 (0.15)	0.71 (0.12)	2.7 (1.13)
Mnet	0.0054 (0.002)	141.644 (114.79)	91.126 (8.44)	0.3 (0.15)	0.91 (0.12)	2.36 (0.82)
Snet	0.0051 (0.001)	139.26 (118.42)	91.537 (6.28)	0.4 (0.3)	0.78 (0.24)	2.61 (0.92)
LASSO	0.0063 (0.002)	131.735 (93.12)	91.058 (6.26)	0.3 (0.15)	0.89 (0.03)	3.32 (1.62)
MCP	0.0142 (0.004)	176.021 (147.67)	**88.406** (9.71)	0.1 (0.15)	0.98 (0.02)	1.87 (0.72)
SCAD	0.0137 (0.007)	184.436 (142.09)	89.972 (9.48)	0.1 (0.15)	0.97 (0.03)	2.08 (0.84)
ORACLE	0.0066 (0.003)	149.993 (124.53)	91.684 (6.72)	1 (0)	1 (0)	2.76 (1.2)

Note: Methods in negative binomial regression with scenario 3.

challenging with highly correlated data since the parameter estimation often becomes more difficult. Penalized statistical models and efficient computing algorithms such as the EMSnet-penalty and coordinate descent algorithm are suitable for the analytical challenge. Hospital LOS is frequently used to assess severity of patients and hospital management. As LOS is often recorded at discrete time, count data regression including Poisson and NB regression is an important alternative to linear regression. The NB regression can be more powerful than Poisson regression when overdispersion occurs, which is frequently encountered in real data. This paper implements novel variable selection methods for count data regression. Simulations suggest that with highly correlated variables, the EMSnet-penalized count regression can generate more accurate parameter estimation. In an extreme

Table 8. Medians and robust standard deviations (in parentheses) of MSE, PE, deviance, sensitivity, specificity, and $\hat{\theta}$.

Method	MSE	PE	Deviance	Sensitivity	Specificity	$\hat{\theta}$
		$p = 20$				
Enet	0.0022 (0.001)	36.365 (23.79)	185.913 (6.78)	1 (0)	0.36 (0.27)	2.26 (0.63)
Mnet	**0.002** (0.001)	38.071 (23.87)	185.475 (6.71)	1 (0)	0.64 (0.13)	2.17 (0.54)
Snet	0.0049 (0.006)	36.212 (22.98)	185.643 (6.28)	1 (0)	0.55 (0.27)	2.27 (0.6)
LASSO	0.064 (0.01)	38.232 (24.09)	185.89 (6.44)	0.33 (0)	0.64 (0.27)	2.23 (0.61)
MCP	0.0803 (0.009)	**34.043** (20.88)	185.418 (5.74)	0.33 (0)	1 (0)	2.15 (0.63)
SCAD	0.0809 (0.009)	36.336 (22.26)	**185.087** (5.72)	0.33 (0)	0.96 (0.07)	2.14 (0.63)
ORACLE	20.452 (10.93)	37.979 (24.07)	184.909 (6.58)	1 (0)	1 (0)	2.37 (0.76)
		$p = 200$				
Enet	7e–04 (2e–04)	56.444 (43.41)	181.794 (10.73)	1 (0)	0.92 (0.05)	1.42 (0.75)
Mnet	**4e–04** (1e–04)	**40.205** (27.71)	184.006 (8.65)	1 (0)	0.95 (0.03)	2.1 (0.67)
Snet	6e–04 (2e–04)	45.099 (31.19)	**181.281** (9.36)	1 (0)	0.92 (0.05)	2.03 (0.81)
LASSO	0.0045 (9e–04)	40.598 (29.82)	184.652 (8.37)	0.33 (0.1)	0.96 (0.02)	2.09 (0.61)
MCP	0.0085 (0.002)	53.754 (36.67)	186.816 (13.06)	0.33 (0)	0.99 (0.01)	2.21 (0.74)
SCAD	0.0081 (0.001)	42.687 (29.09)	184.979 (11.56)	0.33 (0)	0.98 (0.02)	2.26 (0.62)
ORACLE	1.811 (1.250)	39.583 (21.81)	183.601 (9.35)	1 (0)	1 (0)	2.4 (0.58)

Note: Methods in negative binomial regression with scenario 4.

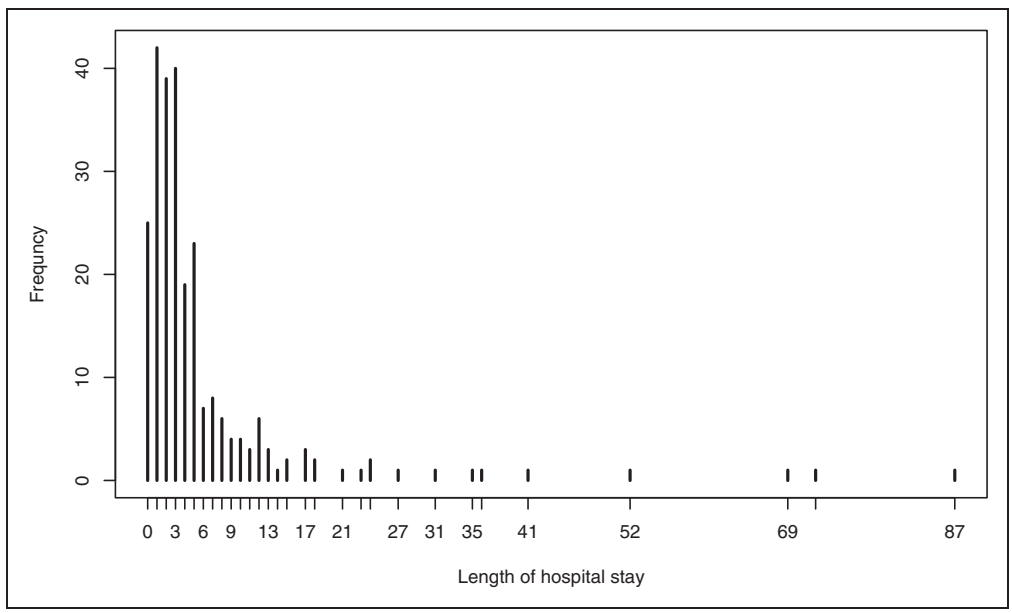

Figure 1. Empirical distribution of hospital stay for the TRIBE study.

Table 9. Penalized Poisson regression (–PR) and penalized negative binomial regression (–NB) for the TRIBE–AKI hospital length of stay.

	Enet-PR	Mnet-PR	Snet-PR	Enet-NB	Mnet-NB	Snet-NB
(Intercept)	3.45 (1.00)	3.39 (1.26)	3.66 (1.27)	2.66 (0.75)	3.85 (1.14)	3.40 (1.10)
Age	2e-05 (0.04)		0.01 (0.07)			
genderM						
Weight						
Site9	0.50 (0.27)	0.61 (0.38)	0.50 (0.39)	0.63 (0.21)	0.74 (0.33)	0.86 (0.30)
Site11	0.40 (0.27)	0.41 (0.37)	0.37 (0.36)	0.21 (0.22)		0.09 (0.29)
RACHS	0.01 (0.13)		0.01 (0.15)	0.07 (0.12)	0.09 (0.16)	0.07 (0.19)
CPB time				0.03 (0.12)		0.05 (0.18)
Cross-clamp time	0.29 (0.19)	0.26 (0.23)	0.29 (0.23)	0.24 (0.15)	0.10 (0.17)	0.12 (0.23)
ACE inhibitorsYes	0.29 (0.18)	0.30 (0.24)	0.27 (0.24)	0.40 (0.18)	0.51 (0.30)	0.64 (0.32)
aspirinYes				0.11 (0.15)		
eGFR pre						
SER_CRE pre						
SER_CRE 0–6 hrs	0.23 (0.25)	0.15 (0.31)	0.20 (0.31)	0.01 (0.13)		
SER_CYSC 0–6 hrs						
UR_IL18 0–6 hrs						
UR_IL18 6–12 hrs	0.07 (0.06)	0.08 (0.07)	0.06 (0.06)	0.13 (0.06)	0.17 (0.11)	0.23 (0.13)
UR_NGAL 0–6 hrs	0.04 (0.02)	0.10 (0.04)	0.03 (0.03)	0.02 (0.02)	0.01 (0.02)	0.003 (0.03)
UR_NGAL 6–12 hrs	0.07 (0.06)	0.001 (0.08)	0.07 (0.08)	0.04 (0.05)		
UR_CRE 0–6 hrs						0.06 (0.18)
UR_CRE 6–12 hrs	−0.67 (0.23)	−0.68 (0.34)	−0.72 (0.35)	−0.62 (0.18)	−0.93 (0.33)	−0.93 (0.33)
UR_KIM1 0–6 hrs				0.05 (0.06)		
UR_KIM1 6–12 hrs	0.36 (0.12)	0.37 (0.17)	0.39 (0.18)	0.30 (0.10)	0.48 (0.19)	0.46 (0.19)
UR_LFABP 0–6 hrs			0.01 (0.02)	0.01 (0.02)		
UR_LFABP 6–12 hrs						
UR_CYSC 0–6 hrs			0.01 (0.09)			
UR_CYSC 6–12 hrs						
UR_AC 0–6 hrs						
UR_AC 6–12 hrs						
UR_MALB 0–6 hrs				0.01 (0.03)		
UR_MALB 6–12 hrs						
CV log-likelihood	−123.78	−123.25	−123.48	−67.89	−67.20	−67.03
Log-likelihood	−1032.56	−1034.91	−1031.30	−645.03	−647.14	−643.45
AIC	2091.12	2091.81	2092.61	1326.05	1314.27	1312.91
BIC	2136.90	2130.55	2145.43	1389.44	1349.48	1358.69
$\hat{\theta}$				1.52	1.47	1.53

Note: Estimated coefficients and bootstrap standard errors in parentheses, cross validated (CV) log-likelihood, log-likelihood for full data, AIC and BIC values. Estimated NB regression dispersion parameter $\hat{\theta}$.

case with grouped variables, the traditional count regression, in particular the NB regression, may generate quite different parameter estimation than the true values, even if the effective variables were assumed known in advance. Such a discrepancy further supports that it is very important to take into account the highly correlated variables.

Acknowledgement

We thank the two referees whose comments improved an earlier draft of this article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: ZW is partially supported by a grant from the Charles H. Hood Foundation, Inc., Boston, MA. CYW is partially supported by National Institutes of Health grants CA53996, ES017030, and HL121347.

References

1. Parikh CR, Devarajan P, Zappitelli M, et al. Postoperative biomarkers predict acute kidney injury and poor outcomes after pediatric cardiac surgery. *J Am Soc Nephrol* 2011; **22**: 1737–1747.
2. Austin PC, Rothwell DM and Tu JV. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Service Outcome Res Methodol* 2002; **3**: 107–133.
3. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc Ser B* 1996; **58**: 267–288.
4. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *Ann Stat* 2010; **38**: 894–942.
5. Fan J and Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 2001; **96**: 1348–1360.
6. Zou H and Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc Ser B* 2005; **67**: 301–320.
7. Huang J, Breheny P, Ma S, et al. The Mnet method for variable selection. Technical Report #402, Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA, 2010.
8. Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software* 2010; **33**: 1–22.
9. Fan J and Lv J. Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transact Inform Theory* 2011; **57**: 5467–5484.
10. Breheny P and Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann Appl Stat* 2011; **5**: 232–253.
11. Wu TT and Lange K. Coordinate descent algorithms for lasso penalized regression. *Ann Appl Stat* 2008; **2**: 224–244.
12. Breiman L. Statistical modeling: the two cultures. *Stat Sci* 2001; **16**: 199–231.
13. Jenkins KJ, Gauvreau K, Newburger JW, et al. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thoracic Cardiovasc Surg* 2002; **123**: 110–118.