



Published in final edited form as:

Int J Stat Med Res. 2021 ; 10: 118–131. doi:10.6000/1929-6029.2021.10.11.

Fixed Effects High-Dimensional Profiling Models in Low Information Context

Jason P. Estes¹, Damla entürk², Esra Kürüm³, Connie M. Rhee⁴, Danh V. Nguyen^{4,*}

¹Mountain View, CA 94043, USA

²Department of Biostatistics, University of California, Los Angeles, CA 90095, USA

³Department of Statistics, University of California, Riverside, CA 92521, USA

⁴Department of Medicine, University of California Irvine, Orange, CA 92868, USA

Abstract

Profiling or evaluation of health care providers, including hospitals or dialysis facilities, involves the application of hierarchical regression models to compare each provider's performance with respect to a patient outcome, such as unplanned 30-day hospital readmission. This is achieved by comparing a specific provider's estimate of unplanned readmission rate, adjusted for patient case-mix, to a normative standard, typically defined as an "average" national readmission rate across all providers. Profiling is of national importance in the United States because the Centers for Medicare and Medicaid Services (CMS) policy for payment to providers is dependent on providers' performance, which is part of a national strategy to improve delivery and quality of patient care. Novel high dimensional fixed effects (FE) models have been proposed for profiling dialysis facilities and are more focused towards inference on the tail of the distribution of provider outcomes, which is well-suited for the objective of identifying sub-standard ("extreme") performance. However, the extent to which estimation and inference procedures for FE profiling models are effective when the outcome is sparse and/or when there are relatively few patients within a provider, referred to as the "low information" context, have not been examined. This scenario is common in practice when the patient outcome of interest is cause-specific 30-day readmissions, such as 30-day readmission due to infections in patients on dialysis, which is only about ~ 8% compared to the > 30% for all-cause 30-day readmission. Thus, we examine the feasibility and effectiveness of profiling models under the low information context in simulation studies and propose a novel correction method to FE profiling models to better handle sparse outcome data.

This is an open access article licensed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited.

*Address correspondence to this author at the Department of Medicine, University of California Irvine, 333 City Blvd. West, City Tower, Suite 400, Orange, CA 92868, USA; danhvn1@hs.uci.edu.

Online Supplementary Materials: Analysis Example and R Codes

Example dataset, R codes, and tutorial for fitting the uncorrected and corrected models are publicly available at <https://sites.google.com/view/usrds-modeling/software>.

Keywords

End-stage renal disease; fixed effects; high-dimensional parameters; logistic regression; infrequent events; Firth's correction

1. INTRODUCTION

Unplanned readmissions following a hospital discharge are a major source of morbidity and mortality risk for patients on dialysis. The burden of hospitalization is particularly high for patients on dialysis, where the latest U.S. national data shows that the frequency of 30-day readmissions is 31.1%, which is more than double the frequency of readmissions seen in older Medicare beneficiaries without kidney disease (United States Renal Data System/USRDS [1]).

Profiling or evaluation of health care providers, such as hospitals, dialysis facilities, and nursing homes among others, serves multiple purposes, including (1) identifying providers with performance below standard by government agencies for regulatory or payment purposes and (2) conveying information and feedback to stakeholders (e.g., the public, patients, providers) regarding the quality of care among providers. The main focus of our work is objective (1), specifically with respect to the goal of identifying providers whose performances (e.g., 30-day readmission) are exceptionally worse (W) and not different (ND) relative to a reference, such as a national “average” standard. Also related to the inferential process of identifying/flagging providers with 30-day readmission rates W and ND from the national rate, it is of interest to obtain accurate estimates of provider-specific effects and associated quality metrics.

When the outcome, such as 30-day readmission, is not frequent and/or when there are relatively few patients within a provider, referred to as the “low information” context [2], estimation and inference for profiling models are understandably more challenging. This is the situation when the patient outcome of interest is cause-specific 30-day readmissions, such as 30-day readmission due to infections in patients on dialysis, which is only about ~8% compared to greater than 30% for all-cause 30-day readmission. Infection-related hospitalizations are serious adverse events that are oftentimes preventable. Hence, it is an important performance indicator that is carefully monitored in dialysis facilities.

Respecting the data structure that patients are nested within providers, current profiling models for 30-day unplanned hospital readmission are hierarchical logistic regressions of the form *outcome* ~ *provider effects* + *patient case-mix effects*. Thus, patient outcomes vary across providers due to variation in providers' quality of care (provider-specific effects) and variation in patient-level case-mix effects, which include demographics, comorbidities, and the type of index admission. Because of the nested data structure and the need to stabilize estimation, modeling provider effects as random effects (RE) has been used [2-7].

A justification for the use of RE models is that they provide stable provider effect estimates through shrinkage, although several inherent disadvantages have been noted. In particular, RE estimates are biased toward the overall provider average and biased in the presence

of confounding between patient risk factors and provider effects [8]. Also, although the overall *average error* in estimation of provider effects is smaller because mean square error is minimized over the full set of provider effects in the RE approach, fixed effects (FE) estimates have smaller error for outlier ‘providers whose effects are exceptionally large or small’ [8], which are the providers we wish to identify. Our previous works also have shown that the benefit of stabilization comes at a severe cost in substantially biased provider effects estimation and, perhaps more important, at a substantial reduction in the power to identify W providers [9, 10]. Our works and others have used high-dimensional FE models to identify sub-standard (“extreme”) performance, especially for profiling 30-day readmission for dialysis facilities where the outcome is not sparse [3, 8-15]. However, the extent to which FE models are useful in the low information context has not been studied, which is the focus of this work. Thus, we assess the relative performance of the FE model proposed by He *et al.* [15], including the stability of provider-specific estimates and the ability to identify extreme providers in simulation studies. Briefly, the FE model of He *et al.* [15] is a high-dimensional parameter model with a unique fixed intercept for each provider and is used in assessing the performance of dialysis facilities [3, 8, 15]; see also Chen *et al.* [14] and Estes *et al.* [11, 12] for recent dialysis facility profiling applications. Furthermore, in this work, we also propose and examine the performance of a novel corrected FE model estimation approach geared towards estimation under low information context, where the (uncorrected) FE model estimates of some provider-specific effects may be unreliable.

2. METHODS: HIGH-DIMENSIONAL FE PROFILING MODELS

We introduce the FE profiling model using the context of hospital readmission as an illustrative example. Let the binary outcome Y_{ij} equal 1 if patient index discharge j in provider i results in a readmission within 30 days, for patient discharge $j = 1, 2, \dots, N_j$ in provider (dialysis facility) $i = 1, 2, \dots, F$. The FE profiling model (He *et al.* [15]) is

$$g(\mu_{ij}) = \gamma_i + \beta^T Z_{ij}, \quad i = 1, \dots, F, \quad (1)$$

where $\gamma = (\gamma_1, \dots, \gamma_F)$ are the provider-specific fixed effects, $\mu_{ij} \equiv E(Y_{ij} | Z_{ij}) = \Pr(Y_{ij} = 1 | \beta, \gamma_i, Z_{ij}) = p_{ij}$ is the expected readmission for patient index discharge $j = 1, 2, \dots, N_j$ in provider $i = 1, 2, \dots, F$, and $g(p_{ij}) = \log\{p_{ij}/(1 - p_{ij})\}$ is the logit function. In profiling model (1), the r patient risk adjustment factors for discharge j in provider i are denoted by the covariate vector $Z_{ij} = (Z_{ij1}, \dots, Z_{ijr})^T$ corresponding to parameters $\beta^T = (\beta_1, \dots, \beta_r)$. In practice, the process of risk adjustment is complex and depends, in part, on policy objectives and the specific patient population (e.g., general population, dialysis population). However, we point out that it is critical to adequately risk-adjust for patient-level factors and *avoid* inclusion of variables (e.g., provider-level or patient-level variables) that are/may be related to the process of care (e.g., see [2, 3, 13]).

To avoid confusion, we emphasize that the model shown in (1) is not a collection of individual models (i.e., not a separate model for each provider), but rather a single model with high-dimensional parameters and requires simultaneous estimation for thousands of provider-specific effects/parameters ($\{\gamma_i\}_{i=1}^F$ and β). For example, for profiling dialysis facilities the dimension of $\gamma = (\gamma_1, \dots, \gamma_F)^T$ is $> 6,000$ dialysis facilities across the U.S., and

the dimension of β is ~ 40 . Standard estimation (e.g., maximum likelihood) and software fails; thus, He *et al.* (2013) proposed a feasible estimation method based on an alternating one-step Newton-Raphson that iterates between estimation of β and γ_i .

The summary performance index for each provider which incorporates patient-level risk factors (Z 's) used in practice is the standardized readmission ratio (SRR). For FE model (1), given the provider and the patient case-mix effect estimates, denoted by $\hat{\gamma}_i$ and $\hat{\beta}$, respectively, the estimated SRR for provider i is

$$SRR_i = \frac{\sum_{j=1}^{N_i} \hat{p}_{ij}}{\sum_{j=1}^{N_i} \hat{p}_{M,ij}}, \quad (2)$$

where $\hat{p}_{ij} = g^{-1}(\hat{\gamma}_i + \hat{\beta}^T Z_{ij})$ is the estimated probability of readmission for patient j in provider i and $\hat{p}_{M,ij} = g^{-1}(\hat{\gamma}_M + \hat{\beta}^T Z_{ij})$. The aggregate parameter $\hat{\gamma}_M$ in the denominator is taken to be the median of the $\{\hat{\gamma}_i\}_{i=1}^F$. Thus, the numerator of SRR_i is the expected total number of readmissions for provider i and the denominator is the expected total number of readmissions for an “average” facility (taken over the population of all providers), adjusted for the particular case-mix of the *same* patients in provider i . Note that SRR_i estimates the true quantity $\tilde{SRR}_i = \sum_{j=1}^{N_i} p_{ij} / \sum_{j=1}^{N_i} p_{M,ij}$, where $p_{M,ij} = g^{-1}(\gamma_M + \beta^T Z_{ij})$.

3. ESTIMATION AND INFERENCE PROCEDURES

In addition to the challenge of high-dimensional parameters, compounding difficulties are encountered in the low information context where the outcome is sparse, resulting in providers with few readmissions or even no readmission. For very small providers with few patients, there is very low information to assess performance. In extreme cases of providers with no or very low readmission, the FE estimation method [15] leads to unstable estimates for those providers. Thus, in the low information context, we propose a correction to the FE estimates for provider-specific effects.

3.1. FE Model Estimation

To describe our proposed FE corrected estimation for provider-specific effects, we first set the notation for the likelihood of the FE model (1) and briefly summarize the alternating Newton-Raphson algorithm proposed by He *et al.* [15]. For the FE model (1), $\Pr(Y_{ij} = 1 \mid Z_{ij}) = p_{ij}^{y_{ij}}(1 - p_{ij})^{1 - y_{ij}}$, and the likelihood function is given by

$$L(\gamma, \beta) = \prod_{i=1}^F \prod_{j=1}^{N_i} \frac{\exp\{(\gamma_i + \beta^T Z_{ij})y_{ij}\}}{1 + \exp(\gamma_i + \beta^T Z_{ij})}. \quad (3)$$

Because direct maximization of (3) is not feasible with standard methods when F is large (e.g., $F \sim 6,000$), He *et al.* (2013) proposed an effective iterative algorithm that alternates between estimation of γ_i given β and estimation of β given γ_i using one-step Newton-

Raphson updates. More precisely, estimation of the high-dimensional parameters (γ, β) are feasible since the likelihood (3) can be written as $L(\gamma, \beta) = \prod_i L_i(\gamma_i, \beta)$ where $L_i(\gamma_i, \beta) = \prod_j \exp\{(\gamma_i + \beta^T Z_{ij}) y_{ij}\} / x \{1 + \exp(\gamma_i + \beta^T Z_{ij})\}$ for provider i . Thus, given β , γ_i can be estimated via a Newton-Raphson procedure that depends only on one variable in the maximization of $L_i(\gamma_i, \beta)$. Briefly, the estimation procedure proposed by He *et al.* (2013) is as follows.

- i. Set the initial values $\beta^{(0)}$ and $\gamma_i^{(0)}$ of β and γ_i , respectively.
- ii. The $(m + 1)$ th maximization step for β , given $\gamma_i^{(m)}$, is

$$\beta^{(m+1)} = \beta^{(m)} + I_\beta^{(m)-1} U_\beta^{(m)},$$

where $I_\beta^{(m)} = -\frac{\partial^2}{\partial \beta \partial \beta^T} \log L(\gamma^{(m)}, \beta) |_{\beta = \beta^{(m)}}$ and $U_\beta^{(m)} = \frac{\partial}{\partial \beta} \log L(\gamma^{(m)}, \beta) |_{\beta = \beta^{(m)}}$.

- iii. The $(m + 1)$ th maximization step for γ_i , given $\beta^{(m)}$, is

$$\gamma_i^{(m+1)} = \gamma_i^{(m)} + I_i^{(m)-1} U_i^{(m)}$$

where $I_i^{(m)} = -\frac{\partial^2}{\partial \gamma_i^2} \log L(\gamma_i, \beta^{(m+1)}) |_{\gamma_i = \gamma_i^{(m)}}$ and

$$U_i^{(m)} = \frac{\partial}{\partial \gamma_i} \log L(\gamma_i, \beta^{(m+1)}) |_{\gamma_i = \gamma_i^{(m)}}$$

- iv. The above steps are repeated until convergence, defined by

$$\max_{i,j} | p_{ij}^{(m+1)} - p_{ij}^{(m)} | < \varepsilon, \text{ where } p_{ij}^{(m)} = g^{-1}(\gamma_i^{(m)} + \beta^{(m)T} Z_{ij}) \text{ and } \varepsilon \text{ is some}$$

prespecified tolerance level. Denote these final uncorrected provider-specific estimates as $\hat{\gamma}^U = (\hat{\gamma}_i^U, \dots, \hat{\gamma}_F^U)$.

Expressions for $I_\beta^{(m)}$, $U_\beta^{(m)}$, $I_i^{(m)}$, and $U_i^{(m)}$ are given in He *et al.* (2013)

and they are provided here for convenience: $I_\beta^{(m)} = \sum_{i=1}^F \sum_{j=1}^{N_i} p_{ij}^{(m)} (1 - p_{ij}^{(m)}) Z_{ij} Z_{ij}^T$,
 $U_\beta^{(m)} = \sum_{i=1}^F \sum_{j=1}^{N_i} (y_{ij} - p_{ij}^{(m)}) Z_{ij}$, $I_i^{(m)} = \sum_{j=1}^{N_i} p_{ij}^{(m)} (1 - p_{ij}^{(m)})$, and $U_i^{(m)} = \sum_{j=1}^{N_i} (y_{ij} - p_{ij}^{(m)})$.

Programs in R, sample data, and tutorial are provided in the online Supplementary

Materials. In our implementation, we choose $\beta^{(0)} = 0$ and $\gamma_i^{(0)} = \log\{\hat{p}_i^* / (1 - \hat{p}_i^*)\}$ where

$\hat{p}_i^* = (N_i + 1)^{-1} \{0.5 + \sum_{j=1}^{N_i} y_{ij}\}$, the Jeffreys' prior estimated proportion for facility i (i.e., posterior mean of a Beta distribution, $Beta(0.5 + \sum_{j=1}^{N_i} y_{ij}, 0.5 + N_i - \sum_{j=1}^{N_i} y_{ij})$).

3.2. Corrected Estimation of Provider Effects

As described earlier, estimation of provider effects, γ_i for the FE model can be unstable for some providers in the low information context. Thus, we consider an approach to “correct”

or stabilize FE estimates. We adapt the Firth correction in (standard) logistic regression [16, 17] to the high-dimensional FE model (1). Recall that for the standard (non-hierarchical data) logistic regression model with N independent units, $j = 1, \dots, N$, $\Pr(Y_j = 1 | Z_j, \theta) = \{1 + \exp(-\sum_{r=1}^p Z_{jr}\theta_r)\}^{-1} \equiv \pi_j$, where $\theta = (\theta_1, \dots, \theta_p)$ are regression coefficients for covariates $Z_j = (Z_{j1}, \dots, Z_{jp})$. Firth's modified score equations [16] for estimation to reduce small sample bias is $U^*(\theta_r) \equiv U(\theta_r) + 0.5 \text{trace}[I(\theta)^{-1} \{ I(\theta) / \theta_r \}] = 0$, for $r = 1, \dots, p$, where $U(\theta_r) = \log L / \theta_r$, $I(\theta)$ is the information matrix, and $L = L(\theta)$ denotes the likelihood. This is equivalent to using a penalized likelihood $L^*(\theta) = L(\theta) | I(\theta) |^{-1/2}$ [17], where the penalty term $|I(\theta)|^{-1/2}$ is equivalent to Jeffreys' prior [18]. Applying this to logistic regression yields the modified estimation equations $U(\theta_r)^* = \sum_{j=1}^N \{y_j - \pi_j + h_j(0.5 - \pi_j)\} Z_{jr} = 0$ for $r = 1, \dots, p$, with h_j as the j th diagonal element of the "hat" matrix $H = W^{1/2} Z(Z^T W Z)^{-1} Z^T W^{1/2}$, with $W = \text{diag}\{\pi_1(1 - \pi_1), \dots, \pi_N(1 - \pi_N)\}$ and Z denotes the $N \times p$ data matrix. For binary outcome with small sample size, Firth's logistic regression has become a standard approach to reduce bias in the estimated regression coefficients.

We adapt this penalized estimation to the high-dimensional FE model (1) to correct for unstable estimation of γ_i for providers with low information. We first note that β can be precisely estimated because it is based on data from all providers; therefore, penalization on patient-level risk factors is unnecessary. Direct application of the Firth's modified score to penalize $\gamma = (\gamma_1, \dots, \gamma_F)$ is not feasible for FE profiling model (1) due to the challenge of calculating the score penalties. These are obtained via the diagonals of the $N \times N$ hat matrix, which in dialysis population applications are in the order of $N \sim 500,000$ or larger. The size of N is many orders of magnitude larger for profiling applications in the general population. However, estimating β with Firth's correction, for a fixed β , is equivalent to sequentially estimating γ_i individually, for a fixed β , using Firth's correction. This is seen as follows. For a fixed β , the hat matrix used in the estimation of β with Firth's correction is $H = W^{1/2} X(X^T W X)^{-1} X^T W^{1/2}$, where $W = W_1 \oplus \dots \oplus W_F$, $X = X_1 \oplus \dots \oplus X_F$, $W_i = \text{diag}\{p_{i1}(1 - p_{i1}), \dots, p_{iN_i}(1 - p_{iN_i})\}$ are provider-specific weight matrices, X_i are $N_i \times 1$ provider-specific design matrices of ones, and \oplus denotes the matrix direct sum operator, e.g., $A \oplus B$ is the block diagonal matrix $[A, 0; 0, B]$. As shown in the Supplementary Appendix section, $H = W^{1/2} X(X^T W X)^{-1} X^T W^{1/2} = W_1^{1/2} X_1(X_1^T W_1 X_1)^{-1} X_1^T W_1^{1/2} \oplus \dots \oplus W_F^{1/2} X_F(X_F^T W_F X_F)^{-1} X_F^T W_F^{1/2}$.

Thus, the diagonal of H may be obtained sequentially via the diagonals of $W_i^{1/2} X_i(X_i^T W_i X_i)^{-1} X_i^T W_i^{1/2}$ for each provider i .

The i th provider hat matrix

$$\begin{aligned} \text{reduces to } H_i &= W_i^{1/2} X_i(X_i^T W_i X_i)^{-1} X_i^T W_i^{1/2} = W_i^{1/2} X_i \{ (W_i^{1/2} X_i)^T \\ & (W_i^{1/2} X_i) \}^{-1} (W_i^{1/2} X_i)^T = (w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2})^T \{ (w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2}) \\ & (w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2})^T \}^{-1} (w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2}) \text{ where } w_{ij} = p_{ij}(1 - p_{ij}). \end{aligned}$$

$$\text{Thus, } \text{diag}(H_i) = (\sum_{j=1}^{N_i} w_{ij})^{-1} \text{diag}\{w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2}\}^T$$

$$(w_{i1}^{1/2}, \dots, w_{iN_i}^{1/2}) = (w_{i1} / \sum_{j=1}^{N_i} w_{ij}, \dots, w_{iN_i} / \sum_{j=1}^{N_i} w_{ij}) \text{ and for a fixed } \beta, \text{ the}$$

estimation of β using Firth's correction can be reduced to a sequence of estimations of a single parameter γ_i by penalizing the score U_i using the weights

$$h_{ij} = w_{ij} / \sum_{j=1}^{N_i} w_{ij}. \text{ More specifically, the provider-specific penalized score equations are } U_i^* = \sum_{j=1}^{N_i} \{y_{ij} - p_{ij} + h_{ij}(0.5 - p_{ij})\} = 0, \text{ for } i = 1, \dots, F.$$

We propose a simple correction to adjust the estimates from Section 3.1 of provider-specific effects, γ_i 's, using the modified score U_i^* . More precisely, first, β is fixed at the estimate resulting from Section 3.1, namely $\hat{\beta}_U$. The provider effects γ_i 's are then reestimated using the estimation procedure outlined in 3.1 with the following modifications. In Step (i), $\beta^{(0)}$ is fixed at $\hat{\beta}_U$ and $\gamma_i^{(0)}$ is set to $\log\{\hat{p}_i / (1 - \hat{p}_i)\} - N_i^{-1} \sum_{j=1}^{N_i} \hat{\beta}_U^T Z_{ij}$. Note that when $\beta^{(0)}$ is set to the zero vector, the initial value of $\gamma^{(0)}$ reduces to value previously noted in Step (i) in Section 3.1. In Step (ii), $\beta^{(m+1)}$ is set equal to $\beta^{(m)}$. In other words, β is no longer estimated. Finally, the score in Step (iii) is modified by replacing U_i with U_i^* .

3.3. Inference: Identifying Extreme Providers

In profiling, one of the main interests is to identify/flag providers that significantly deviate from the national norm (e.g., national average). The current public policy in the U.S. penalizes providers that perform significantly W than the national standard ($SRR > 1$). Thus, in practice, the goal is to flag/identify providers as W or ND from the national standard (SRR not different than 1). Better (B) providers ($SRR < 1$) are not penalized nor incentivized.

First, note that for a provider with an adjusted event rate that does not differ from the national norm, $\gamma_i = \gamma_M$, which implies $SRR_i = 1$. When $SRR_i > 1$ or $SRR_i < 1$, the event rate for provider i is greater than or less than the national norm, respectively. Thus, testing the null hypothesis $H_0 : \gamma_i = \gamma_M$ is of interest and a test statistic is $T_i = \sum_{j=1}^{N_i} \hat{p}_{ij}$ where \hat{p}_{ij} is an estimate of p_{ij} .

Simultaneously testing the null hypothesis for thousands of providers is computationally expensive. However, one can take advantage of the fact that β and γ_M can be estimated based on the large data from all providers. Hence, these parameters are estimated and fixed throughout the proposed algorithm below which is based on resampling responses under the null hypothesis. Since the global parameters β and γ_M are fixed, model fitting to the resampled data only requires estimation of provider-level effects γ_i . This reduces the computational burden substantially since each γ_i is estimated using only data from each provider separately. The steps of the procedure for each provider i are as follows.

1. Draw B samples $\{Y_{ij}^b : j = 1, \dots, N_i\}_{b=1}^B$, where each sample and observation is drawn independently from a Bernoulli distribution under the null:
 $Y_{ij}^b \sim \text{Bern}(g^{-1}\{\hat{\gamma}_M + Z_{ij}^T \hat{\beta}\})$, for $b = 1, \dots, B$. (We used $B = 500$.)
2. Calculate the test statistics for datasets generated/simulated under the null:
 $T_i^b = \sum_{j=1}^{N_i} \hat{p}_{ij}^b$ where, $\hat{p}_{ij}^b = g^{-1}(\hat{\gamma}_i^b + Z_{ij}^T \hat{\beta})$ and estimation of $\hat{\gamma}_i^b$ only involves

steps (iii)-(iv) in Section 3.1 for the uncorrected FE model since β is fixed. For the correction method, the estimation proceeds as described earlier in Section 3.2; that is, the corrected estimation algorithm is applied to the b th dataset to obtain \hat{p}_{ij}^b .

3. A nominal two-sided p -value for the i th provider, P_i is calculated as

$$P_i = 2 \cdot \min \left[\begin{array}{l} B^{-1} \sum_{b=1}^B \{0.5I(T_i^b = T_i^O) + I(T_i^b > T_i^O)\}, \\ B^{-1} \sum_{b=1}^B \{0.5I(T_i^b = T_i^O) + I(T_i^b < T_i^O)\} \end{array} \right],$$

where T_i^O is calculated based on the original/observed data and $I(A)$ denotes the indicator function for event A .

4. SIMULATION STUDY DESIGN

We designed simulation studies to assess the performance of the uncorrected and corrected FE model estimation methods, mainly with respect to (A) estimation of provider-specific effects, γ_i 's and SRR_i 's; and (B) identification of extreme providers relative to a reference. Data were generated from the model

$$g(\mu_{ij}) = \gamma_0 + \gamma_i + \beta_1 Z_{ij1} + \dots + \beta_{15} Z_{ij15} \quad (4)$$

with $i = 1, \dots, F = 5,000$ providers and $\beta = (.25, .25, -.25, -.25, .5, .25, .25, .25, .25, -.25, -.25, -.25, .5, .5, .5)^T$. For the patient case-mix vector, Z_{ij} , the dependence/correlation structure among variables were based on the observed correlations among patient-level variables in real USRDS data. More specifically, Z^* is generated from a multivariate normal distribution with means zero and covariance $Cov(Z^*) = V^{1/2} R V^{1/2}$, where $V^{1/2} = \text{diag}\{\sqrt{\text{Var}(Z_1^*)}, \dots, \sqrt{\text{Var}(Z_{15}^*)}\}$ and R is the correlation matrix. The first 5 covariates were taken to be continuous: $Z_1 \equiv Z_1^*, \dots, Z_5 \equiv Z_5^*$. The remaining 10 covariates, Z_6, \dots, Z_{15} are binary variables, generated by thresholding corresponding Z_r^* so that $\Pr(Z_r = 1) = E(Z_r)$'s are equally spaced between 0.2 and 0.8 (for $r = 6, \dots, 15$). The correlation matrix and standard deviations of the 15 variables are provided in the Supplementary Appendix.

For the provider effects, $\{\gamma_i\}_{i=1}^F$, 2.5% were under-performers (W: "worse") and 2.5% were over-performers (B: "better") whose effects, γ_i 's, were equally spaced in the intervals $[0.4, 1.0]$ and $[-1.0, -0.4]$, respectively. The remaining 95% of providers, with effects not different (ND) from the reference, were generated from a $\mathcal{N}(0, \sigma^2)$ distribution with $\sigma^2 = 0.2^2$. Note that a constant γ_0 has been added to simulation model (4) to conveniently control the baseline rate of readmission (outcome data sparsity), where baseline rates of readmission considered were 20%, 10%, 5%, and 3% corresponding to $\gamma_0 = \log(1/13.5)$, $\log(1/33)$, $\log(1/73)$, and $\log(1/126)$, respectively. This setup conveniently regulates the level of outcome data sparsity. For each baseline readmission rate setting, 200 datasets were

generated and the estimation (Section 3) and inference procedure (Section 3.3) was applied to each simulated dataset.

The provider volume of each generated dataset range from a minimum of 48 to a maximum of 195 patients on average, similar to real USRDS data in applications (e.g., see [14]). More specifically, the number of patients were generated from a truncated Poisson distribution following He *et al.* (2013), where the number of patients was taken to be $N_i = \sum_{h=1}^{1000} m_{ih} 1\{m_{ih} \leq 7\}$ with $m_{ih} \sim \text{Poisson}(15)$. This process mimics the sparse data structure of dialysis facility (provider) i in practice.

5. RESULTS

5.1. Estimation of Provider-Specific Effects and SRRs

The results for provider-specific estimates of γ_i 's for the 125 (2.5%) under-performers ($\gamma_i > 0$) and 125 over-performers ($\gamma_i < 0$) for the case of 3% overall event rate (most sparse) are provided in Figure 1A where averages of γ_i estimates over 200 simulated data sets are plotted. As expected, under this extremely low information context, provider effect estimates are unstable for the uncorrected FE method. However, note that these providers are mainly the over-performers ($\gamma_i < 0$) with low or zero events ($\sum_j y_{ij}$'s are small) leading to "explosion" of the estimates (Figure 1A). It is important to note that these unstable estimates are in the direction of the true effect (negative direction for negative γ_i 's, where $\hat{\gamma}_i \rightarrow -\infty$). Also as expected, estimates for under-performers ($\gamma_i > 0$) are less unstable and more on target for the uncorrected FE method. The corrected estimation approach, which adapts the Firth's modified score equation for the FE model, largely eliminates the instability and estimates are on more target for the true γ_i 's (Figure 1B).

Figure 2 (left column) shows estimates of γ_i 's for increasing percentage of overall events, from 3% to 20% for the uncorrected FE method. Clearly, the frequency of unstable estimates for $\gamma_i < 0$ decreased with increasing overall events, although unstable estimates are apparent even at a 10% event rate. However, the magnitude of the unstable estimates declined quickly ($\hat{\gamma}_i < 0$) as the overall event rate increased (e.g., at 20%).

Next, we summarize results for estimation of the provider-specific SRRs. As describe in Section 2, SRR is the summary performance index for each provider used in practice which incorporates patient-level risk factors Z_{ij} and their estimated effects, $\hat{\beta}$. More specifically, given the provider and the patient case-mix effect estimates for each approach, denoted by $\hat{\gamma}_i^*$ and $\hat{\beta}$, respectively, the estimated SRR for provider i is

$$SRR_i^* = \frac{\sum_{j=1}^{N_i} \hat{p}_{ij}^*}{\sum_{j=1}^{N_i} \hat{p}_{M,ij}^*}, \quad (5)$$

where $\hat{p}_{ij}^* = g^{-1}(\hat{\gamma}_i^* + \hat{\beta}^T Z_{ij})$, $\hat{p}_{M,ij}^* = g^{-1}(\hat{\gamma}_M^* + \hat{\beta}^T Z_{ij})$, $*$ and denotes the uncorrected and corrected approach, namely U and C . Figure 3 (left column) summarizes the uncorrected

FE model estimates of SRR for 3% to 20% overall outcome event. We note that even though specific $\gamma_i < 0$ were unstable for highly sparse data (e.g., at 3% - 10%; Figure 2), corresponding estimates of SRR's are stable overall and targets the true SRR, because SRR incorporates patients characteristics, their effects, as well as provider-specific effects as shown in (5); see Figure 3 (left column). Average SRR estimates for the corrected estimation performed well and are summarized in Figure 3 (right column). However, we note that for extremely sparse data (e.g., at 3%), the uncorrected approach slightly overestimate SRRs while the corrected approach slightly underestimate SRR for truly worse providers (true SRR > 1 ; Figure 4 - top). For truly better providers (true SRR < 1), both methods slightly over estimate the true SRRs, although more so with the corrected method. Differences in SRR estimates between the two methods are negligible as the overall percent of events increases (e.g., at 20%; Figure 4 - bottom).

5.2. Flagging Extreme Providers/Facilities

The overall performance of the uncorrected and corrected FE methods to identify extreme providers are assessed in terms of sensitivity (SEN) to correctly identify providers that under-perform (W: "worse"), over-perform (B: "better") relative to the reference standard (e.g., national reference), and specificity. Specificity (SPEC) refers to the correct identification/flagging of providers whose performances are not different from the reference standard (ND: "not different"). We note that provider assessment policies in practice focus on identifying under-performing providers (W providers) as those are tied to payment policy or regulatory goals. Figure 5 summarizes the distribution of SEN-W, SEN-B, and SPEC for varying levels of outcome sparsity, ranging from 3% to 20% overall outcome rate. For extremely sparse data of 3% and 5%, the uncorrected method has highest sensitivity to detect under-performing providers (higher SEN-W; left column). This is expected since the for truly worse providers, there are more outcome events ($\sum_j y_{ij}$); see Figure 5 (left column). SEN-W rates were similar between uncorrected and corrected methods at 20% overall overall outcome rate.

Because the event counts are zero or low for truly better providers in the context of sparse outcome data, the unstable/poor estimation of provider effects from the uncorrected method results in lower sensitivity to detect over-performing providers (lower SEN-B) compared to the corrected method (Figure 5 - middle column). However, note that the nominal SEN-B rates are low overall, as expected, compared to nominal SEN-W rates. This is expected in the low information context since B providers would have fewer readmissions, making it difficult to correctly identify B providers when the outcome is sparse. SPEC rates were high and similar between uncorrected and corrected methods (Figure 5 - right column).

As mentioned earlier, the main current objective of flagging "extreme" providers in profiling analysis focuses on *identifying W providers and ND providers*. Providers that over-perform (B providers) are not relevant to current payment policy or regulatory objectives. Therefore, under this regime, it is of interest to ensure that there are no (or low rate of) false negatives that misclassify/flag B provider as W provider ($FN_{B \rightarrow W}$). Indeed, there are none, i.e., $FN_{B \rightarrow W} = 0$ across all levels of data sparsity (Figure 6), which is not surprising since W and B providers are on the opposite tails of the distribution of providers. This is true with

the uncorrected FE model (as well as the corrected estimation method) since the direction of unstable estimates of γ_i 's are in the same (negative) direction of true γ_i (as pointed out earlier), despite the unstable provider-specific estimates. However, it is not uncommon for false negative classification of a B provider as a ND provider ($FN_{B \rightarrow ND}$). Although $FN_{B \rightarrow ND}$ decreases with increasing percentage of overall outcome events as expected, $FN_{B \rightarrow ND}$ is common for the extremely low information context (e.g., 3%, 5% overall event rate; Figure 6). We emphasize that high $FN_{B \rightarrow ND}$ does not affect current public policy because over-performers are not incentivized and are considered "ND" providers anyway. Therefore, the FE profiling model, even uncorrected, is still useful in the low information context with respect to the current public policy goal of identifying W and ND providers. However, if the public policy goal evolves to also incentivize for better performance, then novel methods able to correctly identify B providers with high sensitivity are needed.

6. DISCUSSION

Seminal works by Kalbfleisch and Wolfe [8] and He *et al.* [15] show that FE model estimates have smaller error for outlier providers whose effects are exceptionally large or small, and these extreme providers are precisely the ones we wish to identify in profiling analysis. The high-dimensional FE models were then used to assess the performance of dialysis facilities (providers) with respect to all-cause hospital readmissions which are frequent outcomes in dialysis patients. Subsequently, our own works have elucidated several operating characteristics [9, 10] of the FE profiling models and have been applied to assess the performance of dialysis facilities with respect to all-cause 30-day readmissions [11, 12, 14]. However, to date there is no work that examines the performance of FE models in the low information context where the outcome is sparse. The current study starts to fill this gap in knowledge. Several findings from this study have important practical impact in the low information context. First, even though the provider-specific estimates with true $\gamma_i < 0$ (truly B providers) are unstable, they are in the same direction as the true effects and the instability has moderated effects on the estimation of SRRs; i.e., SRRs are reasonably well-estimated and are the relevant quantities used in practice as they incorporate patient case-mix. However, if the provider-specific estimates, $\hat{\gamma}_i$'s, are themselves of interest, then our proposed correction method can be used to provide better estimates, especially corresponding to uncorrected $\hat{\gamma}_i$ that are substantially less than zero. Second, the consequence of sparse outcome data impacts more directly inference for B providers because true over-performers are the ones that contribute no or few events (readmissions); however, this "deficit" in estimation does not greatly impact the identification of W providers/under-performers and ND providers, which is the current focus of profiling in practice. Development of novel methods that have better sensitivity for flagging B providers would be useful when public policies or regulatory goals incorporate an incentive for over-performers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

This study was supported by a National Institute of Diabetes and Digestive and Kidney Diseases grant R01DK092232. The interpretation/reporting of the results presented are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the U.S. government.

APPENDIX

Provider-Specific Modified Score Equation Penalties

We describe details of calculating the penalties for the Firth’s modified score equations, adapted for the high-dimensional FE profiling model in Section 3.2. For a fixed β , the hat matrix used in the estimation of γ with Firth’s correction is $H = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}$ where $W = W_1 \oplus \dots \oplus W_F$, $X = X_1 \oplus \dots \oplus X_F$, $W_j = \text{diag}\{p_{j1}(1-p_{j1}), \dots, p_{jN_j}(1-p_{jN_j})\}$ are provider-specific weight matrices (see Section 3.2). Direct calculation yields

$$\begin{aligned} H &= W^{1/2}X(X^TWX)^{-1}X^TW^{1/2} \\ &= W^{1/2}X\{(X_1 \oplus \dots \oplus X_F)^T(W_1 \oplus \dots \oplus W_F)(X_1 \oplus \dots \oplus X_F)\}^{-1}X^TW^{1/2} \\ &= W^{1/2}X\{X_1^TW_1X_1 \oplus \dots \oplus X_F^TW_FX_F\}^{-1}X^TW^{1/2} \\ &= W^{1/2}X\{(X_1^TW_1X_1)^{-1} \oplus \dots \oplus (X_F^TW_FX_F)^{-1}\}X^TW^{1/2} \\ &= (W_1^{1/2}X_1 \oplus \dots \oplus W_F^{1/2}X_F)\{(X_1^TW_1X_1)^{-1} \oplus \dots \oplus (X_F^TW_FX_F)^{-1}\}X^TW^{1/2} \\ &= \{W_1^{1/2}X_1(X_1^TW_1X_1)^{-1} \oplus \dots \oplus W_F^{1/2}X_F(X_F^TW_FX_F)^{-1}\}X^TW^{1/2} \\ &= \{W_1^{1/2}X_1(X_1^TW_1X_1)^{-1} \oplus \dots \oplus W_F^{1/2}X_F(X_F^TW_FX_F)^{-1}\}(X_1^TW^{1/2} \oplus \dots \oplus X_F^TW^{1/2}) \\ &= W_1^{1/2}X_1(X_1^TW_1X_1)^{-1}X_1^TW^{1/2} \oplus \dots \oplus W_F^{1/2}X_F(X_F^TW_FX_F)^{-1}X_F^TW^{1/2}. \end{aligned}$$

Thus,

$$\text{diag}(H) = [\text{diag}\{W_1^{1/2}X_1(X_1^TW_1X_1)^{-1}X_1^TW^{1/2}\}, \dots, \text{diag}\{W_F^{1/2}X_F(X_F^TW_FX_F)^{-1}X_F^TW^{1/2}\}].$$

Dependence Structure of Covariates in Simulation Model

The correlation matrix and the standard deviation of the patient case-mix variables, Z_{ij1}, \dots, Z_{ij15} , are summarized in Table 1.

Table 1:

Correlation Matrix of Z_1, \dots, Z_{15} and their Standard Deviations

Correlation														
Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}
	-0.06	-0.01	-0.31	0.04	0.08	-0.12	-0.05	0.28	-0.04	-0.13	-0.22	-0.06	0.02	-0.01
	1	0.03	-0.16	0.20	0.33	-0.05	0.08	0.07	0.01	-0.04	-0.05	-0.04	-0.05	0.00
		1	0.00	0.03	0.03	0.00	0.10	0.00	0.07	0.08	0.01	0.02	0.02	0.10
			1	-0.01	-0.67	0.00	0.09	-0.03	0.11	0.17	0.04	0.01	0.04	0.04
				1	0.06	0.00	-0.10	0.02	0.05	0.02	-0.06	-0.09	0.02	0.01
					1	-0.07	0.21	-0.03	-0.01	-0.12	-0.04	-0.02	-0.06	0.01

Correlation														
Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉	Z ₁₀	Z ₁₁	Z ₁₂	Z ₁₃	Z ₁₄	Z ₁₅
						1	-0.02	-0.03	0.00	0.07	0.06	0.25	0.02	-0.02
							1	0.06	0.13	0.09	0.06	0.02	-0.01	0.08
								1	0.48	0.15	0.11	0.06	0.18	0.06
									1	0.33	0.17	0.07	0.19	0.14
										1	0.21	0.29	0.10	0.10
											1	0.34	0.03	0.03
												1	0.02	0.01
													1	0.01
														1
Standard Deviation														
1.59	1.03	0.87	1.43	0.50	0.49	0.10	0.24	0.41	0.36	0.32	0.20	0.19	0.16	0.19

LIST OF ABBREVIATIONS

CMS	Centers for Medicare and Medicaid Services
FE	Fixed effects
US	United States
USRDS	United States Renal Data System
RE	Random effects
SRR	Standardized readmission ratio
W	Worse
ND	Not different
B	Better
SEN	Sensitivity
SPEC	Specificity

REFERENCES

- [1]. United States Renal Data System. Annual Data Report: Epidemiology of Kidney Disease and in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD. [cited 2020]: Available from <https://adr.usrds.org/2020/>
- [2]. Ash AS, Fienberg SE, Louis TA, Normand ST, Stukel TA, Utts J. Statistical issues in assessing hospital performance. The COPSS-CMS White Paper Committee, CMS, Washington D.C. [cited 2012]: Available from <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/statistical-issues-in-assessing-hospital-performance.pdf>.

- [3]. Centers for Medicare & Medicaid Services (CMS)/UM-KECC. Report for the standardized readmission ratio. [cited 2017]; Available from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/SRR_Methodology_Report_June2017.pdf
- [4]. Horwitz L, Partovain C, Lin ZQ, et al. Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. *Annals of Internal Medicine* 2014; 161: S66–75. 10.7326/M13-3000 [PubMed: 25402406]
- [5]. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circulation Cardiovascular Quality and Outcomes* 2011; 4: 243–252. 10.1161/CIRCOUTCOMES.110.957498 [PubMed: 21406673]
- [6]. Normand S, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association* 1997; 92: 803–814. 10.1080/01621459.1997.10474036
- [7]. Normand ST, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Statistical Science* 2007; 22: 206–226. 10.1214/088342307000000096
- [8]. Kalbfleisch JD, Wolfe RA. On monitoring outcomes of medical providers. *Statistics in Biosciences* 20113; 5: 286–302. 10.1007/s12561-013-9093-x
- [9]. Chen Y, Senturk D, Estes JP, Campos LF, Rhee CM, Dalrymple LS, Kalantar-Zadeh K, Nguyen DV. Performance characteristics of profiling methods and the impact of inadequate case-mix adjustment. *Communications in Statistics - Simulation and Computation* 2021; 50: 1854–1871. 10.1080/03610918.2019.1595649
- [10]. Senturk D, Chen Y, Estes JP, Campos LF, Rhee CM, Kalantar-Zadeh K, Nguyen DV. Impact of case-mix measurement error on estimation and inference in profiling of health care providers. *Communications in Statistics - Simulation and Computation* 2020; 49: 2206–2224. 10.1080/03610918.2018.1515360 [PubMed: 33311842]
- [11]. Estes JP, Chen Y, Senturk D, Rhee CM, Kurum E, You AS, Streja E, Kalantar-Zadeh K, Nguyen DV. Profiling dialysis facilities for adverse recurrent events. *Statistics in Medicine* 2020; 39: 1374–1389. 10.1002/sim.8482 [PubMed: 31997372]
- [12]. Estes JP, Nguyen DV, Chen Y, Dalrymple LS, Rhee CM, Kalantar-Zadeh K, Senturk D. Time-dynamic profiling with application to hospital readmission among patients on dialysis (with discussion). *Biometrics* 2018; 74: 1383–1394. 10.1111/biom.12908 [PubMed: 29870064]
- [13]. Estes JP, Nguyen DV, Chen Y, Dalrymple LS, Rhee CM, Kalantar-Zadeh K, Senturk D. Rejoinder: Time-dynamic profiling with application to hospital readmission among patients on dialysis. *Biometrics* 2018b; 74: 1404–1406. 10.1111/biom.12905 [PubMed: 29870066]
- [14]. Chen Y, Rhee CM, Senturk D, Kurum E, Campos LF, Li Y, Kalantar-Zadeh K, Nguyen DV. Association of U.S. dialysis facility staffing with profiling of hospital-wide 30-day unplanned readmission. *Kidney Diseases* 2019; 5: 153–162. 10.1159/000496147 [PubMed: 31259177]
- [15]. He K, Kalbfleisch JD, Li Y, Li Y. Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime Data Analysis* 2013; 19: 490–512. 10.1007/s10985-013-9264-6 [PubMed: 23709309]
- [16]. Firth D Bias reduction of maximum likelihood estimates. *Biometrika* 1993; 80: 27–38. 10.1093/biomet/80.1.27
- [17]. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Statistics in Medicine* 2002; 21: 2409–2419. 10.1002/sim.1047 [PubMed: 12210625]
- [18]. Jeffreys H An invariant form of the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A* 1946; 186(1007): 453–461. 10.1098/rspa.1946.0056

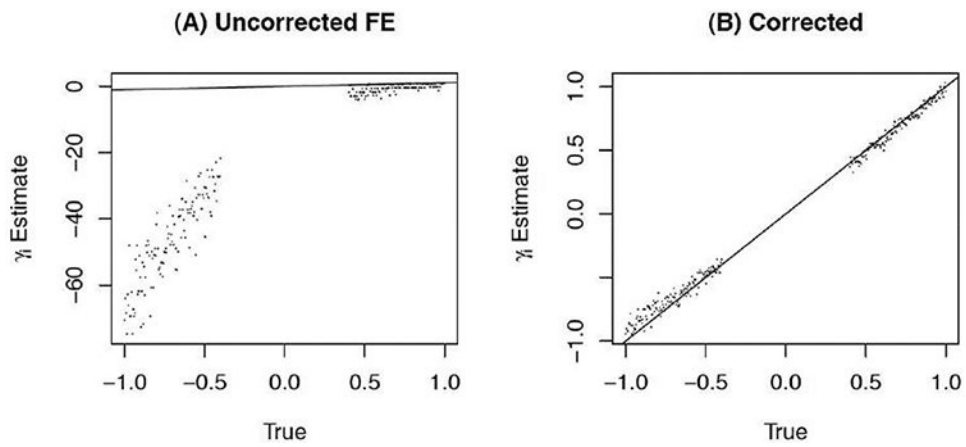


Figure 1: Estimates of provider-specific effects, $\gamma_i < 0$ (over-performers) and $\gamma_i > 0$ (under-performers) (A) for the uncorrected high-dimensional fixed effects (FE) model and (B) for the corrected method at high-level of outcome data sparsity of 3%. Displayed is average for each γ_i estimate, averaged over 200 simulated data sets.

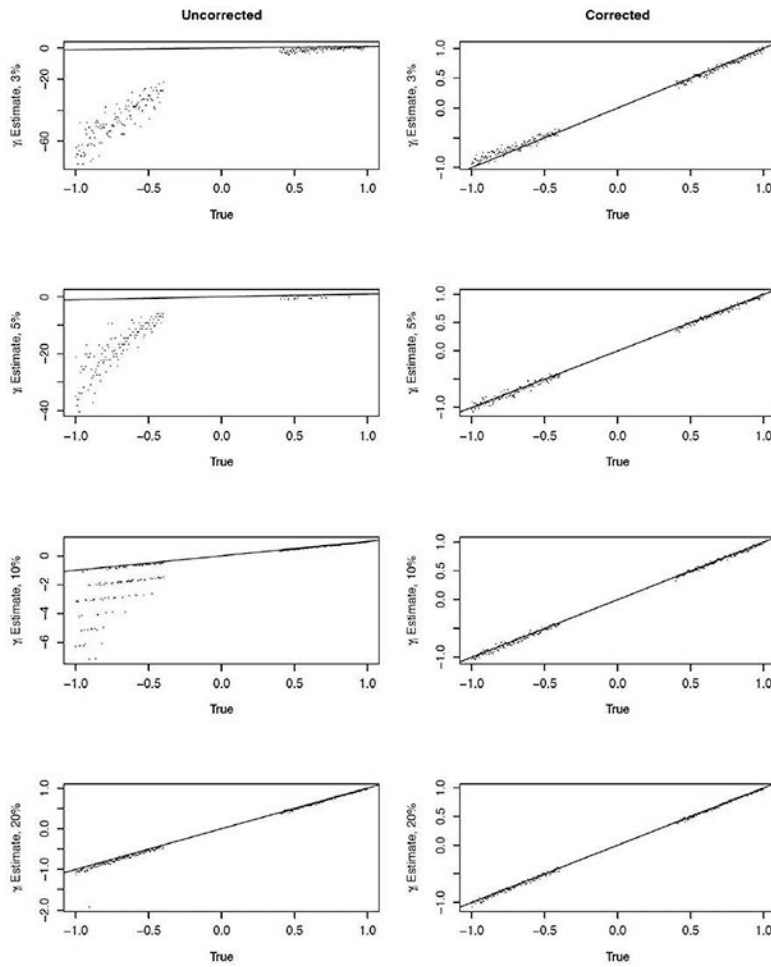


Figure 2: Uncorrected (left column) and corrected (right column) estimation of provider-specific effects, γ_i 's, for 3%, 5%, 10%, and 20% overall outcome event rate. Displayed is average for each γ_i estimate, averaged over 200 simulated data sets.

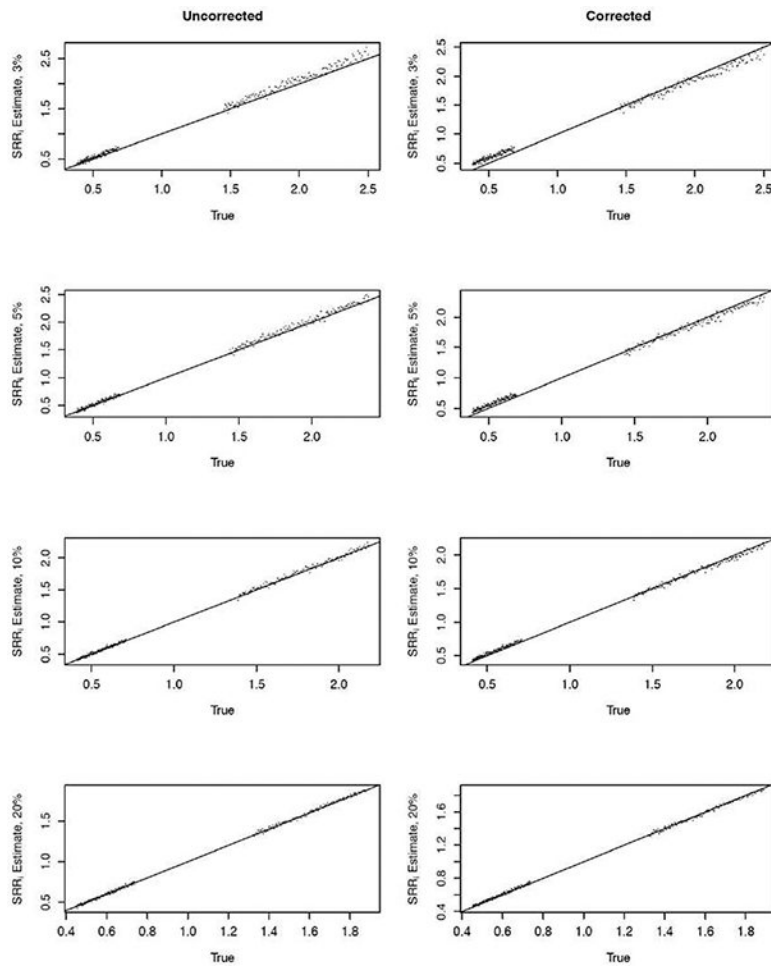


Figure 3: Uncorrected (left column) and corrected (right column) estimates of standardized readmission ratios (SRRs) for 3%, 5%, 10%, and 20% overall outcome event rate. Displayed is average for each *SRR*_{*i*} estimate, averaged over 200 simulated data sets.

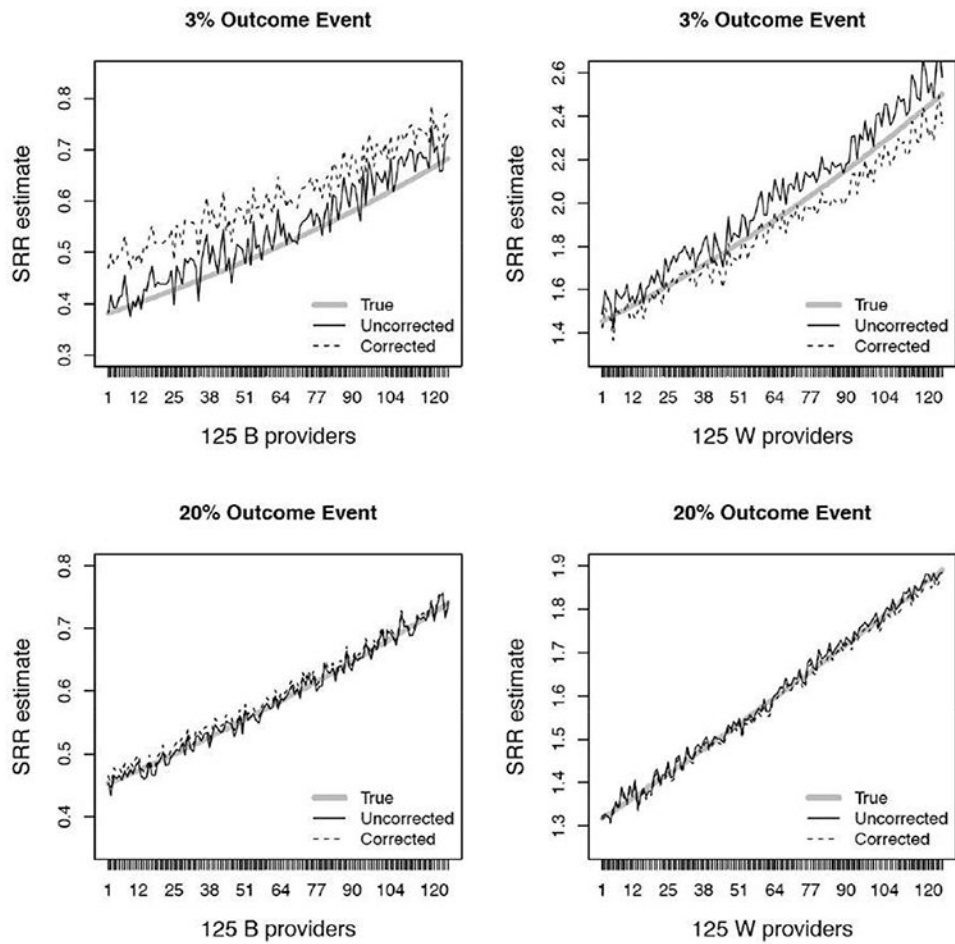


Figure 4: Estimation of standardized readmission ratios (SRRs) for 3% and 20% overall outcome event rate for corrected and uncorrected methods among the 125 better (B) and 125 worse (W) providers. Displayed is average for each SRR_i estimate, averaged over 200 simulated data sets.

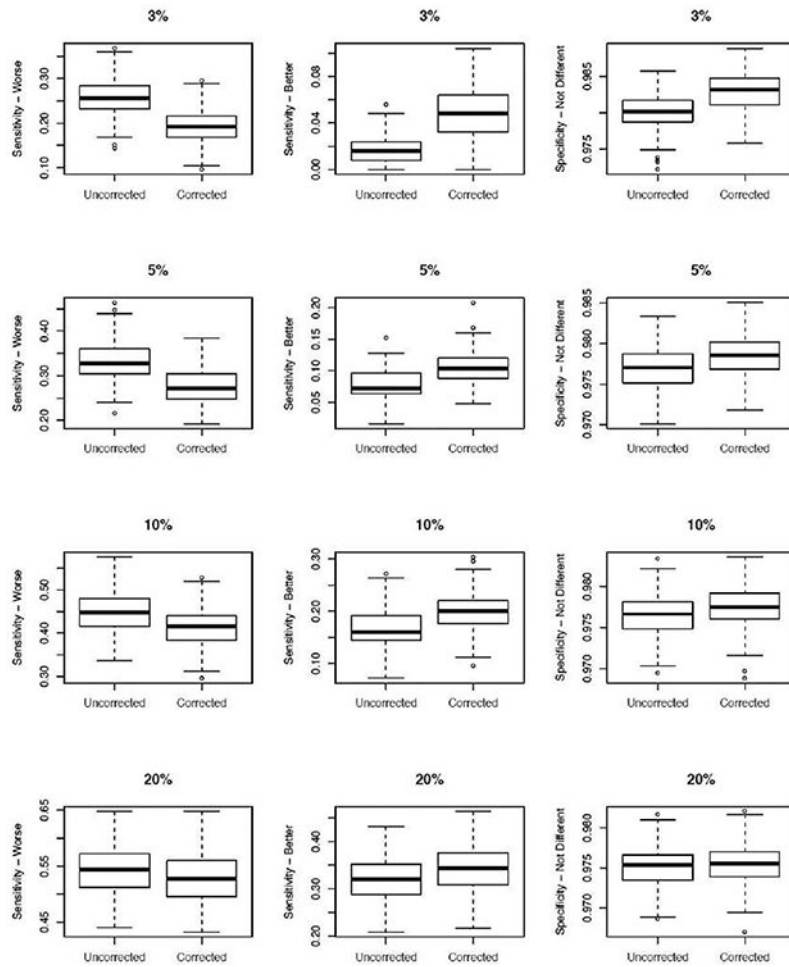


Figure 5: Overall performance of the uncorrected and corrected estimation methods to identify truly worse (sensitivity - worse), truly better (sensitivity - better), and specificity (providers not different from the reference) across data sparsity of 3%, 5%, 10%, and 20% overall outcome event rate. Displayed is average for each SRR_i estimate, averaged over 200 simulated data sets.

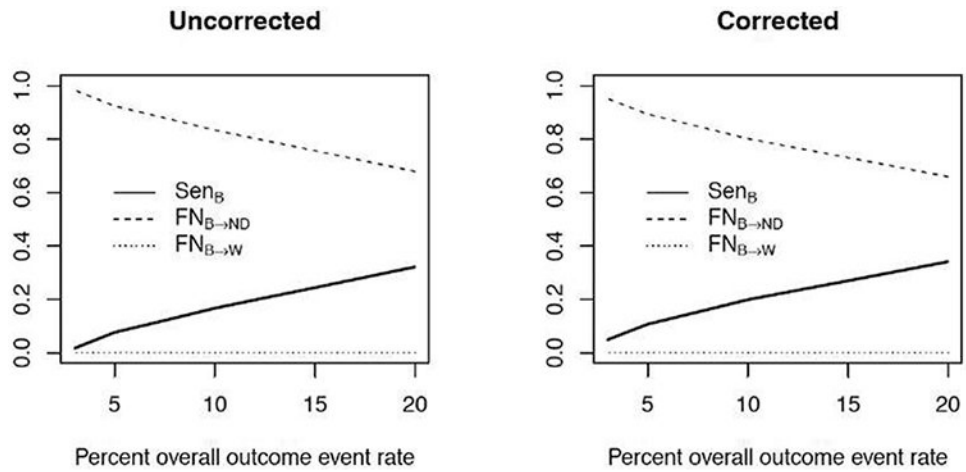


Figure 6: Rate of false negative (FN) for incorrectly flagging better (B) providers as worse (W) providers ($FN_{B \rightarrow W}$) and incorrectly flagging B providers as providers not different (ND) from the reference ($FN_{B \rightarrow ND}$) for the uncorrected and corrected estimation methods across data sparsity of 3%, 5%, 10%, and 20% overall outcome event rate. Displayed is average rates, averaged over 200 simulated data sets.