Supplemental Materials

Performance Characteristics of Profiling Methods and the Impact of Inadequate Case-mix Adjustment

Yanjun Chen[1], Damla Şentürk[2], Jason P. Estes[3], Luis F. Campos[4], Connie M. Rhee[5], Lorien S. Dalrymple[6], Kamyar Kalantar-Zadeh[5] and Danh V. Nguyen[7] *

[1]Institute for Clinical and Translational Science, University of California, Irvine, CA 92687, U.S.A.
[2]Department of Biostatistics, University of California, Los Angeles, CA 90095, U.S.A.
[3]Research, Pratt & Whitney, East Hartford, CT 06118, U.S.A.
[4]Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.
[5]Harold Simmons Center for Chronic Disease Research and Epidemiology, University of California Irvine School of Medicine, Orange, CA 92868, U.S.A.
[6]Epidemiology and Research, Fresenius Medical Care, Waltham, MA 02451, U.S.A.
[7]Department of Medicine, University of California Irvine, Orange, CA 92868, U.S.A.

*Corresponding author. Email: danhvn1@uci.edu

1 Analysis Example and R Codes

This supplemental materials document provides details on the FE flagging procedures, based on nominal p-values and empirical null adjustment, and the RE (CMS) bootstrap procedure. In addition, a sample dataset with 1,000 providers and a tutorial on the implementation along with R codes are provided for both the RE and FE methods here. The example (simulated) dataset containing 1,000 providers and 15 patient case-mix along with R functions and documentation for fitting the RE and FE models can be downloaded at `www.http://faculty.sites.uci.edu/nguyen/supplement/`. Note that the input dataset is sorted by provider ID's (fid).

1.1 Fitting RE and FE Models

The following usages fit the FE and RE/CMS models, respectively.

```
fit.FE.m1(ds, xvar.names, starting.val, numBSRuns)
fit.RE.CMS(ds, xvar.names, numBSRuns)
```

The input arguments, `ds` and `xvar.names` are the dataset name and list of patient case-mix covariate names, respectively. For the FE model, `starting.val` is the starting value for γ_i in the Newton-Raphson estimation. This can be taken to be the baseline readmission rate, for instance. The argument, `numBSRuns` is the number of resamples for testing the hypothesis $H_0 : \gamma_i = \gamma_M$, which is taken to be 500. For the RE/CMS model, `numBSRuns` is the number of bootstrap samples to construct the 95% confidence interval for each provider; also taken to be 500. Details are provided in the R script file (`analysis_script.R`) at `www.http://faculty.sites.uci.edu/nguyen/supplement/` .

1.2 Output Objects from Fitted Models

The output object from each fitted model has two elements, the provider-level results and the patient-level case-mix coefficient estimates. For the FE model, the two elements are:

```
> names(fit.FE)
[1] "FE.result" "betaEstF"
```

The `FE.result` contains provider-level results ($F \times 5$) and `betaEstF` contains the coefficient estimates vector $\widehat{\boldsymbol{\beta}}$ ($r \times 1$). The columns of `FE.result` are

1. `sRREstsF`: SRR_i estimate, using the sum of readmissions as the numerator of SRR.
2. `sRREstsF`: SRR_i estimate, using the sum of estimated probabilities of readmission as the numerator SRR.
3. `pValsF`: P-values from hypothesis test of $H_0 : \gamma_i = \gamma_M$, $i = 1, 2, \ldots, F$.
4. `gammaEstF`: $\widehat{\gamma}_i$ estimate for each provider.
5. `SRR.category.F.emp`: Flagging indicator based on the empirical null distribution (ND: not different, B: better, W: worse).

```
> head(fit.FE$FE.result)
   sRREstsF sRREstsF2 pValsF gammaEstF SRR.category.F.emp
1 0.9368067 0.9368075  0.294 -1.465062                 ND
2 0.8197385 0.8197400  0.000 -2.058169                  B
3 0.8414432 0.8414448  0.020 -2.068698                 ND
4 0.7378985 0.7379005  0.000 -2.416828                  B
5 0.7887469 0.7887485  0.010 -1.918674                 ND
6 0.8696912 0.8696927  0.044 -1.639124                 ND
```

```
> head(fit.FE$betaEstF)
        [,1]
z1 0.5135744
z2 0.5241660
z3 0.5052750
z4 0.5145239
z5 0.5090906
z6 0.5069612
```

The output for the RE/CMS fitted model are similar, with two elements.

```
> names(fit.RE)
[1] "RE.result" "betaEstC"
```

The second element (`betaEstC`) contain the coefficient estimates vector $\widehat{\boldsymbol{\beta}}$ ($r \times 1$). The provider-level output, `RE.result`, consists of the five columns:

1. `sRREstsc`: SRR_i estimate from the RE/CMS model.

2. `sRREstsC_Lower`: Lower bound of the 95% CI for $\widetilde{SRR}_i$ via 500 bootstrap samples.

3. `sRREstsC_Upper`: Upper bound of the 95% CI for $\widetilde{SRR}_i$.

4. `BetterC`: Indicator of providers flagged as "better" (1:yes; 0:no).

5. `WorseC`: Indicator of providers flagged as "worse" (1:yes; 0:no).black

```
> head(fit.RE$RE.result)
  sRREstsC sRREstsC_Lower sRREstsC_Upper BetterC WorseC
1 0.9803385      0.9149993      1.0566237       0      0
2 0.9034610      0.8284067      0.9758525       1      0
3 0.9456762      0.8741779      1.0063479       0      0
4 0.8969990      0.8147081      0.9803536       1      0
5 0.9232880      0.8105824      1.0432741       0      0
6 0.9386103      0.8620731      1.0212551       0      0

> head(fit.RE$betaEstC)
[1] -0.9665475  0.5007279  0.5103138  0.4921882  0.5007380  0.4949895
```

2 Appendix

2.1 RE/CMS Bootstrap Confidence Interval Estimation Procedure

We summarize here FE and RE/CMS procedures for identifying under- (over-) performing providers ("outlier" providers). The RE procedure is based on the following bootstrap CI approach, which we summarize here based on Horwitz et al. (2011) and Ash et al. (2012).

0. Fit the generalized linear mixed effects model (1), i.e., the RE model. Denote the provider-specific estimates by $\widehat{\gamma}_i$, $i = 1, 2, \ldots, F$, with overall mean $\widehat{\gamma}_0$. Also, denote the variance and patient case-mix estimates by $\widehat{\sigma}^2$ and $\widehat{\boldsymbol{\beta}}$, respectively. Calculate estimated standardized readmission ratio, SRR_i as given by (3) for the RE model.

1. Generate a bootstrap dataset by sampling F providers with replacement from the original dataset. Denote the unique set of providers sampled by $\mathbb{F}^{(b)}$, where b indexes bootstrap dataset.

2. Fit the RE model (1) to the bootstrap dataset and treat each resampled provider as distinct. Calculate

   (a) The patient case-mix effects, $\widehat{\boldsymbol{\beta}}^{(b)}$.

   (b) The mean and variance of the distribution of provider effects, $\widehat{\gamma}_0^{(b)}$, and $\widehat{\sigma}^{2(b)}$.

   (c) The provider-specific effects and variances, $\{\widehat{\gamma}_i^{(b)}, \widehat{\mathrm{Var}}^{(b)}(\gamma_i)\}$, $i = 1, 2, \ldots, F$. (If a provider is sampled more than once, then randomly select one set of the provider-specific estimates and variances.)

3. Generate a provider random effect from the provider-specific distribution from step 2(c) for each unique provider sampled in step 1. The posterior distribution of each random effect is approximated by a normal distribution, $\widehat{\gamma}_i^{(b)*} \sim N(\widehat{\gamma}_i^{(b)}, \widehat{\mathrm{Var}}^{(b)}(\gamma_i))$.

4. Calculate SRR_i for each unique provider i sampled in step 1: $SRR_i^{(b)} = \sum_{j=1}^{n_i} \widehat{p}_{ij}^{(b)} / \sum_{j=1}^{n_i} \widehat{p}_{M,ij}^{(b)} = \sum_{j=1}^{n_i} g^{-1}(\widehat{\gamma}_i^{(b)*} + \widehat{\boldsymbol{\beta}}^{(b)^{\mathrm{T}}} \mathbf{Z}_{ij}) / \sum_{j=1}^{n_i} g^{-1}(\widehat{\gamma}_0^{(b)} + \widehat{\boldsymbol{\beta}}^{(b)^{\mathrm{T}}} \mathbf{Z}_{ij})$, for $i \in \mathbb{F}^{(b)}$.

5. Repeat bootstrap procedure, step 1 - step 4, 500 times ($b = 1, \ldots, 500$) and form the 95% confidence interval estimate of $\widetilde{SRR}_i$ for each provider $i = 1, 2, \ldots, F$.

2.2 FE Hypothesis Testing Procedure

The FE inference procedure is based on testing the hypothesis $H_0 : \gamma_i = \gamma_M$ (i.e., $\widetilde{SRR}_i = 1$) for the ith provider. For convenience, we summarize this procedure proposed by He et al. (2013) below. (For fitting the high-dimensional FE model using the iterative one-step Newton-Raphson algorithm, see He et al. (2013) for details.)

1. Estimate FE model parameters and fix $\boldsymbol{\beta}$ and γ_M at their estimated values $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}_M$.

2. For the ith provider, draw $B = 500$ samples under the null hypothesis, $\{Y_{ij}^{(b)} : j = 1, 2, \ldots, n_i\}_{b=1}^{B}$, where each sample and observations are independently drawn from a

Bernoulli distribution: $Y_{ij}^{(b)} \sim Ber(\overline{p}_{ij})$, where $\overline{p}_{ij} = \exp(\widehat{\gamma}_M + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_{ij})/(1 + \exp(\widehat{\gamma}_M + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_{ij}))$

3. Calculate total number of readmissions in the resampled data: $Y_{i\cdot}^{(b)} = \sum_{j=1}^{n_i} Y_{ij}^{(b)}$.

4. Calculate the p-value for testing $H_0 : \gamma_i = \gamma_M$ as follows. Compute $SL_i^+ \equiv B^{-1}\sum_{b=1}^{B}[0.5\mathbb{I}(Y_{i\cdot}^{(b)} = O_i) + \mathbb{I}(Y_{i\cdot}^{(b)} > O_i)]$, where O_i is the observed number of readmissions for provider i in the original data and $\mathbb{I}()$ denotes the indicator function; similarly, compute $SL_i^- \equiv B^{-1}\sum_{b=1}^{B}[0.5\mathbb{I}(Y_{i\cdot}^{(b)} = O_i) + \mathbb{I}(Y_{i\cdot}^{(b)} < O_i)]$. The p-value is $P = 2 \times \min\{SL_i^+, SL_i^-\}$.

5. Repeat steps 2 - 4 for each provider, $i = 1, 2, \ldots, F$.

3 Supplementary Table and Figures

Table S1: Simulation settings, design parameters, and models.

Description	**Name**	**Parameters/summary**
Provider Effect Size (P-ES) *(Smaller P-ES)*	P-ES 1	2.5% W: $\gamma_i \sim U(0.4, 1.5)$; 2.5% B: $\gamma_i \sim -U(0.4, 1.5)$; 95% ND: $\gamma_i \sim N(0, 0.2^2)$
Provider Effect Size (P-ES) *(Larger P-ES)*	P-ES 2	2.5% W: $\gamma_i \sim U(0.6, 1.5)$; 2.5% B: $\gamma_i \sim -U(0.6, 1.5)$; 95% ND: $\gamma_i \sim N(0, 0.2^2)$
Case-mix Effect Size (CM-ES) *(Smaller CM-ES)*	CM-ES 1	$\boldsymbol{\beta}_A : \beta_1 = \cdots = \beta_{10} = 0.5; \beta_{11} = \beta_{15} = 1$ (15 covariates)
Case-mix Effect Size (CM-ES) *(Larger CM-ES)*	CM-ES 2	$\boldsymbol{\beta}_B = 2 \times \boldsymbol{\beta}_A$
CM correlation/dependence	DEP	Five cases: (1-4) $\rho = 0, 0.2, 0.5, 0.8$ for all variables; (5) general dependence/corr. structure
Baseline readmission rate	BRR	Low, medium, high: 14.3%, 27.3%, and 41.7%
Num. of simulation studies combinations		45: 3 (P-ES1+CM-ES1, P-ES1+CM-ES2, P-ES2+CM-ES1) $\times 5$ (DEP) $\times$ 3 (BRR)
Num. of providers Num. of datasets		1,000 per dataset 200 per combination (9,000 = 45 $\times$ 200 total)
RE and FE models fitted to each dataset *($2 \times 5 = 10$ models)*	$\mathcal{M}_0$	Intercept only (no case-mix adjustment)
	$\mathcal{M}_1$	Adjustment for $\{Z_1, Z_2\}$
	$\mathcal{M}_2$	Adjustment for $\{Z_1, \ldots, Z_{10}\}$
	$\mathcal{M}_3$:	Adjustment for $\{Z_{11}, \ldots, Z_{15}\}$
	$\mathcal{M}_f$	Full model with complete case-mix adjustment

$U(a, b)$ denotes uniform distribution; $N(\mu, \sigma^2)$ denotes normal distribution.
W = Worse, B = Better, and ND = not different relative to reference standard.

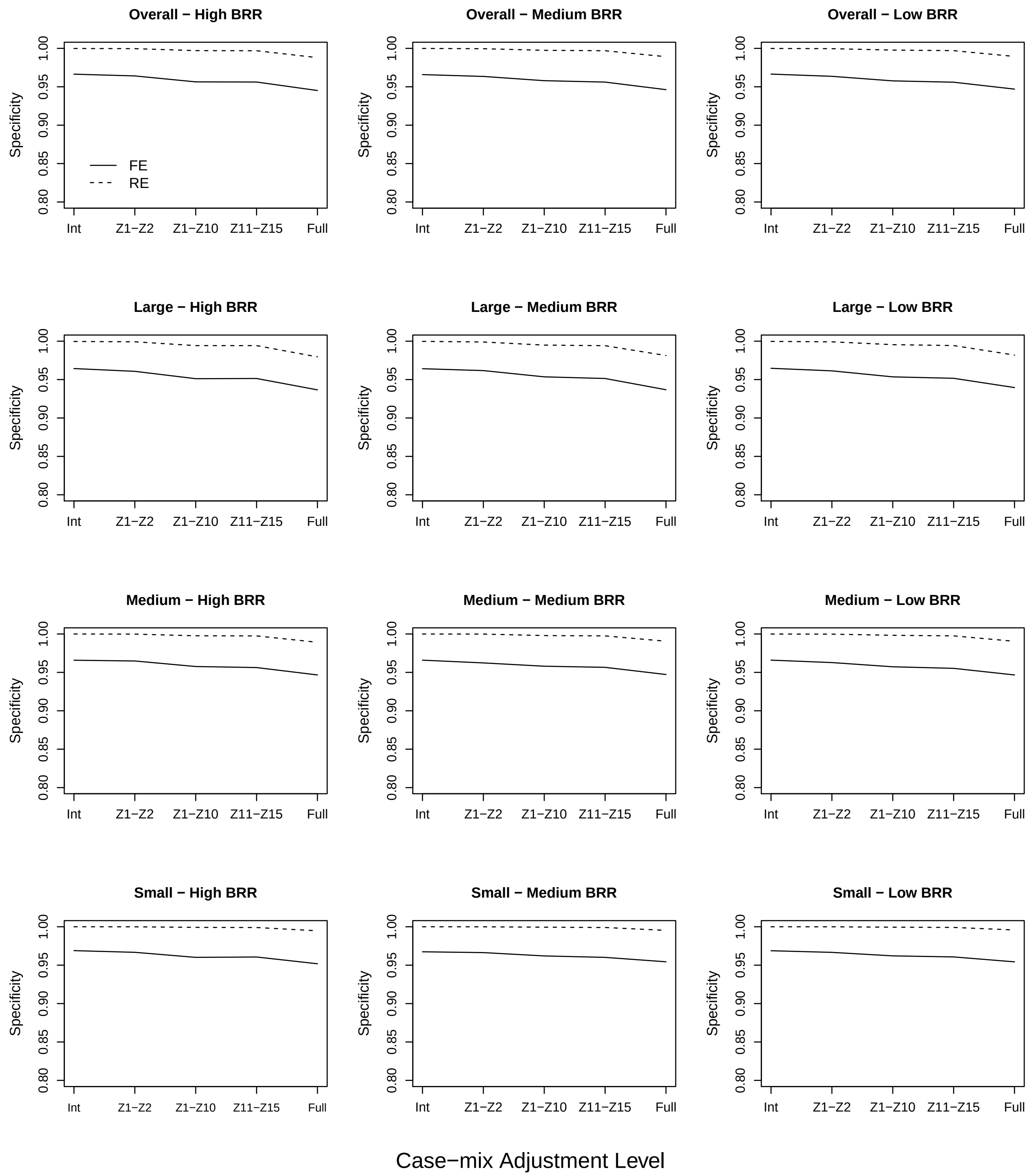

Figure S1: Impact of inadequate case-mix (CM) adjustment levels (Int: Intercept only, $Z_1 - Z_2$, $Z_1 - Z_{10}$, $Z_{11} - Z_{15}$, Full: Full model) on RE and FE models' specificity to identify providers not different from the national reference rate for CM correlation $\rho = 0.2$ across low, medium, and high baseline readmission rates (BRR). Results are presented across all providers (overall) and stratified by provider volume (large, medium, small).

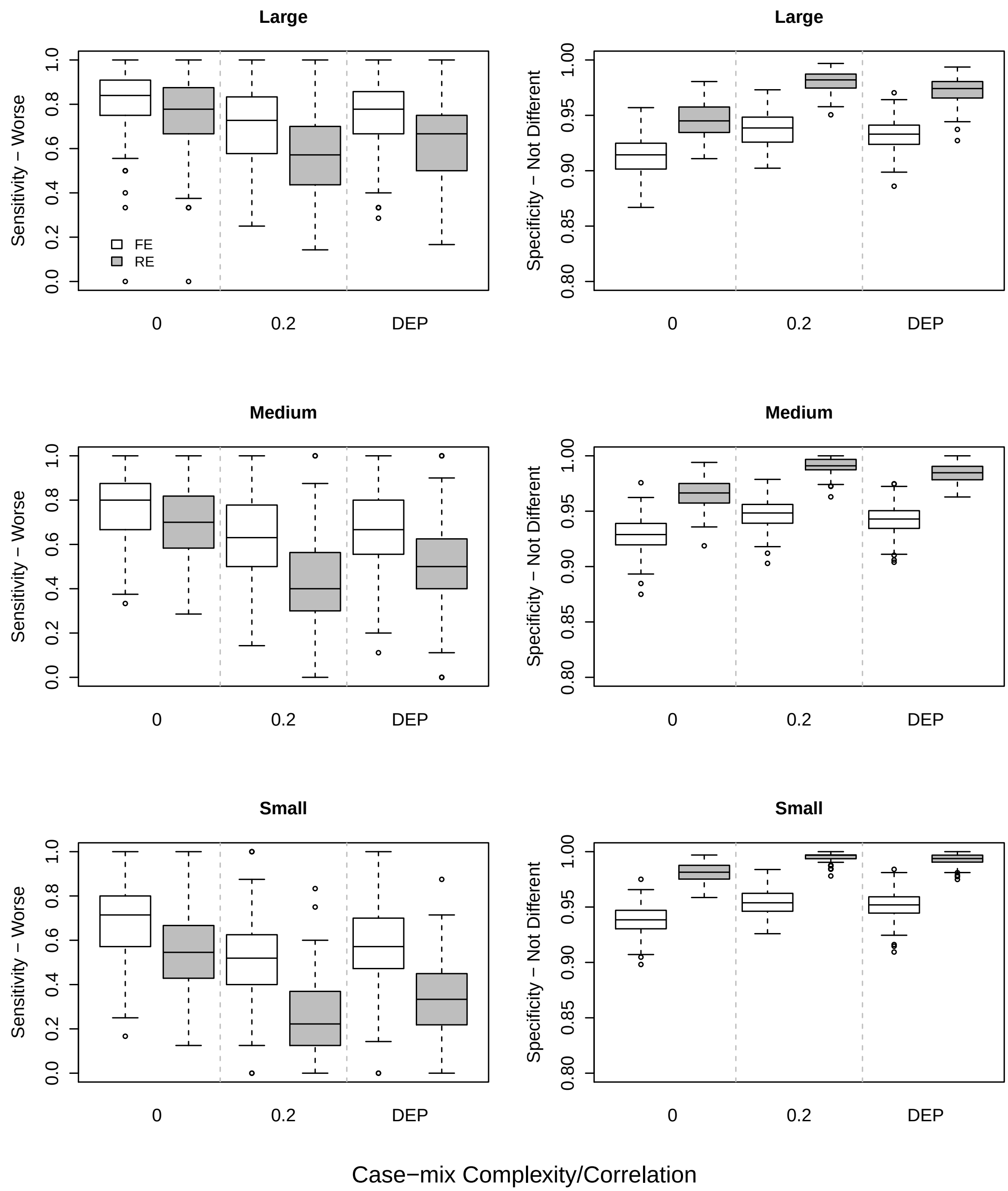

Figure S2: Overall performance of the full (benchmark) RE and FE models by provider volume/size: Small, medium, and large providers defined by tertiles: (Left) Sensitivity, (Right) Specificity. Data generated were from the more general dependence (DEP) structure with unequal correlation among case-mix risk variables. For reference, given also are results for the uncorrelated and equally correlated ($\rho = 0.2$) case-mix scenarios.

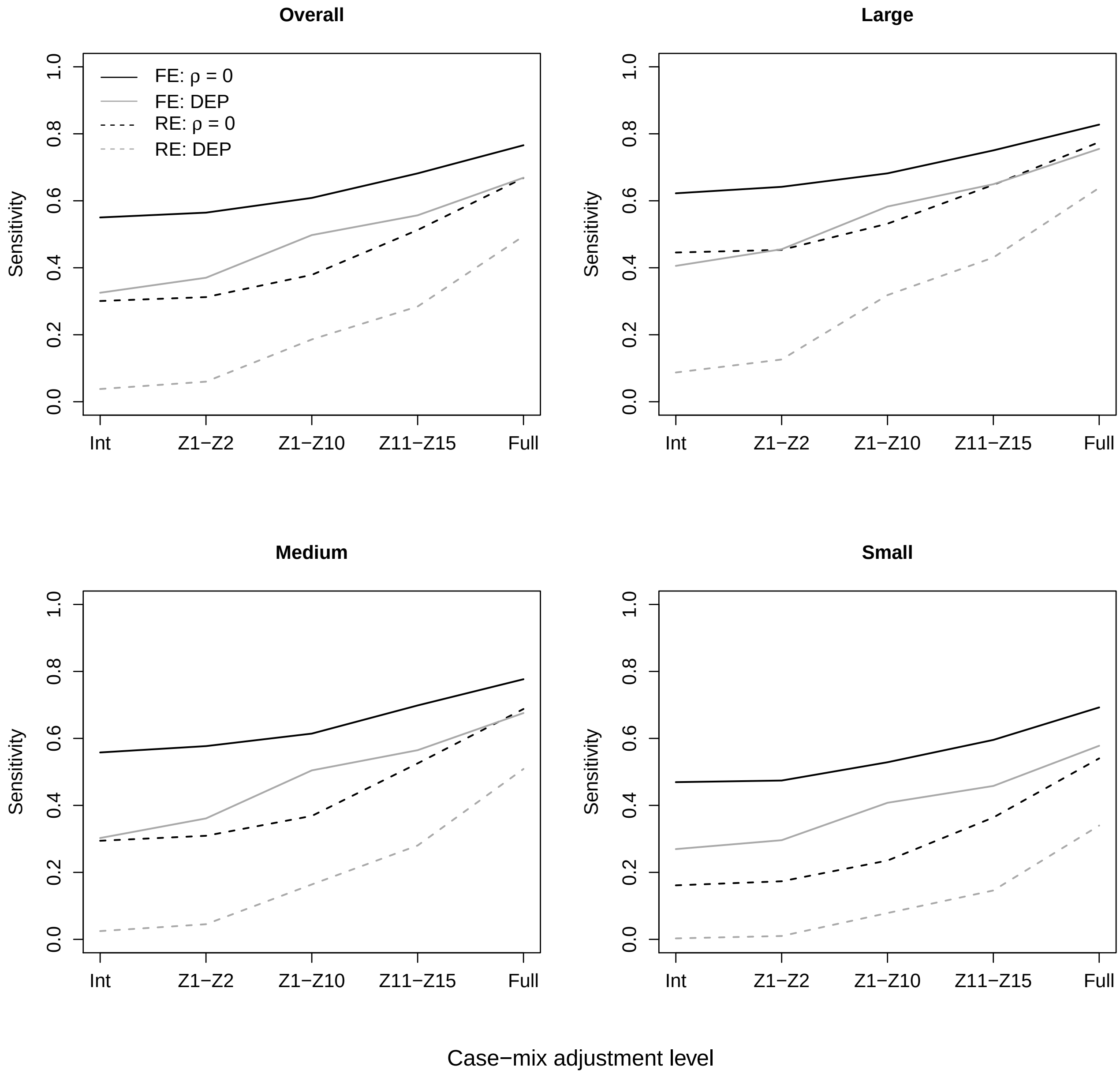

Figure S3: Impact of inadequate case-mix (CM) adjustment levels (Int: Intercept only, $Z_1 - Z_2$, $Z_1 - Z_{10}$, $Z_{11} - Z_{15}$, Full: Full model) on sensitivities of RE and FE models for the more general dependence (unequal) correlation structure (DEP) among case-mix variables. Given are sensitivities to detect under-performing providers for medium baseline readmission rates. Results are presented across all providers (overall) and stratified by provider volume (large, medium, small). The case of uncorrelated case-mix is provided as a reference.

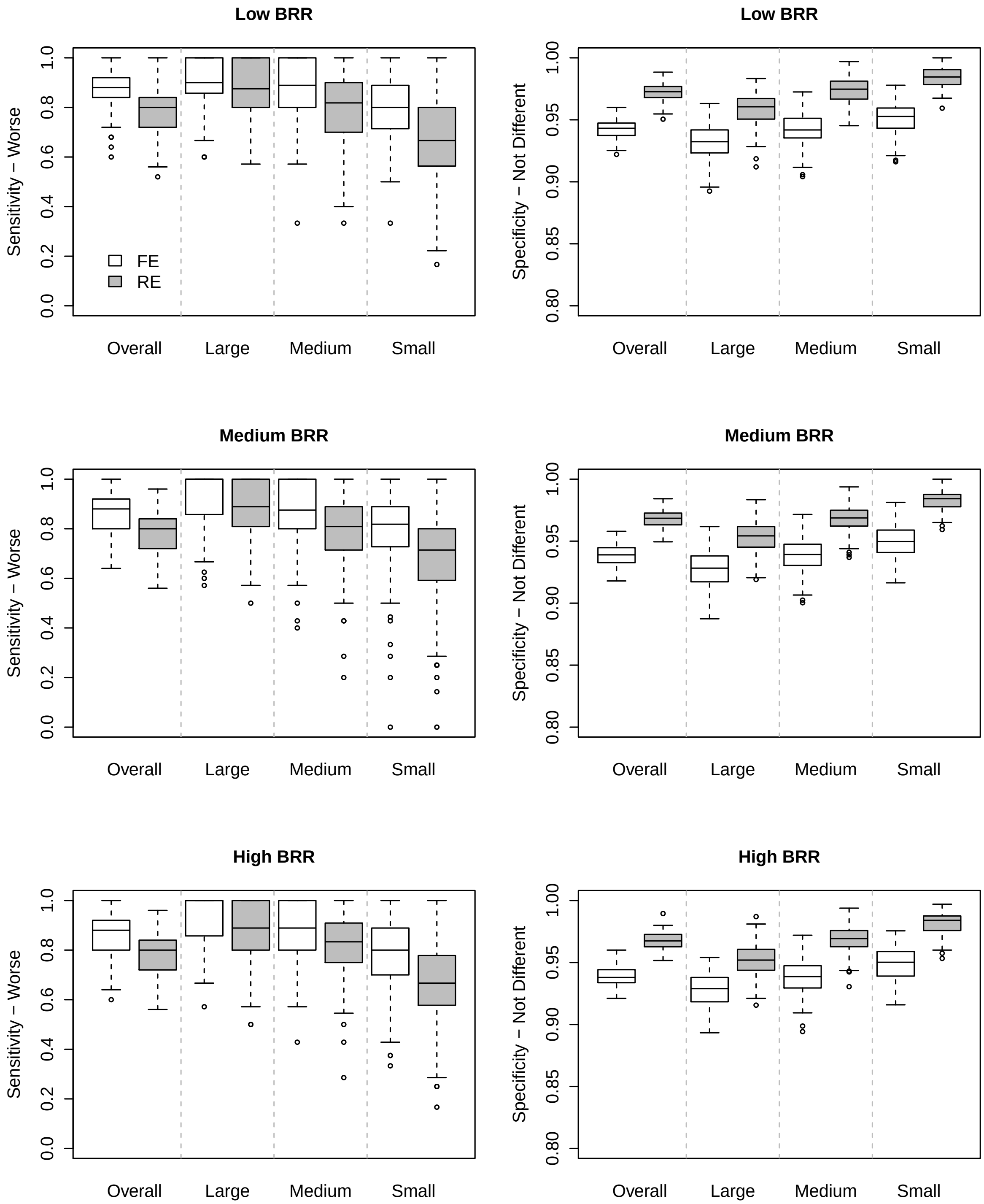

Figure S4: Overall performance of the full (benchmark) RE and FE models by provider volume (small, medium, and large) and by baseline readmission rates (BRR, low, medium and high) for simulation study based on USRDS data.

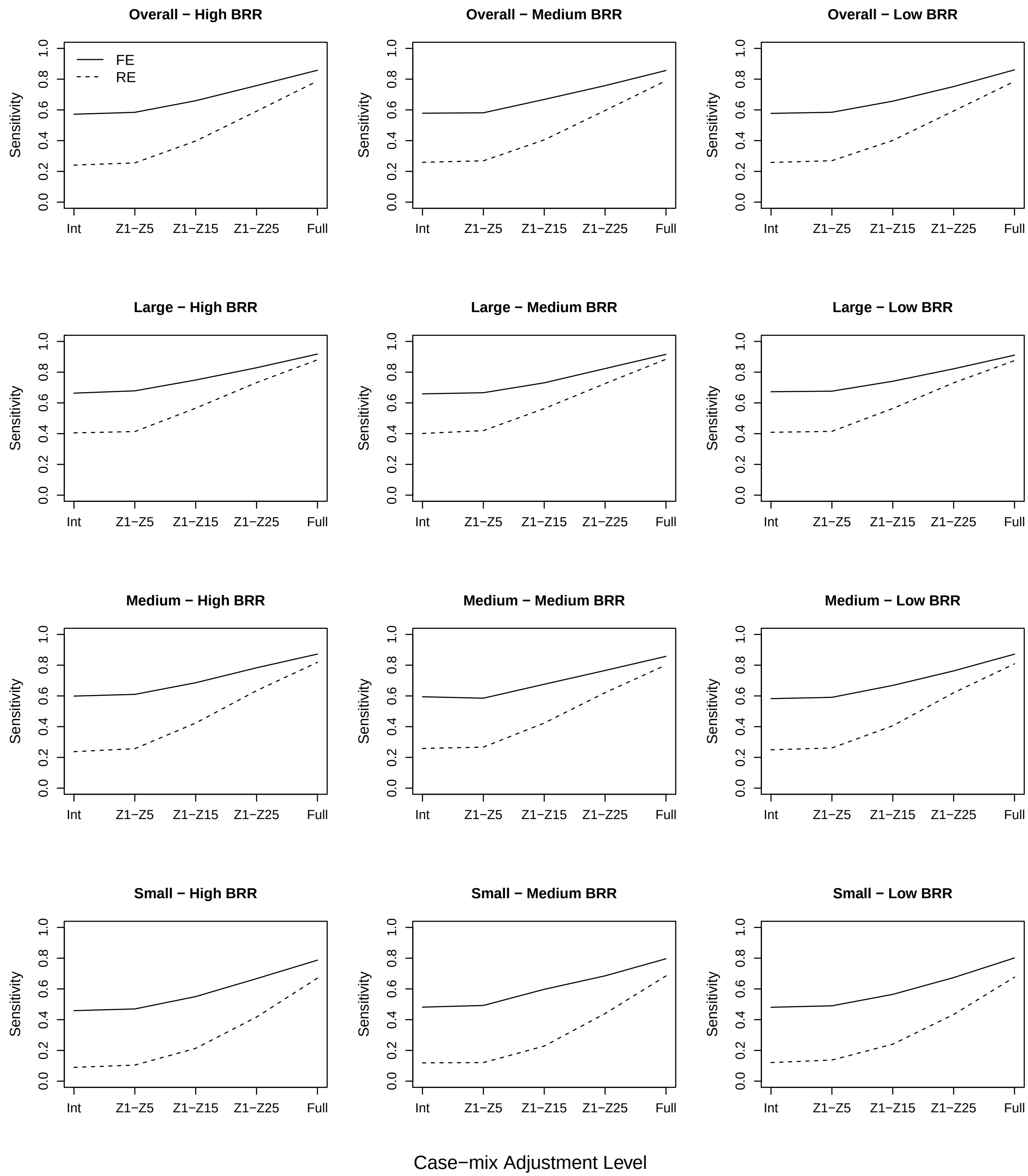

Figure S5: Impact of inadequate case-mix (CM) adjustment levels (Int: Intercept only, $Z_1 - Z_2$, $Z_1 - Z_{10}$, $Z_{11} - Z_{15}$, Full: Full model) on sensitivities of RE and FE models for simulation study based on USRDS data. Given are sensitivities to detect under-performing providers for medium baseline readmission rates (BRR). Results are presented across all providers (overall, row 1) and stratified by provider volume (large, medium, small - rows 2, 3, 4).