Lifetime Data Anal
DOI 10.1007/s10985-013-9264-6

Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects

Kevin He · Jack D. Kalbfleisch · Yijiang Li · Yi Li

Received: 6 November 2012 / Accepted: 15 May 2013
© Springer Science+Business Media New York 2013

Abstract Motivated by the national evaluation of readmission rates among kidney dialysis facilities in the United States, we evaluate the impact of including discharging hospitals on the estimation of facility-level standardized readmission ratios (SRRs). The estimation of SRRs consists of two steps. First, we model the dependence of readmission events on facilities and patient-level characteristics, with or without an adjustment for discharging hospitals. Second, using results from the models, standardization is achieved by computing the ratio of the number of observed events to the number of expected events assuming a population norm and given the case-mix in that facility. A challenging aspect of our motivating example is that the number of parameters is very large and estimation of high-dimensional parameters is troublesome. To solve this problem, we propose a structured Newton-Raphson algorithm for a logistic fixed effects model and an approximate EM algorithm for the logistic mixed effects model. We consider a re-sampling and simulation technique to obtain p-values for the proposed measures. Finally, our method of identifying outlier facilities involves converting the observed p-values to Z-statistics and using the empirical null distribution, which accounts for overdispersion in the data. The finite-sample properties

K. He (✉) · J. D. Kalbfleisch · Y. Li · Y. Li
Department of Biostatistics, University of Michigan, 1420 Washington Hts., Ann Arbor, MI 48109-2029, USA
e-mail: kevinhe@umich.edu

J. D. Kalbfleisch
e-mail: jdkalbfl@umich.edu

Y. Li
e-mail: yili@umich.edu

Y. Li
Google Inc., Mountain View, CA94043, USA
e-mail: yijiang@umich.edu

Published online: 26 May 2013

of proposed measures are examined through simulation studies. The methods developed are applied to national dialysis data. It is our great pleasure to present this paper in honor of Ross Prentice, who has been instrumental in the development of modern methods of modeling and analyzing life history and failure time data, and in the inventive applications of these methods to important national data problem.

Keywords Mixed effects · Flagging · Readmission · Standardization

1 Introduction

It is an honor to contribute to this issue of Life time Data Analysis in honor of Dr. Ross Prentice, whose work has been influential to all of us and who is a close colleague and friend to two of us. Ross has, of course, been absolutely instrumental in the development of modern methods of modeling and analyzing life history and failure time data, and in the inventive applications of these methods to important national data problems. We are very pleased to offer this paper that deals with hospital readmission, a topic with many policy-making implications, as a part of this very fitting tribute.

An unplanned hospital readmission is defined as any unplanned hospital admission that occurs within 30 days of discharge from a previous admission. Readmissions are an important indicator of patient morbidity and quality of life, and hospitalizations and readmissions are often costly, particularly among the patients with end stage renal disease (ESRD) being treated in dialysis facilities. Dialysis patients are admitted to the hospital nearly twice a year and hospitalizations account for approximately 38 % of total Medicare expenditures for dialysis patients (USRDS 2012). Furthermore, a significant percentage (30 %) of ESRD patients discharged from the hospital have an unplanned readmission within 30 days (USRDS 2012). Clinical studies have demonstrated that improved care coordination and discharge planning may reduce readmissions, while some studies (Goldfield et al. 2008) also confirm that a sizable portion of unplanned readmissions are preventable. Hence, a systematic measure of the rate of unplanned readmissions at dialysis facilities can help to identify potential problems and provide cost-effective health care.

In developing a readmission measure for dialysis facilities, it should be noted that the discharging hospital and the receiving dialysis facility share responsibility. Dialysis patients in the United States are admitted to many different hospitals and discharged to many different dialysis facilities. Thus, a dialysis facility may be treating patients discharged from multiple hospitals, and conversely, a hospital may admit patients from multiple dialysis facilities. Hospitals vary in their readmission rates as documented in the Hospital Compare measure of the Centers for Medicare and Medicaid Services (CMS) (see Horwitz et al. 2011). It can be argued that a fair evaluation of hospital readmissions for dialysis facilities should take into account the potential confounding effects of the discharging hospitals. This argument is strengthened by the fact that dialysis facilities often cannot dictate the patient's choice of hospital and, in cases when they do provide the referral, have few hospitals from which to choose. Several previous studies have suggested an influence of multiple types of providers on treatment practices and outcomes among ESRD patients (Turenne et al. 2008; Hirth et al.

2009, 2010; Turenne et al. 2010a,b). This report evaluates the impact of adjusting for discharging hospital on the estimation of facility-level readmission ratios.

The estimation of provider-level readmission rates typically consists of two steps (e.g., Horwitz et al. 2011). First, we model the dependence of readmission events on facilities and patient-level characteristics; to evaluate the impact of including discharging hospitals, we run the model both with and without an adjustment for discharging hospitals. Second, we use results from these models to compute a facility-level Standardized Readmission Ratio (SRR), defined as the ratio of the number of observed events to the number of expected events given the case-mix in that facility where the case-mix may or may not involve hospital effects. In this, the expectation is computed based on the average of national experience and the case-mix may ir may not include hospital effects.

To model the readmissions, a regression method (e.g., logistic regression) is necessary to model the dependence of readmission rates on facilities and patient-level characteristics. One approach would be to use a hierarchial generalized linear model, in which the group-specific (in our case, facility-specific) intercepts are modeled as a random effect, typically from normal distributions. Based on this hierarchial model, a risk-adjusted hospital-wide readmission (HWR) measure has been developed that reflects this aspect of care at hospitals in the United States (Horwitz et al. 2011). Although the hierarchial model is sometimes used for profiling, Kalbfleisch and Wolfe (2013), via extensive comparisons of fixed effects models (FEMs) and random effects models (REMs), argue that it is preferable to use FEMs in the context of profiling medical providers. Specifically, they found that:

(i) The FEM yields estimates of extreme values of facility effects that are less biased and have smaller mean squared error than the REM; this is important since identifying these extreme facilities is the main purpose of this kind of study.
(ii) The FEM method has higher statistical power to identify exceptional facilities, for a given false positive rate.
(iii) The REM estimates are shrunk toward the overall mean and, hence, reduce the reported variation of facility performance. This is sometimes noted as a property of 'stability' of the estimates, but the smaller variance is achieved at the cost of bias, especially for extreme effects.
(iv) The REM usually requires the assumption that the facility effect is independent of case mix. When there is correlation between patient characteristics and facility attributes (e.g., patients with less favorable measurements of health status are referred to facilities with poorer treatment strategies), REM estimates of regression coefficients are biased.

Based on this work, we prefer to use fixed effects to profile facilities. To evaluate the impact of including discharging hospitals on the estimation of facility-level SRRs, we consider two models, one with and one without hospital adjustments. In **Model 1**, no adjustments are made for hospitals; this leads to a logistic regression model for readmission rates with facilities accounting with fixed effects. In **Model 2**, hospitals are represented as random effects, which leads to a mixed-effects logistic regression model. In this, we again consider facilities as fixed effects, because facilities are the entities to be profiled. The model, however, adjusts for the potential confounding

effects of hospitals using random effects. Our general aim is to adjust each facility's readmission measures for the potential effects of the hospitals that are treating its patients. Hospitals are not being profiled and we are interested in the simultaneous estimation of their effects on the facility measures. In this case, it is natural to include hospitals as random effects in the model and to make adjustments taking account of the distribution of hospital effects across the population. With this approach, each facility's measure is adjusted for our best estimate of the true effect of each hospital, taking account of the distribution from which these effects arise. This has the advantage of circumventing problems with identifiability that would arise if hospitals were included as fixed effects and also tends appropriately to dampen the effects of hospitals with extreme outcomes.

The remainder of this article is organized as follows. In Sects. 2 and 3, we describe the models and the methods used to fit them, and the measures resulting from each. We then study issues associated with fitting these models when the numbers of facilities and/or hospitals are large. Specifically, in Sect. 2, we propose an iterative profile-likelihood-based algorithm to fit the fixed-effects model with a large number of facilities. In Sect. 3, we develop an approximate EM algorithm for fitting the more general mixed-effects model. We propose inference procedures for conducting statistical tests and constructing p-values. Finally, to 'flag' outlier facilities, we convert the p-values to Z-statistics and use methods based on the empirical null hypothesis, which accounts for overdispersion in the data (Efron 2004; Kalbfleisch and Wolfe 2013). Simulation studies are provided in Sect. 4 to illustrate and evaluate our proposed methods. Sect. 5 applies the proposed methods to the national data on readmissions for dialysis facilities. We provide some discussion of the proposed and related methods in Sect. 6.

2 Model 1: fixed effects model without adjusting for hospitals

We use subscript i to represent facility and k to represent discharge. Let F be the total number of dialysis facilities. The total number of discharges is denoted by $n = \sum_{i=1}^{F} n_i$, where n_i is the number of discharges in facility i. Let Y_{ik} denote the observed outcome for the kth discharge within the ith facility, where $i = 1, 2, \ldots, F$ and $k = 1, 2, \ldots, n_i$. In the context of our motivating example, Y_{ik} equals 1 if the kth discharge in facility i results in a readmission within 30 days, and Y_{ik} equals 0 otherwise.

We consider first a logistic model in which facilities are represented as fixed effects, and no adjustment is made for hospitals. This leads to a regression model of the form:

$$\text{logit}(p_{ik}) = \log \frac{p_{ik}}{1 - p_{ik}} = \gamma_i + \boldsymbol{\beta}^T \mathbf{Z}_{ik}, \qquad \text{(Model 1)}$$

where the parameters γ_i correspond to the fixed facility effects, $\mathbf{Z}_{ik}^T = (Z_{ik1}, Z_{ik2}, \ldots, Z_{ikr})$ is an r dimensional vector of covariates associated with the kth discharge from facility i, and $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \ldots, \beta_r)$ is a vector of regression parameters. In this, γ_i measures the facility effect in the sense that a large value of γ_i would indicate that the ith facility performs relatively poorly. Note that $p_{ik} = P\{Y_{ik} = 1 |\text{ facility } i, Z_{ik}\}$. We assume that, given the covariates, Y_{ik} and $Y_{i'k}$ are conditionally independent. It should

be noted that in our motivating example, some patients experience multiple readmissions during the observation period, and so this independence assumption would be violated. However, our preliminary results suggest that a much more complicated alternative that takes the repeated aspect into account would have negligible adjustment effects on the analysis.

2.1 Model fitting of the fixed-effects model

Fitting **Model 1** is challenging given the high-dimensional nature of the fixed parameters, as in our motivating example where F, the number of facilities, is around 5,000. Based on the observed data, the likelihood function corresponding to **Model 1** is

$$L(\gamma, \boldsymbol{\beta}) = \prod_{i=1}^{F} \prod_{k=1}^{n_i} \frac{\exp\{(\gamma_i + \boldsymbol{\beta}^T \mathbf{Z}_{ik}) Y_{ik}\}}{1 + \exp(\gamma_i + \boldsymbol{\beta}^T \mathbf{Z}_{ik})}, \tag{1}$$

where $\gamma = (\gamma_1, \ldots, \gamma_F)^T$. When F is large (e.g., $F = 5,000$), standard software fails due to the large design matrix. Note, however, that (1) can be written as

$$L(\gamma, \boldsymbol{\beta}) = \prod_{i=1}^{F} L(\gamma_i, \boldsymbol{\beta}),$$

where $L(\gamma_i, \boldsymbol{\beta}) := \prod_{k=1}^{n_i} [\exp\{(\gamma_i + \boldsymbol{\beta}^T \mathbf{Z}_{ik}) Y_{ik}\} / \{1 + \exp(\gamma_i + \boldsymbol{\beta}^T \mathbf{Z}_{ik})\}]$. On account of this, it is straightforward to estimate γ_i given $\boldsymbol{\beta}$. On the other hand, estimating $\boldsymbol{\beta}$ for given γ_i can also be routinely done. The following iterative algorithm is easily implemented:

(i) Set initial values for $\boldsymbol{\beta}^{(0)}$ and $\gamma_i^{(0)}$ and $\ell = 0$
(ii) For fixed $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\ell)}$, update γ_i using a one-step Newton-Raphson iteration as

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

where $U_i^{(\ell)}$ and $I_i^{(\ell)}$ are defined in the Appendix 1.
(iii) Now update $\boldsymbol{\beta}$ by carrying out one step of the Newton-Raphson iteration

$$\boldsymbol{\beta}^{(\ell+1)} := \boldsymbol{\beta}^{(\ell)} + I_{\boldsymbol{\beta}}^{(\ell)^{-1}} U_{\boldsymbol{\beta}}^{(\ell)},$$

where $U_{\beta}^{(\ell)}$ and $I_{\beta}^{(\ell)}$ are defined in the Appendix 1.
(iv) If $\max \| p_{ik}^{(\ell+1)*} - p_{ik}^{(\ell)*} \| > 10^{-6}$, set $\ell = \ell + 1$ and go back to step (i), where $p_{ik}^{(\ell)*}$ is defined in the Appendix 1.

2.2 Standardized readmission ratio (fixed effects)

A measure of hospital readmissions for patients under the care of a specific dialysis facility is given by the SRR, which is defined as the ratio of the observed number of readmissions to the model-based expected number of readmissions, accounting for patient-level characteristics and assuming a national norm for readmission rates. Under the fixed effects model (**Model 1**), we define

$$SRR_i{}^{(1)} = \frac{O_i}{E_i} = \frac{O_i}{\sum_{k=1}^{n_i} p_{ik}(\hat{\gamma}_M, \hat{\beta})}, \tag{SRR 1}$$

where $O_i = \sum_{k=1}^{n_i} Y_{ik}$ is the observed number of readmissions in facility i and E_i is the corresponding expected number. The latter is the sum of the estimated probabilities of readmission of all patients within this facility, assuming a national norm for the facility effect, which is specified with $\hat{\gamma}_M$ =median$(\hat{\gamma}_1, \ldots, \hat{\gamma}_F)$. In this measure, each facility is compared with an 'average' facility in the population of all facilities, adjusting for its particular case mix. Note that we introduce a median term for the 'average' facility effect; this is more robust to extreme values and avoids problems that would arise in using the mean, for example. Note also that $SRR_i{}^{(1)}$ can be viewed as an estimate of the theoretical quantity

$$\widetilde{SRR}_i^{(1)} = \frac{\sum_{k=1}^{n_i} p_{ik}(\gamma_i, \beta)}{\sum_{k=1}^{n_i} p_{ik}(\gamma_M, \beta)}.$$

An SRR lower than 1.0 indicates that the facility's observed readmission rate is less than expected based on national rates. An SRR greater than 1.0 indicates that the facility has a rate of readmission higher than would be expected based on patient characteristics and the national norm.

2.3 Evaluating the p-values (fixed-effects)

Making statistical inference about SRR is challenging given the high-dimensional nature of fixed parameters, especially when F, the number of facilities, is very large. In the context of our motivating example, some facilities had no readmissions (i.e., the observed number of readmissions $O_i = 0$), which leads to a maximum likelihood estimate of $\hat{\gamma}_i = -\infty$ for the facility effect. The usual Wald test fails in this case. To calculate a p-value, we use an 'exact' method that assesses the probability that the facility would experience a number of readmissions as least as extreme as that observed if the null hypothesis were true; this calculation accounts for each facility's patient mix. To implement this, we exploit the large-scale structure of the data, which allows $\boldsymbol{\beta}$ and γ_M to be estimated extremely precisely when F and $n = \sum n_i$ are large. Therefore, when assessing the significance of the SRR_i or γ_i, we replace $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}$ and γ_M with $\hat{\gamma}_M$ without losing precision. We further evaluate the influence of these replacements in Sect. 4 when the population size and the number of facilities are moderate.

To test $H_0 : \gamma_i = \gamma_M$ (which implies $\widetilde{SRR}_i^{(1)} = 1$), we calculate the nominal p-value for the ith facility as the probability that the observed number of readmissions should be at least as extreme as that expected, if this facility had a readmission rate corresponding to the 'average' facility. Our approach captures the most important aspects of the variability in the proposed estimator, and is described as follows

(i) Obtain $\widehat{\boldsymbol{\beta}}$ and $\hat{\gamma}_M$ and proceed under assumptions that $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\gamma_M = \hat{\gamma}_M$.
(ii) Calculate $P_{H_0}(\sum_{k=1}^{n_i} Y_{ik} \geq O_i)$ or $P_{H_0}(\sum_{k=1}^{n_i} Y_{ik} \leq O_i)$. We do this by simulation and draw B samples, $\{Y_{ik}^b : k = 1, 2, \ldots, n_i\}_{b=1}^B$, where each sample, and each observation is drawn independently from a Bernoulli distribution,

$$Y_{ik}^b \sim Ber\left(\frac{\exp(\hat{\gamma}_M + \mathbf{Z}_{ik}^T\widehat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_M + \mathbf{Z}_{ik}^T\widehat{\boldsymbol{\beta}})}\right), b = 1, 2, \ldots, B; k = 1, 2, \ldots, n_i.$$

(iii) Calculate $Y_{i.}^b := \sum_{k=1}^{n_i} Y_{ik}^b$.
(iv) Compute

$$SL_i^+ := \frac{1}{B}\sum_{b=1}^{B}\left[\frac{1}{2}I(Y_{i.}^b = O_i) + I(Y_{i.}^b > O_i)\right]$$

and

$$SL_i^- := \frac{1}{B}\sum_{b=1}^{B}\left[\frac{1}{2}I(Y_{i.}^b = O_i) + I(Y_{i.}^b < O_i)\right],$$

where O_i is the originally observed number of readmission in facility i, and $I(\dot{})$ is an indicator function. Then calculate the significance level (two tailed test), $P := 2 * \min[SL_i^+, SL_i^-]$.

Note that in step (iv), SL_i^+ the 'mid-p' values as the average of the probabilities $Y_i > O_i$ and $Y_i \leq O_i$. This avoids two-tailed p-values greater than 1 in the center of the distribution. This strategy is particularly useful when we apply methods based on the empirical null hypothesis.

3 Model 2: mixed effects model

We extend **Model 1** to accommodate a random effect representing hospitals. Thus, we use a mixed-effect logistic regression model to estimate the probability of readmission as a function of a random effect for hospital, a fixed effect for dialysis facility and a set of patient characteristics. The model is specified as

$$\log\frac{p_{ijk}}{1 - p_{ijk}} = \gamma_i + \alpha_j + \boldsymbol{\beta}^T\mathbf{Z}_{ijk}, \qquad \text{(Model 2)}$$

where we have introduced a third subscript j to indicate hospital. In this, $i = 1, \ldots, F$; $j = 1, \ldots, H$; and $k = 1, \ldots, n_{ij}$ where $n_{ij} = 0$ is allowed (a common scenario in the motivating example). Note that F is the number of facilities and H is the number of hospitals. Thus, p_{ijk} represents the probability of a readmission for the kth discharge among patients from the ith facility who are discharged from jth hospital. Here, γ_i is the fixed effect for facility i, and α_j is the random effect for hospital j. It is assumed that the α_j's arise as independent normal variables (i.e., $\alpha_j \sim N(0, \sigma_h^2)$).

3.1 Model fitting of the mixed-effects model

We now look at fitting the mixed-effects logistic regression model (**Model 2**) to the readmission data. The complete likelihood is

$$L(\boldsymbol{\alpha}; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{j=1}^{H} \left[\left\{ \prod_{i=1}^{F} \prod_{k=1}^{n_{ij}} \frac{\exp\{Y_{ijk}(\gamma_i + \alpha_j + \mathbf{Z}_{ijk}^T \boldsymbol{\beta})\}}{1 + \exp\{\gamma_i + \alpha_j + \mathbf{Z}_{ijk}^T \boldsymbol{\beta}\}} \right\} \frac{\exp\{-\alpha_j^2/(2\sigma_h^2)\}}{\sqrt{2\pi\sigma_h^2}} \right],$$

where $\boldsymbol{\gamma}^T = (\gamma_1, \gamma_2, \ldots, \gamma_F)$ is the vector of fixed effect parameters for facilities, and $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \ldots, \alpha_H)$ is the vector of (unobserved) random effect parameters for hospitals. The incomplete (or observed) likelihood is

$$L(\sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{j=1}^{H} \int_{-\infty}^{\infty} L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) d\alpha_j, \tag{2}$$

where $L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta})$ is defined implicitly. Our general aim is to use (2) to estimate σ_h, γ and β. Note that the terms "complete" and "incomplete" are used in the sense of the EM algorithm (Dempster et al. 1977), and the random effects α are viewed as missing data. Various approaches have been developed to carry out likelihood inference. Zeger and Karim (1991) used a Gibbs sampling approach based on a Markov chain Monte Carlo (MCMC) algorithm. Breslow and Clayton (1993) obtained approximate estimates based on penalized quasi-likelihood (PQL). Breslow and Lin (1995) and Lin and Breslow (1996) studied the bias in the PQL estimators and developed a bias-correction. These procedures perform reasonably well if the number of facilities, F, is of moderate size. However, when F is large (e.g., F >1,000), these approaches all encounter computational difficulties. In what follows, we present an approximate EM algorithm to address this computational difficulty. In developing this algorithm, we make use of a profile-likelihood-based method similar to that introduced in Sect. 2.1.

Let $\mathbf{Y}$ be the vector of all outcomes of Y_{ijk}. The posterior distribution of α_j, given the data, γ, β and σ_h is

$$\Pi_j(\alpha_j \mid \mathbf{Y}, \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) = L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta})/C_j,$$

where $C_j = \int_{-\infty}^{\infty} L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) d\alpha_j$. The posterior mean and variance of α_j are

$$\alpha_{j0} = \int \alpha_j \Pi_j(\alpha_j \mid \mathbf{Y}, \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) d\alpha_j$$
$$\nu_{j0} = \int (\alpha_j - \alpha_{j0})^2 \Pi_j(\alpha_j \mid \mathbf{Y}, \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) d\alpha_j.$$

To numerically approximate α_{j0} and ν_{j0}, we use a Gauss-Hermite quadrature calculation. The number of quadrature points is pre-specified to be 20, for which the approximation is usually sufficient (Lange 1999).

An approximate EM algorithm is described as follows. The E-step pertains to the calculation of a conditional expectation of the complete log-likelihood. We therefore require

$$\begin{aligned} &E\left[\log L(\boldsymbol{\alpha}; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) \mid Y, \sigma_h^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}\right] \\ &= \sum_{j=1}^{H} E\left[\log L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) \mid Y, \sigma_h^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}\right], \end{aligned} \tag{3}$$

where $(\sigma_h^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{\beta}^{(\ell)})$ are the current estimates of the parameters. Since there is no closed form for (3), we approximate $\log L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta})$ using a 2nd order Taylor expansion about α_{j0} to obtain

$$\begin{aligned} \log L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) \cong -\log\sigma - \frac{\alpha_j^2}{2\sigma_h^2} + \sum_{i=1}^{F}\sum_{k=1}^{n_{ij}} \Big\{ &(\gamma_i + \alpha_j + \boldsymbol{\beta}^T \mathbf{Z}_{ijk})\mathbf{Y}_{ijk} \\ &+ \log(q_{ijk}^0) - (\alpha_j - \alpha_{j0}) p_{ijk}^0 - \frac{(\alpha_j - \alpha_{j0})^2}{2} p_{ijk}^0 q_{ijk}^0 \Big\}, \end{aligned}$$

where $p_{ijk}^0 := p_{ijk}(\gamma_i, \alpha_{j0}, \boldsymbol{\beta})$ and $q_{ijk}^0 := 1 - p_{ijk}^0$. It follows that

$$\begin{aligned} E\left[\log L_j(\alpha_j; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) \mid Y, \sigma_h^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}\right] \cong -\log\sigma_h^{(\ell)} - \frac{(\alpha_{j0}^{(\ell)})^2 + \nu_{j0}^{(\ell)}}{2(\sigma_h^{(\ell)})^2} \\ + \sum_{i=1}^{F}\sum_{k=1}^{n_{ij}} \Big\{ (\gamma_i^{(\ell)} + \alpha_{j0}^{(\ell)} + (\boldsymbol{\beta}^{(\ell)})^T \mathbf{Z}_{ijk})\mathbf{Y}_{ijk} + \log(q_{ijk}^{(\ell)}) - \frac{(\nu_{j0}^{(\ell)})^2}{2} p_{ijk}^{(\ell)} q_{ijk}^{(\ell)} \Big\}, \end{aligned} \tag{4}$$

where $\alpha_{j0}^{(\ell)}$ and $\nu_{j0}^{(\ell)}$ are the posterior mean and variance of α_j given the data, $\sigma_h^{(\ell)}$, $\boldsymbol{\gamma}^{(\ell)}$, and $\boldsymbol{\beta}^{(\ell)}$; $p_{ijk}^{(\ell)} := p_{ijk}(\gamma_i^{(\ell)}, \alpha_{j0}^{(\ell)}, \boldsymbol{\beta}^{(\ell)})$ and $q_{ijk}^{(\ell)} := 1 - p_{ijk}^{(\ell)}$. The M-step involves the maximization of (3) with respect to $(\sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta})$ using the approximation (4). First,

$\sigma_h^{(\ell+1)} = \{\sum_j ((\alpha_{j0}^{(\ell)})^2 + \nu_{j0}^{(\ell)})/H\}^{\frac{1}{2}}$. Second, note that

$$E\left[\log L(\boldsymbol{\alpha}; \sigma_h, \boldsymbol{\gamma}, \boldsymbol{\beta}) \mid Y, \sigma^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}\right] \propto \sum_i \mathcal{L}_i^{(\ell)},$$

where

$$\mathcal{L}_i^{(\ell)} = \sum_{j=1}^{H} \sum_{k=1}^{n_{ij}} \left\{(\gamma_i^{(\ell)} + \alpha_{j0}^{(\ell)} + (\boldsymbol{\beta}^{(\ell)})^T \mathbf{Z}_{ijk})\mathbf{Y}_{ijk} + \log(q_{ijk}^{(\ell)}) - \frac{(\nu_{j0}^{(\ell)})^2}{2} p_{ijk}^{(\ell)} q_{ijk}^{(\ell)}\right\}.$$

Each γ_i can be updated using one step in the Newton-Raphson algorithm listed in the Appendix 2.

3.2 Standardized readmission ratio (mixed effects)

Let $H(i)$ denote the collection of the indices of the discharging hospitals corresponding to the ith facility. The SRR is defined as

$$SRR_i^{(2)} = \frac{O_i}{E_i} = \frac{O_i}{\sum_{j \in H(i)} \sum_{k=1}^{n_{ij}} p_{ijk}(\hat{\gamma}_M, \hat{\alpha}_j, \hat{\beta})}, \qquad \text{(SRR 2)}$$

where, as before, E_i is the expected number of readmission for patients in facility i assuming rates that apply to the 'average' facility and the hats indicate the estimated values. In particular, $\hat{\alpha}_j$ is the estimate of the (random) hospital effect obtained as the mean of the posterior distribution of α_j. Note that $SRR_i{}^{(2)}$ is an estimate of the theoretical quantity

$$\widetilde{SRR}_i^{(2)} = \frac{\sum_{j \in H(i)} \sum_{k=1}^{n_{ij}} p_{ijk}(\gamma_i, \alpha_j, \beta)}{\sum_{j \in H(i)} \sum_{k=1}^{n_{ij}} p_{ijk}(\gamma_M, \alpha_j, \beta)}.$$

Note also that this theoretical quantity depends on the true hospital effects α_j corresponding to the facility i.

3.3 Evaluating the p-values (mixed-effects)

Making statistical inferences about γ_i's under the mixed effects model is complicated. Our aim is to prescribe an approximate yet accurate inference procedure that can be executed in a computationally efficient manner. We provide below a procedure for constructing p-values for the hypothesis $\gamma_i = \gamma_M$. The calculation accounts for facility i's patient mix and the hospitals associated with the facility.

(i) As before, we take the structured parameters β, γ_M and σ_h as fixed, and then estimate $\hat{\alpha}_j$ and $\hat{\nu}_j$ using the approximate EM algorithm developed in Sect. 3.1.

(ii) We generate hospital random effects by sampling from the posterior distribution of the hospital-specific distribution. We approximate the distribution for each random effect with a normal distribution. Thus, we draw an independent sample of size B,

$$\alpha_j^b \sim N[\hat{\alpha}_j, \hat{\nu}_j],\ b = 1, \ldots, B$$

for each hospital $j = 1, \ldots, H$.

(iii) To test $H_0 : \gamma_i = \gamma_M$ (which implies $\widetilde{SRR}_i^{(2)} = 1$), we independently draw B samples, $\{Y_{ijk}^b : j = 1, 2, \ldots, H; k = 1, 2, \ldots, n_{ij}\}_{b=1}^B$. Observations within each sample are drawn independently according to

$$Y_{ijk}^b \sim Ber(\frac{\exp(\hat{\gamma}_M + \alpha_j^b + \widehat{\boldsymbol{\beta}}^T \mathbf{Z}_{ijk}))}{1 + \exp(\hat{\gamma}_M + \alpha_j^b + \widehat{\boldsymbol{\beta}}^T \mathbf{Z}_{ijk})}),$$

where $b = 1, 2, \ldots, B$; $j = 1, 2, \ldots, H$; and $k = 1, 2, \ldots, n_{ij}$.

(iv) We calculate $Y_{i..}^b := \sum_{j \in H(i)} \sum_{k=1}^{n_{ij}} Y_{ijk}^b$; the number of readmissions in facility i from the bth random sample.

(v) We compute $SL_i^+ := \frac{1}{B} \sum_{b=1}^B \left[\frac{1}{2} I(Y_{i..}^b = O_i) + I(Y_{i..}^b > O_i)\right]$ and SL_i^- is defined correspondingly; then, the two sided p-value of $\gamma_i = \gamma_M$ is $P = 2 * \min[SL_i^+, SL_i^-]$.

We remark that this approach implicitly assumes that β, γ_M and σ are estimated with little error. A more elaborate and complex sampling scheme could be derived to account for the uncertainty in these estimates, but this approximation is justified because of the large-scale nature of our motivating example.

3.4 Fitting Model 2 with a two stage approach

As examined in the next simulation section, the proposed **Model 2** works well when the data structure is balanced in that each facility's discharges came from multiple hospitals and vice versa. However, when the data are very sparse so that most of the patients in each facility are discharged from one or only a few hospitals, joint fitting of the fixed effect parameters for facilities along with the random effect for hospitals can cause numerical problems. This is the case in our motivating example, where most of the n_{ij}'s (number of discharges in facility i and hospital j) are zero, and facilities are associated with relatively few hospitals (further details are provided in Sect. 5). In this case, the proposed algorithm for fitting the mixed effects model fails to appropriately capture the hospital effects (as shown in the next section).

To circumvent this problem, we consider a two-stage approach to estimate the mixed effects model:

Stage 1: At the first stage, we fit a double random effects model. This model takes the form of **Model 2**, but both facilities and hospitals are random effects (i.e., $\gamma_i \sim N(0, \sigma_f^2)$ and $\alpha_i \sim N(0, \sigma_h^2)$ where the γ and α are mutually

independent). From this double random effects model, both σ_f and σ_h can be accurately estimated.

Stage 2: Facilities are modeled as fixed effects, and hospitals are modeled as random effects, with the standard deviation, σ_h, for hospital effects taken to be $\hat{\sigma}_h$, its estimate from Stage 1. This stage takes the form of **Model 2**, except that the standard deviation, σ_h, for hospital effect is set as constant.

4 Simulation study

We carried out a simulation study in order to examine the finite sample properties of the proposed methods. We considered three covariates: Z_1 was generated from a Bernoulli distribution with probability 0.5; and Z_2 and Z_3 were generated from independent standard normal distributions. We set $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \beta_3) = (0.5, 0.5, -0.5)$.

4.1 Simulation setting 1

In the first set of simulations, we examined the bias and empirical standard deviation of the proposed estimation procedures with relatively large numbers of facilities and hospitals ($F = H = 1,000$). The sample size n_{ij} was generated from a truncated Poisson distributions (i.e., $n_{ij} = m_{ij} * I(m_{ij} \leq 7)$. where $m_{ij} \overset{\text{iid}}{\sim} Poisson(15)$). For this setting, $Pr(n_{ij} = 0) = 0.982$, so that a large proportion of cells are empty. The observed outcomes Y_{ijk} were generated from a mixed-effect logistic regression model

$$\log \frac{P(Y_{ijk} = 1)}{1 - P(Y_{ijk} = 1)} = log(3/7) + \gamma_i + \alpha_j + \boldsymbol{\beta}^T \mathbf{Z}_{ijk}.$$

where $i = 1,\ dots, 1,000$ and $j = 1, \ldots, 1,000$. Note that we chose $log(3/7)$ to approximate the national readmission rate of about 31 %. We focused on one outlier facility (Facility 1) to evaluate the proposed methods. Specifically, we varied the magnitude of γ_1 from -1 to 1; for the more extreme values, Facility 1 is an outlier facility, which is distinct from the others. All other facility effects were generated as $\gamma_i \overset{\text{iid}}{\sim} N(0, 0.2^2)$ for $i = 2, \ldots, 1,000$. The hospital effects were generated as $\alpha_j \overset{\text{iid}}{\sim} N(0, 0.2^2)$ for $j = 1, \ldots, 1,000$. Note that $\sigma_f = \sigma_h = 0.2$ is close to the estimate from our motivating example.

In Table 1, we compare the estimated SRR from the proposed approaches. Both approaches for fitting **Model 2** performed well in this setting (estimates are the same), in the sense that the biases were small and the coverage probabilities (CPs) were close to the nominal value 0.95. Note that although the bias of **Model 1** was also small, its empirical standard deviation (ESD) was slightly larger than those from **Model 2**. Moreover, the CPs of **Model 1** were lower than those from **Model 2**. The reason for this phenomenon is that the average hospital effect was zero (since mean of α_j equals zero); hence, on average, the fixed effects model is approximately unbiased. However, individual hospital effects vary about zero, and ignoring hospital effects may affect the estimation of facility effects, resulting in a larger ESD and lower CP.

Table 1 Simulation setting 1

γ_1	True SRR	Model 1			Direct approach Model 2			Two-stage Model 2		
		Bias	ESD	CP	Bias	ESD	CP	Bias	ESD	CP
−1.0	0.50	0.008	0.12	0.94	0.000	0.11	0.96	−0.002	0.11	0.96
−0.4	0.80	0.013	0.15	0.93	0.000	0.14	0.96	−0.003	0.14	0.96
0	1.00	0.017	0.16	0.92	0.000	0.15	0.95	−0.004	0.15	0.95
0.4	1.25	0.020	0.17	0.93	0.000	0.15	0.95	−0.005	0.15	0.95
1.0	1.66	0.028	0.19	0.92	0.000	0.15	0.96	−0.007	0.15	0.96

Performance of $\widehat{SRR}$ with various values of γ_1; Number of facilities: $F = 1,000$, Number of hospitals: $H = 1,000$; 1,000 replications; Sample size: $n_{ij} = m_{ij} * I(m_{ij} \leq 7)$, where $m_{ij} \overset{\text{iid}}{\sim} Poi(15)$ ($Pr(n_{ij} = 0) = 0.98$); Facility effects: $\gamma_i \overset{\text{iid}}{\sim} N(0, 0.2^2)$, where $i = 2, \ldots, 1,000$; Hospital effects: $\alpha_j \overset{\text{iid}}{\sim} N(0, 0.2^2)$; *CP* coverage probability; *ESD* empirical standard deviation; Model 1: defined in Sect. 2.1; direct approach for Model 2: defined in the beginning of Sect. 3; two-stage Model 2: defined in Sect. 3.4

In contrast, **Model 2** accounts for the hospital effects; hence, the CPs were closer to 0.95. To further clarify the influence of replacing β with $\hat{\beta}$ and γ_M with $\hat{\gamma}_M$ when the population size is moderate, we used **Model 1** as an example and accounted for the uncertainty in the estimation of β and γ_M by bootstrap. The results busing bootstrap were very close to those based on replacing β with $\hat{\beta}$ and γ_M with $\hat{\gamma}_M$ (e.g., CPs were essentially equivalent; results not shown).

4.2 Simulation setting 2

We also performed a simulation study, mimicking the data structure of the motivating example: the number of facilities, number of hospitals and sample sizes (n_{ij}) are the same as in the real data (F =5,158 and H =5,107). Most of the n_{ij}'s were equal to zero (more than 99.9 %). The data set is sparse, in that most patients in a facility were discharged from relatively few hospitals. Instead of using the whole data set of 5, 158 facilities, we randomly drew 500 facilities and used the sub-data for further simulation. We chose this set up for two reasons. First, it dramatically saved computation time. Second, since the number of facilities was relatively small, standard statistical software, such as R and SAS, can be implemented, to compare our methods with those from the standard procedures (i.e., GLM for fixed effects model and PQL for mixed effects model). All other set ups were the same as those in Setting 1. We focused on one outlier facility (i.e., Facility 1) to evaluate the proposed methods, and we varied the magnitude of γ_1 from −1 to 1.

Using R, the results from our proposed **Model 1** were the same as those from GLM, and the results for our proposed first approach of **Model 2** (direct approach) agreed with those from PQL. The estimated random effects of hospitals from the direct approach of **Model 2** and PQL were both close to zero, which led to the same estimation of facilities effects as those from **Model 1**. In contrast, although the double random effects model led to biased estimation for γ_1, it provided good estimates of the standard deviation of the random hospital effects (bias $\cong -0.003$). This finding motivated the

Table 2 Simulation setting 2

γ_1	True SRR	Model 1			Two-stage Model 2			Double random effects		
		Bias	ESD	CP	Bias	ESD	CP	Bias	ESD	CP
−1.0	0.50	0.031	0.13	0.94	0.000	0.11	0.96	0.249	0.08	0.01
−0.4	0.80	0.046	0.14	0.91	−0.001	0.12	0.94	0.117	0.06	0.75
0	1.00	0.059	0.15	0.80	−0.006	0.13	0.93	−0.000	0.05	1.00
0.4	1.25	0.067	0.15	0.90	−0.006	0.13	0.93	−0.143	0.07	0.24
1.0	1.60	0.085	0.17	0.91	−0.009	0.14	0.95	−0.317	0.10	0.12

Performance of $\widehat{SRR}$ with various values of γ_1; 500 facilities (drawn from original data with $F = 5,158$, $H = 5,107$); 1,000 replications; Sample size: equal to those from real data; Facility effects: $\gamma_i \overset{\text{iid}}{\sim} N(0, 0.2^2)$, where $i = 2, \ldots, 500$; Hospital effects: $\alpha_j \overset{\text{iid}}{\sim} N(0, 0.2^2)$; *CP* coverage probability; *ESD* empirical standard deviation; σ: estimated standard deviation of random effects; Model 1: defined in Sect. 2.1; two-stage Model 2: defined in Sect. 3.4; Double random effects model: random effects for both facilities and hospitals

two-stage approach for fitting **Model 2**. Table 2 summarizes the performances of **Model 1**, the two-stage approach of **Model 2** and the double random effects model. The ESDs of **Model 1** were larger than those from the two-stage approach of **Model 2**. Moreover, the CPs of **Model 1** were substantially lower than those from **Model 2** (two-stage approach); the latter CPs were close to the nominal value, 0.95. Finally, the shrinkage was much more severe when using the double random effects approach. Hence, when the true facility effects were different from the population average, the bias for the double random effects approach was large and the corresponding CPs were substantially lower than the nominal value of 0.95.

5 Application

We evaluated each model using data for ESRD patients hospitalized in calendar year 2009. In all, there were 489,493 discharges from 5,107 hospitals to 5,158 dialysis facilities; 151,147 of these discharges resulted in an unplanned hospital readmission within 30 days of discharge, yielding an overall readmission rate of 30.9 %. The number of discharges per facility varied from 11 to 614, with a mean of 95 and a median of 80 discharges. The number of facilities per hospital varied from 1 to 132, with a mean of 10 and a median of 5. The number of hospitals per facility varied from 1 to 71, with a mean of 10 and a median of 9. Figure 1a, b demonstrate how facilities correspond to hospitals and vice versa. Both models included the same patient-level adjustments for age, sex, body mass index at incidence of ESRD, time since onset of ESRD, diabetes as cause of ESRD, past-year comorbidities, discharge diagnoses that are rare but have a high rate of readmission, and length of hospital stay during the index admission.

The estimated hospital effects from the direct (first) approach **Model 2** were close to zero, which led to the results identical to those from **Model 1**. This agreed with the situation we found in simulation using the same data structure, so that the first

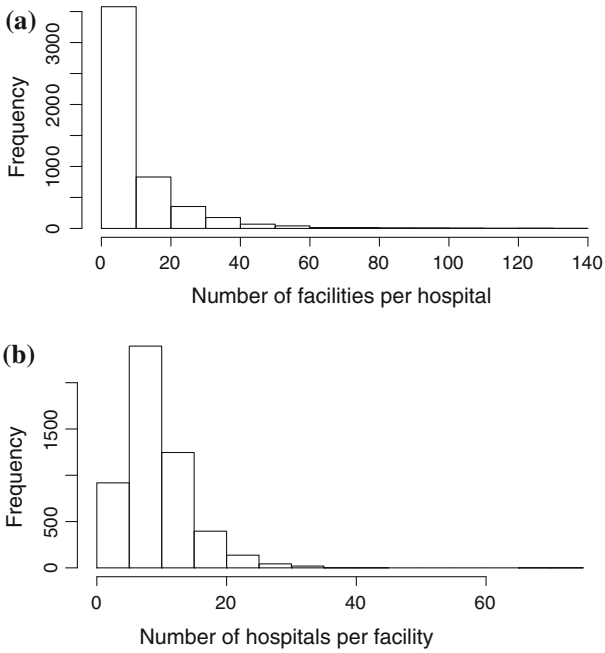


Fig. 1 Number of facilities per hospital and number of hospitals per facility

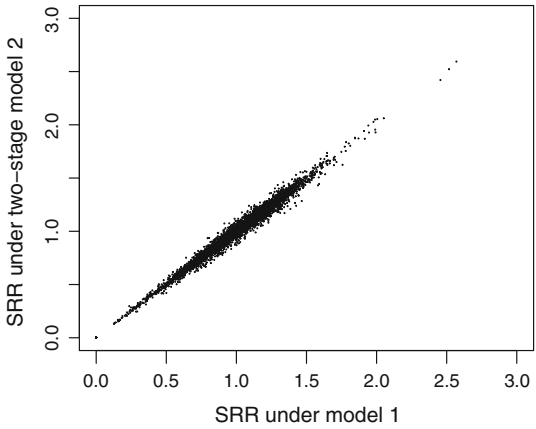


Fig. 2 Comparison of SRRs: Model 1 and two-stage approach for Model 2

approach for fitting **Model 2** did not work. Therefore, we focused on **Model 1** and the two-stage approach for **Model 2** for further comparison. Figure 2 represents the pairwise comparisons of the SRRs from **Model 1** and **Model 2**. Figure 3a, b show the distribution of these two SRRs, stratified by tertiles of numbers of hospital discharges within each facility. As expected, the variation of SRRs decreases as the number of discharges increases in both models.

To facilitate the sensitivity analysis, we also considered a more direct comparison of the the SRRs by studying the pairwise ratios of one SRR with another. For example,

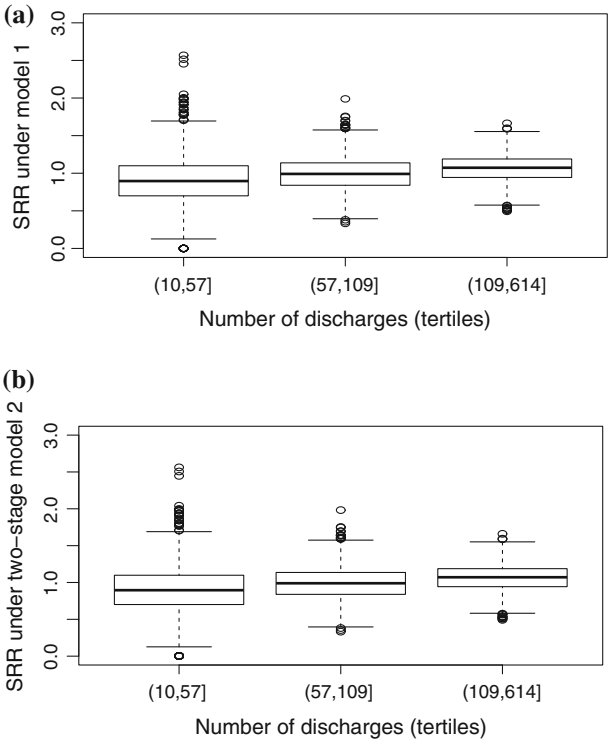


Fig. 3 SRR Distributions, by number of facility discharges (tertile 1: [11, 57]; tertile 2: [58, 109]; tertile 3: [110, 614])

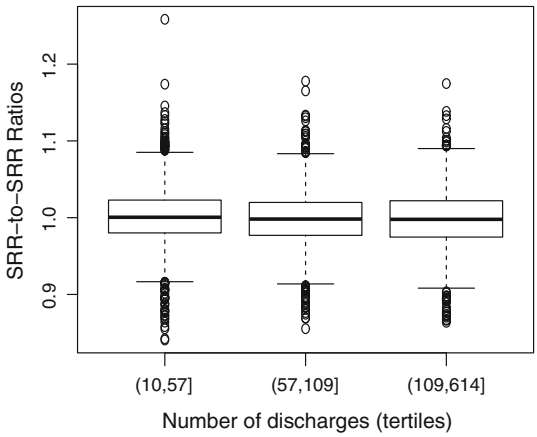


Fig. 4 Distributions of SRR-to-SRR Ratio (two-stage approach for Model 2/Model 1), by number of facility discharges

we considered the variable $R_{2,1} = SRR_i^{(2)}/SRR_i^{(1)}$. Figure 4 presents the distribution of this ratio, stratified by tertiles of numbers of hospital discharges. There are some discrepancies among these two SRRs, but the variation in this ratio is consistent across facility sizes.

Table 3 Number and percentage of outlier facilities

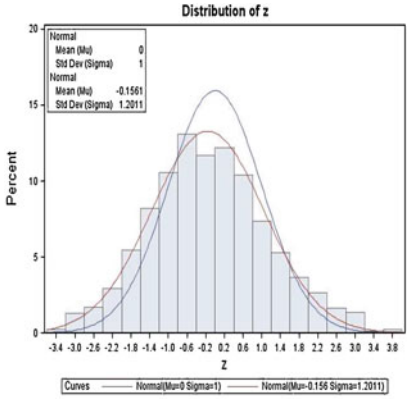
Two-stage Model 2	Model 1			
	Non-significant	Significantly-better	Significantly-worse	Row-sum
Non-significant	4366 (84.6 %)	31 (0.6 %)	60 (1.2 %)	4457 (86.4 %)
Significantly-better	47 (0.9 %)	233 (4.5 %)	0 (0 %)	280 (5.4 %)
Significantly-worse	47 (0.9 %)	0 (0 %)	374 (7.3 %)	421 (8.2 %)
Column-sum	4460 (86.4 %)	264 (5.1 %)	434 (8.5 %)	5158 (100 %)

Model 1: no adjustment for hospitals; Model 2: two-stage approach with random effects for both faculties and hospitals; Significantly-better: SRR < 1 and one-sided p-value $\leq 2.5\%$; Significantly-worse: SRR > 1 and one-sided p-value $\leq 2.5\%$

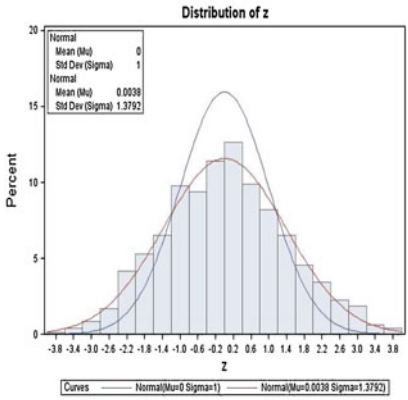
Table 3 presents the pairwise comparison of the numbers and percentages of outlier facilities identified by the p-values corresponding to their SRRs (using a test SRR = 1). For example, a total of 107 facilities changed outlier status when switching between these two models. Specifically, 60 facilities that were significantly worse-than-average based on **Model 1** were not significant based on the two-stage approach for fitting **Model 2**. On the other hand, 47 facilities that were significantly worse-than-average based on **Model 2** were not significant based on **Model 1**. In summary, adjusting for hospitals has some influence, although relatively small, on the estimation of SRRs. Typically, the difference between the two SRRs was less than 10 %, and most SRRs changed less than 5 % between the models.

Finally, to address the problem of simultaneously monitoring a large number of facilities, we used the method discussed in Kalbfleisch and Wolfe (2013). Essentially, the method is based on the empirical null (Efron 2004, 2007), which accounts for unexpected overdispersion in the data. The p-value for each facility was converted to a Z-score, and the corresponding histograms from the two-stage approach of **Model 2** are plotted in Fig. 5, stratified into three groups based on numbers of discharges within each facility. The $N(0, 1)$ density is then superimposed on the histogram along with a normal curve fitted to the center of the histograms using a robust M-estimation method. The overdispersion of the Z-scores is substantial in facilities with a larger number of discharges. It is clear that the departure from the null is related to the number of discharges, which is consistent with the finding in Kalbfleisch and Wolfe (2013). This motivated us to refer from the empirical null distribution to assess outlier facilities and to stratify the adjustment for overdispersion on the number of discharges within facilities. Table 4 presents the flagging numbers and rates for various methods, which includes the proposed **Model 1**, two-stage approach for fitting **Model 2**, a single random effect model (random effect for facilities, and no adjustment for hospital), and a double random effects model (random effects for both facilities and hospitals). The results presented are based on a one-tailed test where the relevant p-value is 2.5% or less. Similar to previous findings, the total proportions of outlier facilities flagged by **Model 1** and **Model 2** are comparable. The standard SRR method based on **Model 2** would flag about 8.1% of facilities overall. Facilities with a larger number of discharges are more likely to be flagged (16.4%) than those facilities with smaller number of discharges (2.4%). In contrast, the empirical null method makes an appro-

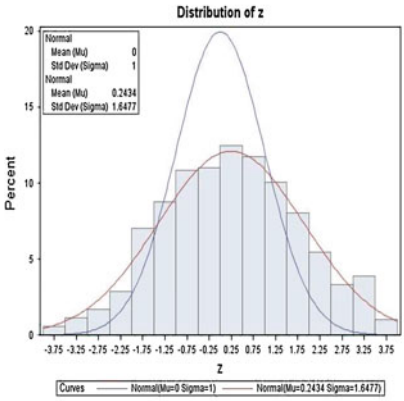
Fig. 5 Histogram of Z Scores (based on two-stage approach for Model 2), by number of facility discharges (tertile 1: [11, 57]; tertile 2: [58, 109]; tertile 3: [110, 614])



(a) (Smallest Third)



(b) (Middle Third)



(c) (Largest Third)

Table 4 Number and percentage of facilities flagged as significantly worse based on the empirical null

Number of discharges	Model 1 test		Two-stage Model 2 test		Single random	double random
	SRR = 1	Emp null	SRR = 1	Emp null	SRR = 1	SRR = 1
[11, 57]	42 (2.5 %)	99 (5.8 %)	41 (2.4 %)	94 (5.5 %)	1 (0.1 %)	0
[58, 109]	111 (6.3 %)	74 (4.2 %)	100 (5.7 %)	79 (4.5 %)	31 (1.8 %)	4 (0.2 %)
[110, 614]	286 (16.9 %)	36 (2.1 %)	278 (16.4 %)	63 (3.7 %)	171 (10.1 %)	43 (2.5 %)
Overall	439 (8.5 %)	209 (4.1 %)	419 (8.1 %)	236 (4.6 %)	203 (3.9 %)	47 (0.9 %)

Flagging is based on a one-sided p-value of 2.5 % or less; single random effects model: random effects for facilities, no adjustment for hospitals; double random effects model: both random effects for facilities and hospitals

priate adjustment in each of the strata and yields fairly consistent flagging rates across all strata. It is interesting to note that the test based on random effects model (either a single random effects model without adjustment for hospitals, or a double random effects model with both random effects for facilities and hospitals) flags fewer outlier facilities. Especially for facilities with a small number of discharges, the single random effects model flagged only one facility and the double random effects model flagged none. The shrinkage in this setting is severe under the random effects approach.

6 Discussion

The purpose of this paper is to develop and evaluate a method for including discharging hospitals on the estimation of facility-level SRRs. To model the dependence of readmission events on facilities and patient-level characteristics, we evaluated two models that are distinguished by the level of control for the influence of the discharging hospital, while accounting for facilities as fixed effects. One purpose of instituting an SRR measure for dialysis facilities is to encourage communication between dialysis facilities and hospitals with respect to the effective treatment of a patient following a hospital discharge. It might be argued that one potential advantage of **Model 1** is that it makes no adjustment for hospitals and hence provides a strong incentive for dialysis facilities to coordinate patient care with the discharging hospitals. It should be noted, however, that there remains a substantial benefit to facilities to coordinate care even with **Model 2**. In addition, **Model 2** recognizes that some aspects of the hospitals' care are outside the dialysis facilities' control. With **Model 1**, some dialysis facilities might argue that the reason for their poor outcomes with respect to patient readmissions is the poor care of the discharging hospital. The adjustment in **Model 2**, however, weakens this rationale and reduces the force of this argument or excuse. We recommend using the SRR computed under **Model 2** because this model specification accounts, to some extent, for the potential confounding effects of hospitals. The inclusion of this effect is perceived to provide a fair presentation of facility effects.

In terms of our choice of regression model for profiling dialysis facilities, we prefer fixed effects to random effects, specifically when identifying facilities with extreme outcomes. Fixed effects provide more precise estimation of the true effects for those facilities with extreme outcomes. In contrast, random effects result in shrinkage estimators (where the estimate for each facility is shifted toward the overall mean), and the shrinkage is particularly large for smaller facilities. This makes identification of poor performance in smaller facilities even more difficult.

Spiegelhalter et al. (2012) discussed strategies for health care regulation. The measure we considered in this report focuses on measuring the deviation of dialysis facilities relative to the overall population 'average'. Our proposed method considered the situation in which the primary goal is to assess whether a threshold has been breached. Another important strategy is to select a proportion of health care providers to inspect, for which particular attention should be paid to the problem of simultaneously monitoring a large number of indicators (i.e., overassertion may occur). We addressed this issue using the methods based on the empirical null hypothesis, which accounts for overdispersion in the data.

It is worth noting that the denominator of the SRR estimates the expected number of readmissions given the observed number of discharges. Therefore, the SRR may unfairly penalize a facility with a low hospitalization rate but many readmissions. Dialysis facilities are also reviewed with respect to their overall hospitalization rates, and the Standardized Hospitalization Rate (SHR) compares the number of hospitalizations to the expected number of hospitalizations in each facility (see, for example, Liu et al. 2012). These two measures, the SHR and the SRR together, help to address this issue.

Acknowledgements The authors would like to thank Dr. Marc Turenne, Professor John Wheeler, Professor Joseph Messana, Ms. Deanna Chyn, Ms. Tempie Shearon and Ms. Valarie Ashby for helpful discussion and comments. We also acknowledge with thanks the comments from the Editors and referees on this paper, which helped to improve the presentation. This work was supported in part by a contract from the Centers for Medicare and Medicaid Services (CMS), although the opinions presented here are not necessarily those of the CMS.

7 Appendices

7.1 Appendix 1: model fitting algorithm of the fixed-effects model

(i) Set initial values for $\boldsymbol{\beta}^{(0)}$ and $\gamma_i^{(0)}$ and $\ell = 0$

(ii) For fixed $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\ell)}$, update γ_i using a one-step Newton-Raphson iteration as

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

where

$$U_i^{(\ell)} := \frac{\partial}{\partial \gamma_i} \log L(\gamma_i, \boldsymbol{\beta}^{(\ell)})\Big|_{\gamma_i=\gamma_i^{(\ell)}} = \sum_{k=1}^{n_i} [Y_{ik} - p_{ik}^{(\ell)}],$$

$$I_i^{(\ell)} := -\frac{\partial^2}{\partial \gamma_i^2} \log L(\gamma_i, \boldsymbol{\beta}^{(\ell)})\Big|_{\gamma_i=\gamma_i^{(\ell)}} = \sum_{k=1}^{n_i} p_{ik}^{(\ell)}[1 - p_{ik}^{(\ell)}],$$

with

$$p_{ik}^{(\ell)} := p_{ik}(\gamma_i^{(\ell)}, \boldsymbol{\beta}^{(\ell)}).$$

(iii) Now update $\boldsymbol{\beta}$ by carrying out one step of the Newton-Raphson iteration

$$\boldsymbol{\beta}^{(\ell+1)} := \boldsymbol{\beta}^{(\ell)} + I_{\boldsymbol{\beta}}^{(\ell)^{-1}} U_{\boldsymbol{\beta}}^{(\ell)},$$

where

$$
\begin{aligned}
U_{\boldsymbol{\beta}}^{(\ell)} &:= \frac{\partial}{\partial \boldsymbol{\beta}} \log L(\gamma^{(\ell+1)}, \boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(\ell)}} \\
&= \sum_{i=1}^{F}\sum_{k=1}^{n_i}\{Y_{ik} - p_{ik}^{(\ell)*}\}\mathbf{Z}_{ik}, \\
I_{\boldsymbol{\beta}}^{(\ell)} &:= -\frac{\partial^2}{\partial\beta\partial\beta^T} \log L(\gamma^{(\ell+1)}, \boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(\ell)}} \\
&= \sum_{i=1}^{F}\sum_{k=1}^{n_i} p_{ij}^{(\ell)*}\{1 - p_{ik}^{(\ell)*}\}\mathbf{Z}_{ik}\mathbf{Z}_{ik}^T,
\end{aligned}
$$

with

$$
p_{ik}^{(\ell)*} := p_{ik}(\gamma_i^{(\ell+1)}, \boldsymbol{\beta}^{(\ell)}).
$$

(iv) If $\max \| p_{ik}^{(\ell+1)*} - p_{ik}^{(\ell)*} \| > 10^{-6}$, set $\ell = \ell + 1$ and go back to step (i).

7.2 Appendix 2: Newton-Raphson algorithm for the mixed-effects model

The γ_i is updated as

$$
\gamma_i^{(\ell+1)} = \gamma_i^{(\ell)} - \mathcal{L}_i''^{(\ell)^{-1}} \mathcal{L}_i'^{(\ell)},
$$

where

$$
\begin{aligned}
\mathcal{L}_i'^{(\ell)} &:= \frac{\partial}{\partial\gamma_i}\mathcal{L}_i^{(\ell)}\Big|_{\gamma_i=\gamma_i^{(\ell)}} = \sum_{j=1}^{H}\sum_{k=1}^{n_{ij}}\{Y_{ijk} - p_{ijk}^{(\ell)} + \frac{\nu_{j0}^{(\ell)}}{2}(p_{ijk}^{(\ell)}q_{ijk}^{(\ell)^2} - p_{ijk}^{(\ell)^2}q_{ijk}^{(\ell)})\} \\
&= \sum_{j=1}^{H}\sum_{k=1}^{n_{ij}} a_{ijk}^{(\ell)}, \\
-\mathcal{L}_j''^{(\ell)} &:= -\frac{\partial^2}{\partial\gamma_i^2}\mathcal{L}_i^{(\ell)}\Big|_{\gamma_i=\gamma_i^{(\ell)}} = \sum_{j=1}^{H}\sum_{k=1}^{n_{ij}}\{p_{ijk}^{(\ell)}q_{ijk}^{(\ell)} + \frac{\nu_{j0}^{(\ell)}}{2}p_{ijk}^{(\ell)}q_{ijk}^{(\ell)}(q_{ijk}^{(\ell)^2} + p_{ijk}^{(\ell)^2} - 4p_{ijk}^{(\ell)}q_{ijk}^{(\ell)})\} \\
&= \sum_{j=1}^{H}\sum_{k=1}^{n_{ij}} b_{ijk}^{(\ell)}.
\end{aligned}
$$

Similarly, $\boldsymbol{\beta}$ is updated as

$$
\boldsymbol{\beta}^{(\ell+1)} = \boldsymbol{\beta}^{(\ell)} - \mathcal{L}_{\boldsymbol{\beta}}''^{(\ell)^{-1}} \mathcal{L}_{\boldsymbol{\beta}}'^{(\ell)},
$$

where

$$\mathcal{L}'^{(\ell)}_{\boldsymbol{\beta}} := \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{F} \mathcal{L}_i^{(\ell)} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(\ell)}} = \sum_{i=1}^{F} \sum_{j=1}^{H} \sum_{k=1}^{n_{ij}} \mathbf{Z}_{ijk} a_{ijk}^{(\ell)},$$

$$-\mathcal{L}''^{(\ell)}_{\boldsymbol{\beta}} := -\frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \sum_{i=1}^{F} \mathcal{L}_i^{(\ell)} = \sum_{i=1}^{F} \sum_{j=1}^{H} \sum_{k=1}^{n_{ij}} b_{ijk}^{(\ell)} \mathbf{Z}_{ijk}^T \mathbf{Z}_{ijk}.$$

The $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are then repeatedly updated until they converge.

References

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25

Breslow NE, Lin X (1995) Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika 82:81–91

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 39:1–38

Efron B (2004) Large-scale simultaneous hypothesis testing: the choice of a null hypotheses. J Am Stat Assoc 99:96–104

Efron B (2007) Size, power and false discovery rates. Ann Stat 35:1351–1377

Goldfield NI, McCullough EC, Hughes JS, Tang AM, Eastman B, Rawlins LK, Averill RF (2008) Identifying potentially preventable readmissions. Health Care Financing Rev 30:75–91

Hirth RA, Turenne MN, Wheeler J, Pan Q, Ma Y, Messana JM (2009) Provider monitoring and pay-for-performance when multiple providers affect outcomes: an application to renal dialysis. Health Serv Res 44(5):1585–1602

Hirth RA, Turenne MN, Wheeler J, Ma Y, Messana JM (2010) Do resource utilization and clinical measures still vary across dialysis chains after controlling for the local practices of facilities and physicians? Medical Care 48(8):726–732

Horwitz L, Partovain C, Lin ZQ, Herrin J, Grady J, Conover M, Montague J, Dillaway C, Bartcazk K, Suter L, Ross J, Bernheim S, Krumholz H, Drye E (2011) Hospital-wide all-cause risk-standardized readmission measure: measure methodology report. Yale New Heaven health services corporation/center for outcomes research and evaluation (YNHHSC/CORE). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/download/MMSHospital-WideAll-ConditionReadmissionRate.pdf. Accessed 27 March 2013

Kalbfleisch JD, Wolfe RA (2013) On monitoring outcomes of medical providers. Stat Biosciences (in press)

Lange K (1999) Numerical analysis for statisticians. Springer, New York

Lin X, Breslow NE (1996) Bias correction in generalized linear mixed models with multiple component of dispersion. J Am Stat Assoc 91:1007–1016

Liu D, Schaubel DE, Kalbfleisch JD (2012) Computationally efficient marginal models for clustered recurrent event data. Biometrics 68:637–647

Spiegelhalter D, Sherlaw-Johnson C, Bardsley M, Blunt I, Wood C, Grigg O (2012) Statistical methods for healthcare regulation: rating, screening and surveilliance. J Royal Stat Soc A 175(1):1–47

Turenne MN, Hirth RA, Pan Q, Wolfe RA, Messana JM, Wheeler J (2008) Using knowledge of multiple levels of variation in care to target performance incentives to providers. Medical Care 46(2):120–126

Turenne MN, Hirth RA, Messana JM, Turner JS, Sleeman KK, Wheeler J (2010a) When payment systems collide: the effect of hospitalization on anemia in renal dialysis patients. Medical Care 48(4):296–305

Turenne M, Hunter S, Wolfe R, Shearon T, Pearson J, Kalbfleisch J, Dahlerus C, Wheeler J (2010b) 30-day hospital readmission among dialysis patients: influence of dialysis facilities versus hospitals. Annual Meeting of the American Society of Nephrology, Denver

US Renal Data System (USRDS) (2012) Annual data report: atlas of chronic kidney disease and end-stage renal disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD. http://www.usrds.org/2012/slides/indiv/v1index.html Accessed 27 March 2013

Zeger SL, Karim R (1991) Generalized linear models with random effects: a Gibbs sampling approach. J Am Stat Assoc 86:79–86