Title
Time-dynamic profiling with application to hospital readmission among patients on dialysis

Journal
Biometrics, 74(4)

ISSN
0006-341X

Authors
Estes, Jason P
Nguyen, Danh V
Chen, Yanjun
et al.

Publication Date
2018-12-01

DOI
10.1111/biom.12908

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2019 February 04.

Published in final edited form as:

Biometrics. 2018 December ; 74(4): 1383–1394. doi:10.1111/biom.12908.

Time-Dynamic Profiling with Application to Hospital Readmission Among Patients on Dialysis

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

[1]Department of Biostatistics, University of Michigan Ann Arbor, MI 48109, U.S.A.

[2]Department of Medicine, University of California Irvine, Orange, CA 92868, U.S.A.

[3]Institute for Clinical and Translational Science, Irvine, CA 92697, U.S.A.

[4]Department of Medicine, University of California Davis, Sacramento, CA 95817, U.S.A

[5]Department of Biostatistics, University of California, Los Angeles, CA 90095, U.S.A.

Summary:

Standard profiling analysis aims to evaluate medical providers, such as hospitals, nursing homes or dialysis facilities, with respect to a patient outcome. The outcome, for instance, may be mortality, medical complications or 30-day (unplanned) hospital readmission. Profiling analysis involves regression modeling of a patient outcome, adjusting for patient health status at baseline, and comparing each provider's outcome rate (e.g., 30-day readmission rate) to a normative standard (e.g., national "average"). Profiling methods exist mostly for non time-varying patient outcomes. However, for patients on dialysis, a unique population which requires continuous medical care, methodologies to monitor patient outcomes continuously over time are particularly relevant. Thus, we introduce a novel time-dynamic profiling (TDP) approach to assess the time-varying 30-day readmission rate. TDP is used to estimate, for the first time, the risk-standardized time-dynamic 30-day hospital readmission rate, throughout the time period that patients are on dialysis. We develop the framework for TDP by introducing the standardized dynamic readmission ratio as a function of time and a multilevel varying coefficient model with facility-specific time-varying effects. We propose estimation and inference procedures tailored to the problem of TDP and to overcome the challenge of high-dimensional parameters when examining thousands of dialysis facilities.

Keywords

End-stage renal disease; Hospital readmission; Multilevel varying coefficient models; Profiling of medical care providers; United States Renal Data System

* dsenturk@ucla.edu.

1. Introduction

End-stage renal disease (ESRD) is kidney failure requiring long-term renal replacement therapy with either dialysis or kidney transplantation. Most recent population data shows that there were over 661,000 individuals with ESRD in the United States (US) as of December 31, 2013 (United States Renal Data System Annual Data Report [USRDS ADR] 2015). About 70% of patients with ESRD were on dialysis. In this population, 3-year mortality is about 55% and, on average, dialysis patients are admitted to a hospital nearly twice a year, and about 30% have an (unplanned) readmission within the 30 days following discharge (ADR 2015). Methodologies to monitor patient outcomes *continuously over time* are of particular relevance for patients on dialysis, since they require continuous medical care on maintenance dialysis, due to the limited treatment options through kidney transplantation. Towards this objective, we introduce a novel time-dynamic monitoring (profiling) method to assess the time-varying 30-day readmission rate. The method is used to estimate, for the first time, the risk-standardized time-varying readmission rate (throughout the time period that patients are on dialysis) for each dialysis facility in the US.

Profiling or evaluation of medical care providers (e.g., hospitals, nursing homes or dialysis facilities) involves the application of statistical models to compare quality of care, service utilization or cost relative to a normative standard (e.g., a national “average” standard). For example, a patient outcome, such as 30-day hospital readmission rate for patients at a specific dialysis facility, may be compared to a norm based on national rates of dialysis facilities. Profiling methods exist mostly for *static or non time-varying patient outcomes*, such as an annual risk-standardized mortality or annual 30-day hospital readmission rate (Normand et al. 1997; Normand and Shahian 2007; Ash et al. 2011; Horwitz et al. 2011; He et al. 2013; Kalbfleisch and Wolfe 2013; CMS 2014 and references therein). The limited time-varying metrics in the literature include the risk-adjusted CU mulative SUMmation (CUSUM) and observed-expected CUSUM techniques (Biswas and Kalbfleisch 2008; Sun and Kalbfleisch 2013; Van Rompaye et al. 2015), although these are tailored to survival time outcomes. In profiling, the modeling process to evaluate a provider includes an important risk-adjustment step to account for patient case-mix (i.e., baseline patient-level comorbidities). Profiling analyses serve many purposes, including identification of providers with below standard performance by government agencies for regulatory or payment policies; informing providers for improving patient care; providing patient outcome information to guide patients’ selection of specific providers; and research that inform patient care. We note that profiling analysis dates back nearly a century (Codman 1916), although the methodologies have received renewed attention, particularly since the Center for Medicare and Medicaid Services (CMS) implementation of a new national patient quality strategy (CMS 2016).

To address the challenge of profiling over time, we introduce a new approach called time-dynamic profiling (TDP) to provide a time-varying metric for assessing performance over time for generalized patient outcomes (e.g., hospital readmission rate). The main goal of TDP is to evaluate a (dialysis) facility’s performance, with respect to a patient outcome measure, as a function of time that patients are on dialysis (vintage). This allows for examining distinct time periods of under-performance during dialysis relative to a reference

standard. Our proposed TDP also conforms to the necessary requirements of traditional profiling analysis: it compares a facility's performance relative to a norm (e.g., a national standard) and accounts for baseline patient case-mix prior to dialysis (risk-standardized). We note here that our illustrative application of TDP using USRDS data (in Section 3) is based on 30-day hospital readmission as the outcome; however, other patient outcomes are also relevant to patients on dialysis, including mortality, transfusion, dialysis adequacy and hypercalcemia rate. The model, estimation, and inference procedures that we propose are of sufficient generality to be applicable to other settings, particularly for chronic disease conditions over time.

For TDP, we propose a multilevel varying coefficient model (MVCM) with fixed facility time-varying effects and subject-specific random effects to accommodate the multilevel data structure with patients nested within dialysis facilities, and observations over time nested within patients. The choice of fixed versus random facility effects has been examined for time-static profiling (Kalbfleisch and Wolf 2013). While the average absolute error in estimation is typically smaller overall for random effects models, this average gain is achieved mostly in the center of the distribution of the outcomes. However, when the main inferential interest is to identify extreme facilities, fixed effects models have been reported to be effective in identifying outliers in dialysis facility profiling (Kalbfleisch and Wolf 2013). Another reason for our choice of fixed facility effects is that fixed effects models do not have the inherent problem of confounding of patient risk with facility-level effects in multilevel modeling for dialysis facility profiling. Furthermore, to accommodate the dependence among the repeated measures on the same subject, we introduce subject-specific random effects in the MVCM.

Currently, available standard estimation and inference tools for MVCMs are not applicable to profiling analysis because they were not designed to handle the high-dimensional parameters associated with fixed effects MVCMs. For instance, the proposed TDP modeling for USRDS data involves approximately 400,000 discharges of over 100,000 patients from approximately 3,000 dialysis facilities. Thus, the problem entails simultaneous estimation of close to 3,000 varying coefficient functions (together with a large number of regression parameters). Estimation of these high-dimensional parameters cannot be handled in a single step via standard local smoothing or basis expansions, where such standard estimation techniques fail due to the high-dimension of the design matrix. To put our proposed MVCM into context, we review some of the relevant standard varying coefficient model (VCM) literature. VCMs were introduced by Cleveland et al. (1991) and Hastie and Tibshirani (1993) to model time-varying effects; see review in Fan and Zhang (2008). The literature on the standard VCMs (Hoover et al. 1998; Qu and Li 2006; Şentürk et al. 2013; Estes et al. 2014) have largely been for "single-level" data. Limited works have considered mixed VCMs (Liang et al. 2003; Zhang 2004), mostly for the analysis of regular longitudinal data (i.e., without the higher-level units). A multilevel functional regression model with functional predictors was considered by Crainiceanu et al. (2009), but for modeling a scalar response. A MVCM for a space- and time-varying response and predictors was proposed by Serban (2011). Although the model was useful for its intended purpose, it is not applicable to the inferential goal of profiling.

We formulate the MVCM for profiling analysis and introduce a time-dynamic facility performance assessment index called *standardized dynamic readmission ratio* (SDRR) in Section 2. To overcome the estimation and inference challenges for a high-dimensional parameter space, we propose an approximate EM algorithm based on an iterative Newton-Raphson approach (Section 2.2), inspired by the work of He et al. (2013). We develop a hypothesis testing procedure in Section 2.3 which is tailored to the goal of profiling and is specifically developed for addressing the challenge of drawing inference on a very large number of varying coefficient functions. A novel application to assess time-dynamic 30-day readmission rates for dialysis facilities based on USRDS data is given in Section 3. Simulation studies examine the efficacy of the proposed method (Section 4). We conclude with a discussion in Section 5.

2. Multilevel Varying Coefficient Model for Time-Dynamic Profiling

Assume that we have a cohort of incident dialysis patients followed over time from the start of dialysis. Let $i = 1,\ldots, I$ denote dialysis facilities and $j = 1,\ldots, N_i$ denote patients receiving dialysis treatment at facility i with N_i total number of patients. Further, let $k = 1,\ldots, N_{ij}$ index hospitalizations for patient j at facility i, where N_{ij} denotes the total number of hospitalizations. Let the outcome variable, denoted by $Y_{ijk} = Y_{ij}(t_{ijk})$, equal 1 if the kth index hospitalization of the jth patient within facility i results in a readmission within 30 days, and equal 0 otherwise. Here, t_{ijk} denotes the time of the index hospitalization after the jth patient has initiated dialysis. There are several important considerations in building models for TDP. First, the model needs to respect the multilevel data structure and needs to allow for potential time-varying effects for each facility. Second, longitudinal patient-level variables or post-dialysis cross-sectional covariates (such as adverse events during dialysis or patient health attributes acquired after the start of dialysis which might be the result of care) need to be excluded to avoid confounding with the time-dynamic facility-level effects, the main quantity of interest. Third, the extent to which the proposed modeling will be useful will depend on having a set of rich baseline covariates (case-mix), $Z_{ij} = (Z_{1ij},\ldots, Z_{rij})^{\mathrm{T}}$, for each patient that when taken together adequately captures each patient's health characteristics prior to dialysis. For this risk adjustment step, we follow CMS-defined risk adjustment variables to adjust for age, sex, body mass index (BMI), 23 comorbidities and presence of high-risk diagnosis at a hospital discharge during the year prior to dialysis (past-year comorbidities), and whether diabetes was the cause of ESRD (CMS 2014). (Additional details are in Web Appendix C.) Note that case-mix risk variables (e.g., high-risk diagnosis) are not defined at the time of the index hospitalization, since this would constitute adjusting for longitudinal covariates with multiple index hospitalizations per patient. We emphasize that all risk variables are determined, instead, based on discharges during the year prior to dialysis.

To achieve the above goals we propose the generalized (logistic) MVCM

$$g\left[E\left\{Y_{ij}(t)\middle|Z_{ij}, b_{ij}, t < S_{ij}\right\}\right] = g\left\{p_{ij}(t)\right\} = \gamma_i(t) + b_{ij} + Z_{ij}^{\mathrm{T}}\beta, \quad i = 1,\ldots, I, \quad \text{(M1)}$$

where g is the logit link function, t denotes the time after initiation of dialysis, S_{ij} denotes death time of subject j, the functions $\gamma_I(t)$ correspond to the fixed time-varying facility-level effects, $b_i = \left(b_{i1}, \ldots, b_{iN_i}\right)^{\mathrm{T}}$ correspond to subject-specific random effects (REs) within the ith facility with variance σ_b^2, $\beta = \left(\beta_1, \ldots, \beta_r\right)^{\mathrm{T}}$ is a vector of regression parameters, and $p_{ij}(t) \equiv E\left\{Y_{ij}(t) | Z_{ij}, b_{ij}, t < S_{ij}\right\} = g^{-1}\left\{\gamma_i(t) + b_{ij} + Z_{ij}^{\mathrm{T}}\beta\right\}$ denotes the 'partly conditional' target of inference, conditional on being alive $t < S_{ij}$. The partly conditional target of inference has been considered before in modeling longitudinal data (Kurland and Heagerty 2005) and more recently in the context of varying coefficient models (Estes et al. 2014; 2015) to model time-varying regression effects in the dynamic cohort of survivors. We alert the reader here that model (M1) involves the collection of high-dimensional parameter space $\left\{\gamma_1(t), \ldots, \gamma_I(t), \beta_1, \ldots, \beta_r, \sigma_b^2\right\}$ that requires simultaneous estimation. For our application, this involves simultaneous estimation and inference for about 2,900 γ_I(t)'s and over 20 covariates, β_r's. One should not confuse model (M1) as a separate model for each facility i.

Several modeling assumptions, choices, and limitations are inherent in the TDP setting. First, similar to the formulation in time-static profiling, the readmission risk, by definition, is conditional on the existence of an index hospitalization. The MVCM in (M1), which we will refer to as Model 1, extends the standard fixed effects logistic regression model of He et al. (2013) for time-static dialysis facility profiling, to model time-dynamic facility effects via facility-level varying coefficient functions, which are the main interest in TDP. Second, we do not make adjustments for longitudinal covariates and target facility-level effects under the assumption that, changes in a patient's longitudinal readmission risk after initiation of dialysis are directly attributable to the care they receive at the dialysis facility, once adjusted for the health status of the patient prior to dialysis, characterized by the rich set of case-mix (Z_{ij}), which are assumed to be fixed over time within a facility. Third, even though, no adjustment are made for time-varying covariates on the causal pathway of facility-level effects, we include subject-specific random effects in Model 1, which are related to longitudinal changes in the readmission risk of the patient. We view inclusion of the subject-specific random effects as a compromise between valid inference in TDP and avoiding confounding of facility effects, since as will be shown (through simulations of Section 4), their exclusion do not produce valid inference in testing for outlier facilities. Fourth, for mathematical convenience in deriving a feasible estimation procedure for Model 1 in Section 2.2, we assume that there are two independent sources for within-subject correlation of the response: the subject-specific REs b_{ij} and the dependence of the response on the subject's death time S_{ij}, i.e. we assume b_{ij} and S_{ij} to be independent. This formulation, while allowing for within-subject correlation, also achieves informative censoring in truncation by death, through the dependence of the response on the death time S_{ij}. For more details on generation of the binary response Y_{ijk} based on the two independent sources of within- subject correlation for simulation studies, we refer the readers to Web Appendix D. Briefly, the binary response is generated based on a continuous latent outcome, denoted by Y_{ijk}^*, which is modeled unconditional on the subject-specific random effects, jointly with survival time. The means of the response conditional on the random effects are recovered using the bisection algorithm. Fifth, the current modeling assumes that once adjusted for the patients'

baseline risks, readmission risks of high- and low-risk index hospitalizations are not different and the TDP setting can not adjust for length of stay. This issue is further discussed in details in the Discussion section and in Web Appendix C.

Finally, we model a partly conditional target of inference, which characterizes the 30-day readmission outcome among dialysis facilities conditional on the patients being alive and does not include patient death as another joint outcome resulting from the patient care process. This is an important limitation in the proposed modeling, since if patient care within facilities leads to differences in mortality rates, this is not directly accounted for in the current TDP set-up. To include death as part of the outcome of care, joint modeling of a longitudinal outcome (e.g. 30 day readmission rate) and survival can be considered for TDP. This extension would require extensive methodological developments to address computational challenges in estimation and inference for a high-dimensional parameter space and is identified as a potential direction for future research.

2.1 Standardized Dynamic Readmission Ratio

To assess the time-dynamic performance of the ith facility relative to a reference and account for patient case-mix, we introduce the standardized dynamic readmission ratio

$$SDRR_i(t) = \frac{\sum_{j \in \mathbb{N}_{it}} p_{ij}(t)}{\sum_{j \in \mathbb{N}_{it}} p_{ij,M}(t)}, \quad (1)$$

where $p_{ij,M}(t) = g^{-1}\left\{\gamma_M(t) + b_{ij} + Z_{ij}^{\mathrm{T}}\beta\right\}$ with $\gamma_M(t)$ denoting the cross-sectional median of $\left\{\gamma_1(t), \ldots, \gamma_I(t)\right\}$, $\mathbb{N}_{it}$ denoting the time-varying index set of all subjects who are alive and receiving care at facility i at time t and t representing time after initiation of dialysis. Similar to model (M1), the proposed $SDRR_i(t)$ also targets the performance of the ith facility for the dynamic cohort of survivors. The $SDRR_i(t)$ is the ratio of the expected total number of readmissions for all patients at facility i at time t relative to the expected total number of readmissions for the same patients based on the reference (national) norm. The denominator is the expected total number of readmissions for an "average" facility (taken over the population of all facilities) at time t, adjusted for the particular case-mix of the *same patients* alive at time t in facility i. Note that it is the sum of readmission probabilities of all patients alive at time t in facility i based on a national norm, specified by $\gamma m(t)$. The definition of $SDRR_i(t)$ requires that $\mathbb{N}_{it}$ is not an empty set, i.e. facility i treats some patients at time t. In our data application with about 2, 900 dialysis facilities, the number of patients treated throughout the three year follow-up does not get smaller than 5. We emphasize again that the risk adjustment (term $Z_{ij}^{\mathrm{T}}\beta$) to account for differences in patients' health status at baseline is required to ensure that variations in patient outcomes among facilities are attributable to the process of care rather than to the differences in the characteristics of the patients.

Estimates of $SDRR_i(t)$ provide key patterns of time-varying patient outcome for each dialysis facility relative to a reference norm and provide a graphical display of the time

periods of under-performance during dialysis. Time periods (t) for which $SDRR_i(t) > 1$ indicate that 30-day readmission rates for facility i are greater than the reference norm. We caution that some practitioners or consumers may be tempted to compare this measure between two facilities (i and i'): $SDRR_i(t)$ versus $SDRR_{i'}(t)$. This is generally wrong because the distributions of patient case-mix between any two facilities may not be the same; a comparison of $SDRR_i$(t) with $SDRR_{i'}(t)$ (i.e., between two facilities) would only be meaningful to the extent that their distributions of case-mixes, Z_{ij} and $Z_{i'j}$, overlap.

A natural estimator of $SDRR^i(t)$ is $\widehat{SDRR}_i(t) = \sum_{j \in \mathbb{N}_{it}} \hat{p}_{ij}(t) / \sum_{j \in \mathbb{N}_{it}} \hat{p}_{ij,M}(t)$ are predicted subject-specific random effects obtained from the means of the posterior distributions of b_{ij}. Estimators of the model parameters β, $\gamma 1\ (t),\ldots,\ \gamma I$(t) and predicted random effects, will be obtained via an approximate EM algorithm due to the large number of facilities, alternating between the estimation of $\gamma_i(t)$, β, and the predicted random effects until convergence as outlined in the next section. The technique is motivated by the work of He et al. (2013).

2.2 Estimation Procedure

To develop the intuition behind the proposed estimation algorithm, first consider a simpler logistic MVCM without subject-specific random effects,

$$g\left[E\left\{Y_{ij}(t)\middle|Z_{ij}, t < S_{ij}\right\}\right] = g\left\{p_{ij}(t)\right\} = \gamma_i(t) + Z_{ij}^{\mathrm{T}}\beta, \quad \text{(M2)}$$

which we refer to as Model 2. Also, this model will be used, in comparison to Model 1, to evaluate the impact of ignoring within-subject correlation. Let $L_{ij}\{\gamma_i(t),\beta\}$ denote the joint distribution of the response $\left(Y_{ij1},\ldots,Y_{ijN_{ij}}\right)$ observed at $t_{ij} = \left(t_{ij1},\ldots,t_{ijN_i}\right)$, such that $t_{ij} < S_{ij}$, given the case-mix Z_{ij}. Then the likelihood function is $L\left\{\gamma_1(t),\ldots,\gamma_I(t),\beta\right\} = \prod_{i=1}^{I}\prod_{j=1}^{N_i} L_{ij}\left\{\gamma_i(t),\beta\right\}$. For estimation in Model 2, facility-specific fixed effect functions $\gamma_i(t)$ can be approximated locally and the derived parameters can be estimated via maximization of the local likelihoods. However, when the number of facilities is large (several thousands), maximizing the local likelihoods poses a serious computational challenge. Standard estimation techniques, such as a global estimation procedure through basis expansions (e.g., via spline approximations), is infeasible due to the high-dimension of the design matrix. Nevertheless, the likelihood $L\{\gamma_1(t),\ldots,\ \gamma_I(t),\ \beta\}$ is separable into I components, where the ith component $\prod_{j=1}^{N_i} L_{ij}\left\{\gamma_i(t),\beta\right\}$ depends only on $\{\gamma_i(t),\ \beta\}$. Hence, given β, $\gamma_i(t)$ can be estimated by maximizing the local likelihood based on data only from the ith facility; and given $\{\gamma_1(t),\ldots,\ \gamma_I(t)\}$, β can be estimated based on a global likelihood without the need for localization. Therefore, an iterative Newton-Raphson algorithm can be implemented for estimation in Model 2. (See Web Appendix A for details.)

Now consider Model 1, the proposed logistic MVCM (M1) with subject-specific REs to account for within-subject correlation. The estimation of Model 1 will similarly rely on the idea of iterating between estimation of the model parameters to deal with the computational

challenges due to the large number of parameters for thousands of dialysis facilities. Let $b = (b_{ij} : i = 1,\ldots,I; j = 1,\ldots,N_i)^{\mathrm{T}}$ denote the vector of independent and identically distributed (i.i.d.) normal variables with mean zero and variance σ_b^2; $b_{ij} \sim N\left(0, \sigma_b^2\right)$. Further let $L_{ij}^*\{\gamma_i(t), \beta\}$ denote the joint distribution of the response $\left(Y_{ij1}, \ldots, Y_{ijN_{ij}}\right)$ observed at t_{ij}, conditional on b_{ij}, $t_{ij} < S_{ij}$ and the case-mix Z_{ij}. Under the independence of b_{ij} and S_{ij}, the joint distribution of $\left(Y_{ij1}, \ldots, Y_{ijN_{ij}}, b_{ij}\right)$ given the case-mix and $t_{ij} < S_{ij}$, denoted by $L\{b_{ij}, \sigma_b, \gamma_i(t), \beta\}$, is given as

$$L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} = L_{ij}^*\left\{\gamma_i(t), \beta\right\} \times \frac{\exp\left\{-b_{ij}^2/\left(2\sigma_b^2\right)\right\}}{\sqrt{2\pi\sigma_b^2}}.$$

Viewing the subject-specific REs as missing data, we propose an approximate EM algorithm. The complete likelihood for Model 1 is $L\left\{b, \sigma_b, \gamma_1(t), \ldots, \gamma_I(t), \beta\right\} = \prod_{i=1}^{I} \prod_{j=1}^{N_i} L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\}$. The incomplete or observed likelihood is

$$L\left\{\sigma_b, \gamma_1(t), \ldots, \gamma_I(t), \beta\right\} = \prod_{i=1}^{I} \prod_{j=1}^{N_i} \left[\int_{-\infty}^{\infty} L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} db_{ij}\right].$$

Also, let Y denote the vector of all outcomes Y_{ijk}, and Y_{ij} denote the vector of all outcomes for the patient j in facility i. The posterior distribution of b_{ij} given Y_{ij} and $\{\gamma_i(t), \beta, \sigma_b\}$ is

$$D_{ij}\left\{b_{ij} \middle| Y_{ij}, \sigma_b, \gamma_i(t), \beta, t_{ij} < S_{ij}\right\} = \frac{L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\}}{\int_{-\infty}^{\infty} L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} db_{ij}}.$$

Hence the posterior mean and variance of $b_{ij0} \equiv C_{ij}^{-1} \int_{-\infty}^{\infty} b_{ij} L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} db_{ij}$ and $v_{ij0} \equiv C_{ij}^{-1} \int_{-\infty}^{\infty} \left(b_{ij} - b_{ij0}\right)^2 L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} db_{ij}$, respectively, where $C_{ij} = \int_{-\infty}^{\infty} L_{ij}\left\{b_{ij}, \sigma_b, \gamma_i(t), \beta\right\} db_{ij}$. We use a Gauss-Hermite quadrature calculation with 20 quadrature points to numerically approximate b_{ij0} and v_{ij0} and utilize a working independence assumption in approximating $L_{ij}^*\left\{\gamma_i(t), \beta\right\}$ by $\prod_{k=1}^{N_{ij}} \left[\exp\left\{\gamma_i\left(t_{ijk}\right) + b_{ij} + Z_{ij}^{\mathrm{T}}\beta\right\} Y_{ijk}\right] / \left[1 + \exp\left\{\gamma_i\left(t_{ijk}\right) + b_{ij} + Z_{ij}^{\mathrm{T}}\beta\right\}\right]$ conditional on the subject-specific REs. Note that the approximation will work well if a weak correlation is introduced due to sharing a death time S_{ij} and that the subject- specific REs account for the majority of the within-subject correlations in the response. The working independence assumption will also be used in the M-step of the algorithm. As explained by Kurland and Heagerty (2005), generalized estimating equations or likelihood- based approaches with non-independence working correlation will not yield valid inference for the partly conditional target.

The E-step of the approximate EM algorithm pertains to the calculation of the conditional expectation of the complete log-likelihood. Let $\{\sigma_b^*, \gamma_1^*(t), \ldots, \gamma_I^*(t), \beta^*\}$ denote the current parameter estimates; b_{ij0}^* and v_{ij0}^* denote the posterior mean and variance of b_{ij} given the current parameter estimates; and let $p_{0,ijk}^* \equiv g^{-1}\{\gamma_i^*(t_{ijk}) + b_{ij0}^* + Z_{ij}^{\mathrm{T}}\beta^*\}$, $q_{0,ijk}^* \equiv 1 - p_{0,ijk}^*$ and $\ell_{ij}\{b_{ij}, \sigma_b, \gamma_i(t), \beta\} \equiv \log L_{ij}\{b_{ij}, \sigma_b, \gamma_i(t), \beta\}$. Because the closed form for $E[\log L\{b, \sigma_b, \gamma_1(t), \ldots, \gamma_I(t), \beta\} | Y, \sigma_b^*, \gamma_1^*(t), \ldots, \gamma_I^*(t), \beta^*, t < S] = \sum_{i=1}^{I}\sum_{j=1}^{N_i} E[\ell_{ij}\{b_{ij}, \sigma_b, \gamma_i(t), \beta\} | Y_{ij}, \sigma_b^*, \gamma_i^*(t), \beta^*, t < S_{ij}] \equiv \mathscr{E}$, the expected log-likelihood, is not available, we approximate $\ell_{ij}\{b_{ij}, \sigma_b, \gamma_i(t), \beta\}$ using a second order Taylor's expansion about b_{ij0}^* to obtain

$$\mathscr{E} \approx \sum_{i=1}^{I}\sum_{j=1}^{N_i}\left(\sum_{k=1}^{N_{ij}}\left[Y_{ijk}\{\gamma_i^*(t_{ijk}) + b_{ij0}^* + Z_{ij}^{\mathrm{T}}\beta^*\} + \log(q_{0,ijk}^*) - \frac{v_{ij0}^*}{2}p_{0,ijk}^* q_{0,ijk}^*\right]\right. \quad (2)$$
$$\left. - \frac{(b_{ij0}^*)^2 + v_{ij0}^*}{2(\sigma_b^*)^2} - \frac{1}{2}\log\{2\pi(\sigma_b^*)^2\}\right) = \sum_{i=1}^{I} L_i\{\sigma_b^*, \gamma_i^*(t), \beta^*\},$$

where $L_i\{\sigma_b^*, \gamma_i^*(t), \beta^*\}$ is defined implicitly. The M-step maximizes the expectation of the complete log-likelihood utilizing the approximation in (2). A key observation is that this approximation is separable into I components, namely $L_i\{\sigma_b^*, \gamma_i^*(t), \beta^*\}, i = 1, \ldots, I$, where the ith component depends only on $\{\sigma_b^*, \gamma_i^*(t), \beta^*\}$ and the posterior mean b_{ij0}^* and variance, v_{ij0}^*, similar to the likelihood of Model 2. Hence, while joint maximization with respect to all model parameters is a substantial computational challenge for a large number of facilities, an iterative alternating approach similar to the one discussed for Model 2 can be feasibly implemented. In the proposed sequential algorithm, we estimate σ_b and β by maximizing the approximation of the expected global log-likelihood given in (2). For estimation of $\gamma_i(t)$, we maximize the approximate expected local log-likelihood (derived by plugging the local linear approximation of $\gamma_i(t)$ into $L_i\{\sigma_b^*, \gamma_i^*(t), \beta^*\}$) using data from only the ith facility. We begin by estimation of σ_b, followed by the estimation of $\gamma_i(t)$ given the current estimates β^*, σ_b^*, and (b_{ij0}^*, v_{ij0}^*) via a one-step Newton-Raphson iteration. Finally, β is estimated utilizing a one-step Newton-Raphson iteration. The algorithm is summarized in Web Appendix A.

2.3 Hypothesis Testing for Fixed Effects: Identifying Extreme Facilities

In addition to estimation, statistical inference about the time-varying readmission rate for dialysis facilities, $\{SDRR_i(t)\}_{i=1}^{I}$, is of interest. More precisely, it is of interest to identify facilities that deviate from the national norm over time. For facilities with time-varying readmission rates not different from the national norm, $SDRR_i(t)$ will be a constant function equal to 1 across time t. Time periods during which $SDRR_i(t) < 1$ or $SDRR_i(t) > 1$ would

indicate that the facility's readmission rates are less than or greater than expected, respectively. Thus, the null hypothesis $H_0 : SDRR_i(t) = 1$ for all t is of interest.

However, making statistical inference about the large number of functions, $\{SDRR_i(t)\}_{i=1}^{I}$, is a non-standard problem and is challenging given the computational cost to estimate thousands of facility-specific varying coefficient functions. To reduce the computational burden, we take advantage of the fact that β, $\gamma_M(t)$ and σ_b can be estimated quite precisely based on data from all facilities. Hence, these quantities are estimated only once and fixed throughout the proposed algorithm which is based on resampling the subject-specific random effects and the responses under the null hypothesis. Since the global parameters β, $\gamma m(t)$ and σ_b are fixed, model refitting to the resampled data only requires estimation of facility-specific parameters b_{ij0}, V_{ij0} and $\gamma_i(t)$. This reduces the computational time substantially, since model fits to the resampled data are carried out using data from each facility separately. Explicit steps of the testing procedure including the test statistic are outlined in Web Appendix B.

3. Time-Dynamic Patterns of Hospital Readmission Among Dialysis Facilities

Our study was conducted using the USRDS, which collects data on nearly all patients receiving care for ESRD, including data on patient demographics and baseline factors prior to the initiation of dialysis. The analysis/study cohort included 113,764 patients receiving dialysis care at 2,896 facilities for a maximum follow-up of three years. Patients experienced a total of 381,922 index discharges with an overall 30-day readmission rate of 29.7%. Details of the study cohort, risk adjustment variables and inclusion/exclusion rules are provided in Web Appendix C. The number of patients per facility varied between 20 to 146.

Overall model fits.

We summarize results from fitting Model 1, the main model, for estimation of the facility-specific time-dynamic 30-day readmission index, SDRR (t), as well as hypothesis testing results from testing the null $H_0 : SDRR_i(t) = 1$ to identify/flag facilities deviating from the national norm over time. The bandwidths used in estimation of $\gamma_i(t)$ were allowed to change across facilities and were selected using 10-fold cross-validation, as described in the simulation results presented in Web Appendix D, where the selected bandwidth values ranged between .8 and 1.3 years. The variance of the subject-specific random effects in Model 1 was estimated to be of $\hat{\sigma}_b^2 = .63$, leading to relatively small to medium differences in the estimated SDRRs. We also compared flagging patterns of Model 1 with Model 2, which ignores within-subject correlation. Both models included a risk adjustment step to account for age, sex, body mass index (BMI), 23 comorbidities and presence of high- risk discharge during the year prior to dialysis and whether diabetes was the cause of ESRD. (The effects of these baseline risk factors on 30-day readmission are given in Web Appendix Table 2.)

Relative time-dynamic hospital readmission patterns.

We identified distinct relative patterns of hospital readmission over time by testing the null hypothesis $H_0 : SDRR_i(t) = 1$ under Model 1, which accounts for within-subject correlation. Note that when H_0 is rejected and the performance of the dialysis facility is found to be significantly different than the national norm, there are three patterns to consider. The identified/flagged facility, relative to the national norm, may be (1) consistently worse over time $\widehat{SDRR}_i(t) > 1$, (2) consistently better over time (i.e. $\widehat{SDRR}_i(t) < 1$ (3) mixed pattern over time, with some time periods of worse or better or not different over the observation time period. Figure 1 illustrates the time-dynamic patterns of 30-day readmission for selected facilities with varying sizes (Model 1 fits). The first row of Figure 1 provides examples of dialysis facilities flagged as significantly worse than the national norm. Also illustrated in Figure 1 are some facilities with SDRR significantly better (row 4), mixed patterns (row 3), and not different (row 2), relative to the national norm. As expected, estimated SDRRs were more variable for smaller facility size (lower volume) for all patterns of SDRR. When the SDRR demonstrates a mixed pattern over time, further investigation will be needed if one is interested in identifying specific time periods of significant over-performance, or under-performance. Note that the proposed hypothesis testing algorithm can be extended to test for significantly worse or better rate of readmission in specific time periods during follow-up by restricting the support of the testing period to these pre-specified time intervals.

Model 1 and Model 2 comparison and flagging rates.

The estimated SDRRs from the two models (with and without accounting for within-subject correlation) were compared using the normalized L2 distance $\left[\int_0^3 \left\{\widehat{SDRR}_{1i}(t) - \widehat{SDRR}_{2i}(t)\right\}^2 dt\right]^{1/2} / \left[\int_0^3 \widehat{SDRR}_{1i}^2(t) dt\right]^{1/2}$ where $\widehat{SDR}_{1i}(t)$ and $\widehat{SDRR}_{2i}(t)$ denote the estimated $SDRR_i(t)$ under Models 1 and 2, respectively. The distributions of the normalized differences between the two estimated SDRRs are displayed in Figure 2(d), stratified by small, medium and large facilities (20–30, 31–42 and 43–146 patients, respectively). Figure 2(d) shows that the difference in SDRRs decreased with increasing facility size as expected since the effects of the predicted random effects on estimated $SDRR_i(t)$ shrink as the predicted probabilities were averaged over larger number of patients within a facility.

The pairwise comparison of the number and percentage of outlier facilities identified by the nominal p-values calculated under Models 1 and 2 are displayed in Table 1(A). Since the proposed test statistic in Section 2.3 is one-sided, the categorization of an outlier facility as better, worse, or mixed was determined by the shape of its estimated $SDRR_i(t)$. Models 1 and 2 identified many more facilities as significantly worse or mixed (12.5% and 15.5% in Models 1 and 2, respectively), than significantly better (1.7% and 1.2% in Models 1 and 2, respectively). The two models did not always identify the same facility as significantly different from the national norm, which was to be expected. For example, 22 out of 125 facilities flagged as significantly worse under Model 1 were not flagged under Model 2. Note that consistent with the decline in the estimated differences in SDRRs between the two models with increasing facility size (Figure 2(d)), the proportion of agreement in flagging results also increased with increasing facility size (results not shown).

Overdisperson, empirical null distribution, and facility volume.

We also considered the empirical null adjustment, discussed in Kalbfleish and Wolfe (2013) and He et al. (2013) (based on Efron (2004)), that accounts for the overdispersion when simultaneously testing a large number of facilities. Accounting for this overdispersion allows for identification of a targeted proportion (e.g., 2.5%) of outlying facilities to further investigate. The degree of overdispersion may depend on the facility size where large facilities will be flagged more frequently. Therefore, the adjustment can be stratified by facility size. The adjustment procedure, similar to He al al. (2013), converts the nominal p-value for each facility to a z-score using the inverse cumulative distribution function (see Web Appendix C for details).

The results are given in Figure 2(a)-(c), which show the histograms of the z-scores, stratified by facility size. Evidently, the distributions are left skewed due to outlier facilities with large negative z-scores. The overdispersion was highest for large facilities, consistent with the findings in Kalbfleisch and Wolfe (2013) who first considered adjustment based on the empirical null for standard (time-invariant) profiling. Also provided in Figure 2(a)-(c) is the density function of a $N(0,1)$ superimposed onto each histogram of the z-scores for reference, as well as a normal density fitted to each histogram representing the empirical null distribution. Since the means of all three empirical null distributions were negative, identifying the 2.5% left-tail of the empirical null distributions for flagged facilities, resulted in much fewer facilities compared to using nominal p-values < .025. Table 1(B) compares the flagging rates of worse or mixed facilities using the nominal p-value and the empirical null distribution. While 12.5% of facilities were flagged as worse or mixed using the nominal p-value, 3.9% were flagged using the empirical null method. The number of flagged facilities were similarly reduced for Model 2. However, Model 1 flagged fewer facilities overall than Model 2, based on the empirical null, consistently for all facility sizes. As expected, a higher percentage of large facilities were flagged using the nominal p-value (16.6%), compared to small (9.3%) and medium (11.7%) facilities. Estimation of the empirical null separately by facility size, reduced the difference in the amount of flagged facilities in each category (2.2%, 4.3%, and 5.2% of the small, medium and large facilities were flagged, respectively).

4. Simulation Studies

We studied estimation of model parameters, including estimation of the key time-varying metric $SDRR_i(t)$, the effects of ignoring within-subject correlation, the validity of the hypothesis testing procedure and the overdispersion when simultaneously testing a large number of facilities through simulation studies. We briefly summarize the main results here where details of the simulation designs and results are deferred to Web Appendix D. The mean squared error (MSE) and the relative mean squared deviation error (MSDE), used to assess estimation of time-invariant and time-varying model parameters, respectively, are given in Table 2, which show that the model parameters were well estimated. Figure 3 illustrates how the estimates track the true $SDRR_i(t)$ (given for Model 1). Given in Table 3 are the estimated levels (acceptance probabilities) of the facility hypothesis test for an outlier facility deviating more from other facilities with varying δ. While the proposed estimation

procedure for Model 1 targeted the parameter estimates and $SDRR_i(t)$ effectively leading to valid inference for identifying outlier facilities, Model 2 which ignores the within-subject correlation resulted in reduced validity (levels substantially less than the 0.95 target). While the first part of Table 3 reports estimated levels of the hypothesis testing from multiple null hypotheses, the second part reports estimated power values. The power increases rapidly with increasing δ and is most effected by facility size rather than the total number of facilities, since the test is largely based on data within a facility. Finally our simulation on overdispersion justified carrying out the empirical null adjustment stratified according to facility size, since a disproportionate number of large facilities were flagged compared to small and medium facilities.

5. Discussion

In this work we proposed a novel method for time-dynamic profiling and solutions to inferential challenges involving a high-dimensional parameter space. Our application of the methodology to assess time-varying 30-day readmission rate over the entire time period from the start of dialysis identified distinct patterns of SDRR. We believe that the focus on the assessment of patient outcome over time among medical care providers, although challenging, is deserving of more research emphasis. For instance, are there common aspects/factors of the patient care processes among providers that exhibit consistently better patient outcomes over time? The first step to answering such a question is the availability of techniques to assess time-varying effects for providers; our work here is a contribution towards this effort.

There are several important limitations and potential extensions to the proposed modeling approach. First, we directly modeled the 30-day readmission outcome among dialysis facilities conditional on the patients being alive via a partly conditional target of inference. To include death as part of the outcome of care, joint modeling of a longitudinal outcome (e.g. 30 day readmission rate) and survival can be considered for TDP as discussed in Section 2. A second limitation involves the case-mix risk adjustment. A factor affecting the readmission risk following an index hospitalization is the discharge reasons/status of the initial index hospitalization. However due to not adjusting for longitudinal covariates on the causal pathway of the facility performance, our time-varying profiling framework cannot adjust for high-risk status of the index hospitalization. Hence an implicit assumption in the current modeling is that once adjusted for the patient's baseline risks, readmission risks of high- and low-risk index hospitalizations are not different. Nevertheless, a sensitivity analysis can be explored excluding the high-risk index hospitalizations; the sensitivity analysis of excluding the low prevalence (1.1%) of high-risk hospitalizations did not have an impact on the results in our application. We note that although we do not adjust for longitudinal case-mix risk variables, we encourage practitioners to expand the risk adjustment to include measurable time-varying confounders that are independent of the effects of the facility's process of care. Also, the impact of time-dependent confounding requires further research.

Finally, several extensions to the proposed modeling framework may be of interest. The model can be extended to include subject-specific time-varying random effects (e.g., see Yao

et al. 2005). The model may also be extended to include calendar year at the initiation of dialysis, especially for a larger cohort initiating dialysis over longer calendar year periods. Also, alternative ways of defining the benchmark for assessing provider performance, other than the national median/average standard, deserves attention. For example, an alternative for defining the benchmark could be based on an external standard of possibly time-specific deviations from the national medium facility performance. The deviation could be based on an "acceptable" tolerance level or on a desired level of patient outcome improvement.

6. Supplementary Materials

Web appendices referenced in Sections 2–4 are available with this article at the Biometrics website on Wiley Online Library.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful for the valuable comments from two referees, an associate editor and the editor. This study was supported by research grants from the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK092232 - DS, DVN, KK, CMR, YC; K23 DK102903 - CMR, KK, DVN). The interpretation and reporting of the data presented here are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the United States government.

References

Ash AS, Fienberg SE, Louis TA, Normand ST, Stukel TA, and Utts J (2011). Statistical issues in assessing hospital performance. The COPSS-CMS White Paper.

Biswas P, and Kalbfleisch JD (2008). A risk-adjusted CUSUM in continuous time based on the Cox model. Statistics in Medicine 27, 3382–3406. [PubMed: 18288785]

Cleveland WS, Grosse E, and Shyu WM (1991). Local Regression Models. In Statistical Models in S (Chambers, J. M. and Hastie, T. J.). Wadsworth & Brooks, Pacific Grove.

Centers for Medicare & Medicaid Services (CMS) (2014). Report for the standardized readmission ratio. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/MeasureMethodologyReportfortheProposedSRRMeasure.pdf Accessed June 5, 2016.

CMS (2016). CMS Quality Strategy. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf Accessed June 5, 2016.

Crainiceanu CM, Staicu AM, and Di CZ (2009). Generalized multilevel functional regression. Journal of the American Statistical Association 104, 1550–1561. [PubMed: 20625442]

Codman E (1916). Hospitalization standardization. Surgery, Gynecology, and Obsterrics 22, 119–120.

Efron B (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. Journal of the American Statistical Association 99, 96–104.

Estes J, Nguyen DV, Dalrymple LS, Mu Y and Şentürk D (2014). Cardiovascular event risk dynamics over time in older patients on dialysis: A generalized multiple-index varying coefficient model approach. Biometrics 70, 754–764. [PubMed: 24766178]

Estes J, Nguyen DV, Dalrymple LS, Mu Y and Şentürk D (2015). Time-varying effect modeling with longitudinal data truncated by death: Conditional models, interpretations and inference. Statistics in Medicine 35(11), 1834–1847. [PubMed: 26646582]

Fan J, and Zhang W (2008). Statistical methods with varying coefficient models. Statistics and its Interface 1, 179–195. [PubMed: 18978950]

Hastie T, and Tibshirani R (1993). Varying coefficient models. Journal of the Royal Statistical Society B 55, 757–796.

He K, Kalbfleisch JD, Li Y, and Li Y (2013). Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. Lifetime Data Analysis 19, 490–512. [PubMed: 23709309]

Hoover DR, Rice JA, Wu CO, and Yang LP (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85, 809–822.

Horwitz L, Partovain C, Lin ZQ, Herrin J, Grady J, Conover M, Montague J, Dillaway C, Bartcazk KL, Ross J, Bernheim S, Drye E, and Krumholz H (2011). Hospital-wide (all-condition) 30 day risk-standardized readmission measure. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/downloads/MMSHospital-WideAll-ConditionReadmissionRate.pdf Accessed June 16, 2016.

Kalbfleisch JD, and Wolfe RA (2013). On monitoring outcomes of medical providers. Statistics in Biosciences 5, 286–302.

Kurland BF and Heagerty PJ (2005). Directly parameterized regression conditioning on being alive: Analysis of longitudinal data truncated by deaths. Biostatistics 6, 241–258. [PubMed: 15772103]

Liang H, Wu H, and Carroll RJ (2003). The relationship between virologic and immuno-logic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. Biostatistics 4, 297–312. [PubMed: 12925523]

Normand ST, Glickman ME, and Gatsonis CA (1997). Statistical methods for profiling providers of medical care: Issues and applications. Journal of the American Statistical Association 92, 803–814.

Normand ST, and Shahian DM (2007). Statistical and clinical aspects of hospital outcomes profiling. Statistical Science 22, 206–226.

Qu A, and Li R (2006). Quadratic inference functions for varying coefficient models with longitudinal data. Biometrics 62, 379–391. [PubMed: 16918902]

Şentürk D, Dalrymple LS, Mohammed SM, Kaysen GA, and Nguyen DV (2013). Modeling time varying effects with generalized and unsynchronized longitudinal data. Statistics in Medicine 32, 2971–2987. [PubMed: 23335196]

Serban N (2011). A space-time varying coefficient model: The equity of service accessibility. The Annals of Applied Statistics 5, 1699–2264.

Sun RJ, and Kalbfleisch JD (2013). A risk-adjusted O-E CUMSUM with monitoring bands for monitoring medical outcomes. Biometrics 69, 62–69. [PubMed: 23383682]

United States Renal Data System (2015). 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.

Van Rompaye B, Eriksson M, and Goetghebeur E (2015). Evaluating hospital performance based on excess cause-specific incidence. Statistics in Medicine 34(8), 1334–1350. [PubMed: 25640288]

Yao F, Muller H-G, and Wang J-L (2005). Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association 100, 577–590.

Zhang D (2004). Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics 60, 8–15. [PubMed: 15032768]

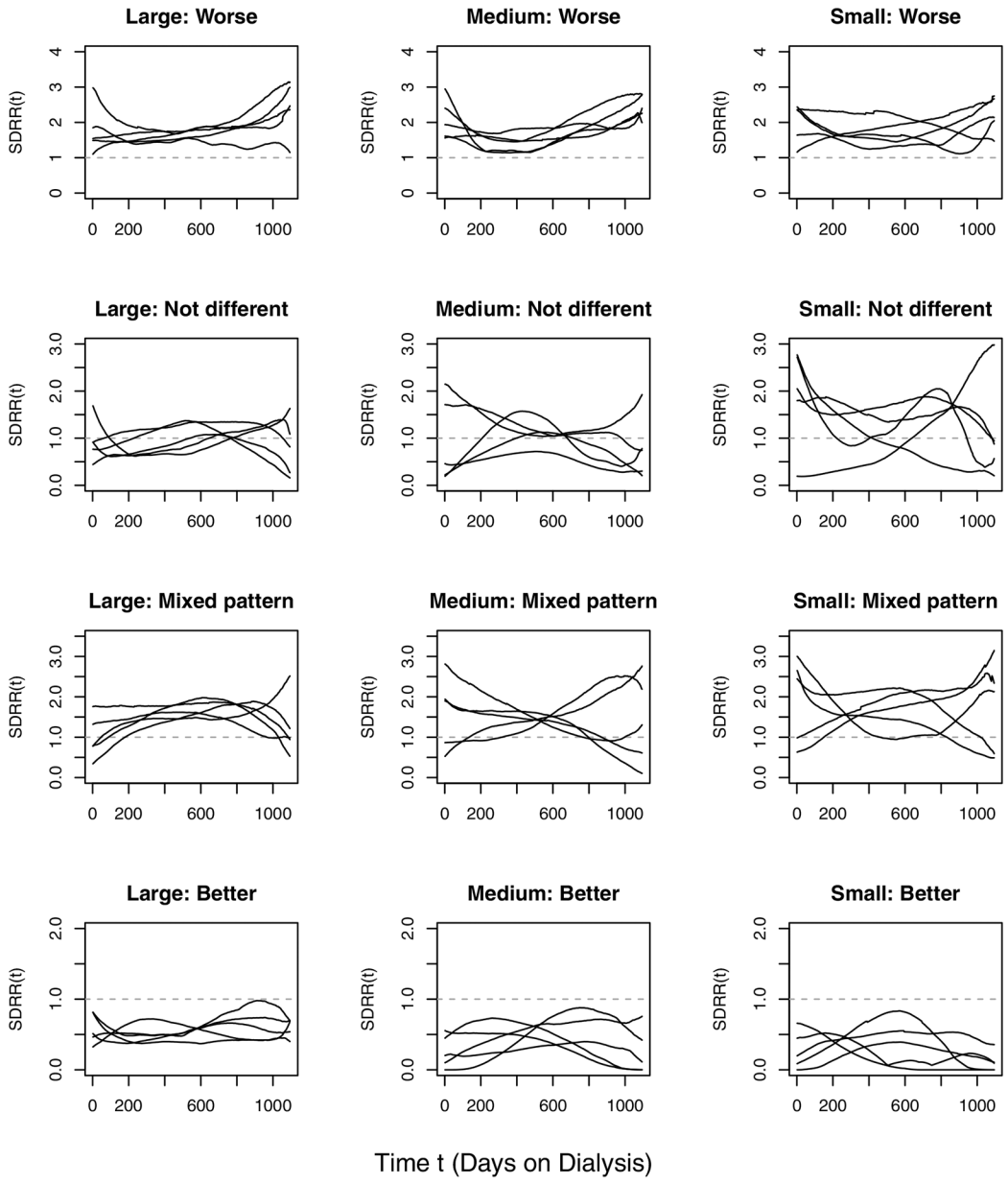

Figure 1:

Illustrative time-dynamic patterns of estimated standardized dynamic readmission ratio $SDRR_i(t)$ for several dialysis facilities flagged as significantly worse (row 1), better (row 4), mixed (row 3) and not significantly different than the national norm (row 2) with varying size (small, medium, large). Displayed are 5 distinct dialysis facilities per plot.

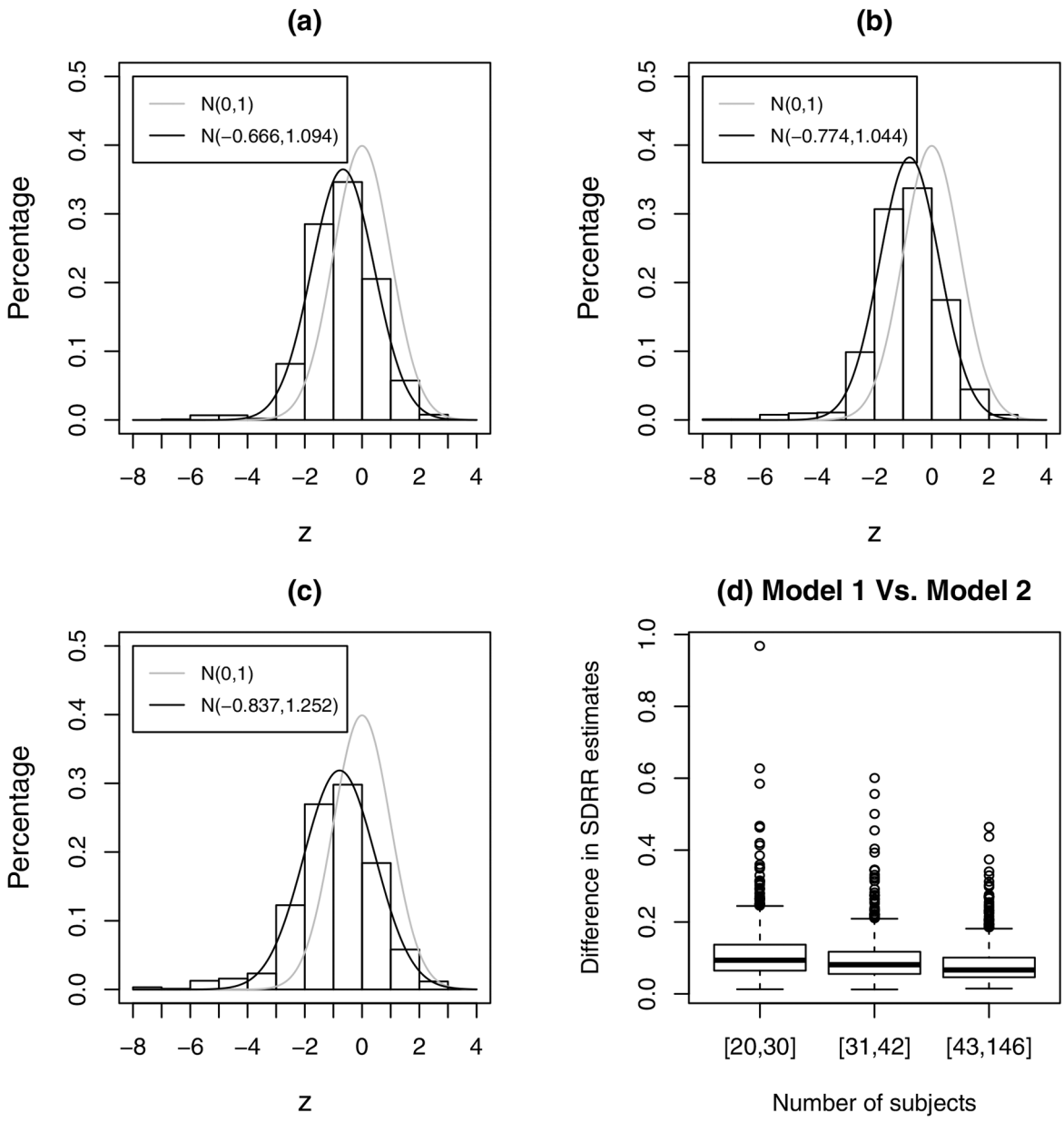

Figure 2:

Histograms of z-scores estimated from p-values of Model 1 in testing H_0: $SDRR_I(t) = 1$, stratified by facility size (a) small, (b) medium, and (c) large. The standard normal density is superimposed on the histograms (grey) along with normal densities (empirical null distribution) fitted to the center of the histograms using a robust M-estimation procedure. (d) Normalized L2-norm of the difference between the estimated facility $SDRR_I(t)$ functions under Model 1 and 2, stratified by facility size.

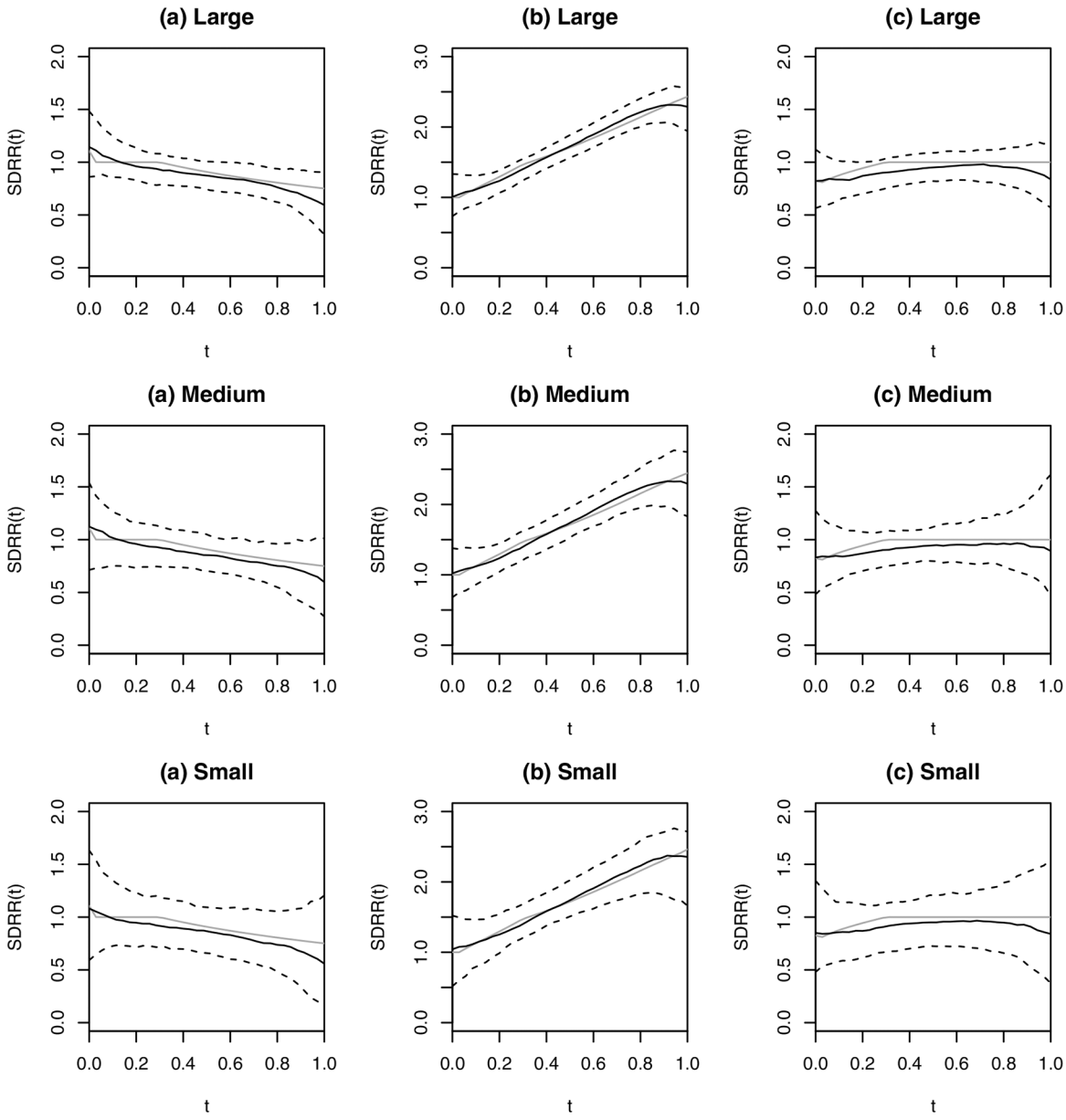

Figure 3:
Estimated $SDRR_i(t)$ functions at the median (black), 10th and 90th percentiles (dashed) of mean squared deviation error (MSDE), stratified by facility size (small, medium and large). The corresponding true average $SDRR_i(t)$ is given in gray. Columns (a), (b) and (c) correspond to the shape of the underlying facility-specific effects $\gamma_i(t)$ as a square root, quadratic or constant function, respectively. See Web Appendix D for details.

Table 1:

(A) Number (percent) of facilities identified as significantly different (better, worse, mixed) from expectation (national norm) over time, determined by nominal p-value < .025. (B) Number (percent) of facilities flagged as worse or mixed based on the nominal p-value and the empirical null distribution, stratified by facility size (small, medium, large).

(A)

	Model 1				
Model 2	**Non-significant**	**Better**	**Worse**	**Mixed**	**Total**
Non-significant	2305 (79.6%)	16 (0.6%)	22 (0.8%)	67 (2.3%)	2410 (83.2%)
Better	6 (0.2%)	28 (1%)	0 (0%)	2 (0.1%)	36 (1.2%)
Worse	53 (1.8%)	0 (0%)	91 (3.1%)	9 (0.3%)	153 (5.3%)
Mixed	123 (4.2%)	4 (0.1%)	12 (0.4%)	158 (5.5%)	297 (10.3%)
Total	2487 (85.9%)	48 (1.7%)	125 (4.3%)	236 (8.1%)	2896 (100%)

(B)

	Model 1		**Model 2**	
# of Subjects	**Nom. p-value**	**Emp. null**	**Nom. p-value**	**Emp. null**
Small [20,30]	96 (9.3%)	23 (2.2%)	114(11.1%)	44 (4.3%)
Med. [31,42]	108 (11.7%)	40 (4.3%)	134 (14.5%)	40 (4.3%)
Large [43,146]	157 (16.6%)	49 (5.2%)	202 (21.4%)	82 (8.7%)
Overall	361 (12.5%)	112 (3.9%)	450 (15.5%)	166 (5.7%)

Table 2:

(A) Estimated bias, standard error (SE), mean squared error (MSE) of the time-invariant parameter estimates and (B) quartiles of the mean squared deviation error (MSDE) of the time-varying estimates of $\gamma_1(t)$ and (C) $SDRR_1(t)$ based on 200 Monte Carlo runs. MSDE are given by overall (all), stratified by facility size (small, medium and large) and by time-varying coefficient function shape (constant, square root and quadratic).

	***I* = 100**						***I* = 1000**					
(A)	Model 1			Model 2			Model 1			Model 2		
Estimate	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\hat{\beta}_1$	.008	.066	.004	−.058	.058	.007	.008	.021	.001	−.057	.019	.004
$\hat{\beta}_2$	−.014	.063	.004	.052	.055	.006	−.006	.021	< .001	.058	.018	.004
$\hat{\sigma}_b^2$	−.025	.041	.002	—	—	—	−.023	.013	.001	—	—	—
(B) $\hat{\gamma}(t)$	Model 1			Model 2			Model 1			Model 2		
MSDE	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
All	.055	.116	.241	.055	.109	.209	.055	.117	.242	.056	.108	.211
Small	.101	.208	.391	.093	.179	.333	.101	.210	.401	.093	.180	.335
Medium	.063	.119	.227	.058	.111	.200	.063	.124	.235	.060	.112	.203
Large	.035	.067	.122	.038	.072	.123	.034	.064	.123	.038	.070	.120
Constant	.052	.105	.218	.052	.103	.193	.052	.107	.211	.053	.101	.191
Square root	.042	.085	.168	.046	.089	.163	.041	.084	.166	.046	.089	.162
Quadratic	.089	.182	.351	.078	.150	.282	.087	.185	.361	.077	.151	.281
(C) $\widehat{SDRR}(t)$	Model 1			Model 2			Model 1			Model 2		
MSDE	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
All	.012	.024	.048	.010	.021	.041	.011	.024	.048	.010	.021	.041
Small	.019	.040	.077	.017	.034	.066	.019	.039	.075	.017	.034	.063
Medium	.012	.025	.046	.011	.022	.040	.012	.025	.048	.011	.022	.041
Large	.008	.015	.027	.007	.014	.024	.007	.014	.026	.007	.013	.023
Constant	.015	.031	.058	.013	.026	.049	.016	.031	.057	.014	.027	.049
Square root	.017	.033	.061	.015	.029	.052	.017	.033	.062	.015	.028	.053
Quadratic	.007	.013	.026	.007	.013	.023	.006	.013	.025	.006	.012	.023

Table 3:

Median mean squared deviation error (MSDE) of $SDRR_i(t)$ estimates and Part I: acceptance probabilities (AP) in testing $H_0 : \gamma_1(t) = \gamma_{0\delta}(t)$ and Part II: estimated power (P) in testing $H_0 : \gamma_1(t) = \gamma_m(t)$, from Section 4. Results are reported from 500 Monte Carlo runs and grouped by facility size (small, medium and large).

Part I

$I = 100$

	Small				Medium				Large			
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 2	
δ	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP
0	.057	.943	.049	.810	.039	.962	.034	.744	.021	.971	.018	.767
.25	.042	.954	.036	.835	.031	.969	.026	.809	.018	.979	.015	.787
.50	.037	.969	.029	.852	.023	.975	.017	.873	.016	.963	.012	.841
.75	.031	.980	.023	.911	.022	.978	.014	.883	.015	.971	.009	.875
1	.024	.965	.017	.882	.020	.971	.011	.878	.012	.963	.007	.870

$I = 1000$

	Small				Medium				Large			
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 2	
δ	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP	MSDE	AP
0	.055	.960	.045	.850	.034	.954	.028	.796	.022	.954	.018	.752
.25	.041	.958	.033	.838	.028	.968	.022	.834	.019	.984	.015	.808
.50	.033	.964	.024	.870	.022	.976	.016	.872	.016	.970	.013	.830
.75	.028	.972	.020	.874	.020	.970	.014	.890	.015	.968	.010	.846
1	.023	.968	.016	.850	.018	.964	.011	.872	.014	.956	.008	.860

Part II

$I = 100$

	Small				Medium				Large			
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 2	
δ	MSDE	P	MSDE	P	MSDE	P	MSDE	P	MSDE	P	MSDE	P
0	.057	.039	.047	.096	.035	.033	.029	.118	.024	.038	.020	.111
.25	.047	.161	.036	.391	.032	.214	.024	.498	.022	.397	.017	.665
.50	.040	.617	.027	.901	.026	.841	.018	.968	.018	.981	.014	.998
.75	.031	.942	.023	.991	.021	1	.018	1	.015	1	.015	1
1	.026	1	.020	1	.019	1	.018	1	.015	1	.017	1

$I = 1000$

	Small				Medium				Large			
	Model 1		Model 2		Model 1		Model 2		Model 1		Model 2	
δ	MSDE	P	MSDE	P	MSDE	P	MSDE	P	MSDE	P	MSDE	P
0	.056	.035	.049	.107	.037	.054	.031	.124	.023	.026	.019	.105
.25	.046	.141	.032	.382	.031	.227	.023	.527	.021	.371	.015	.701
.50	.037	.618	.024	.869	.028	.835	.018	.970	.018	.967	.013	.994
.75	.028	.964	.022	.990	.019	1	.018	1	.016	1	.015	1

1	.025	1	.019	1	.018	1	.017	1	.013	1	.017	1