INQUIRY: The Journal of Health Care Organization, Provision, and Financing
Volume 57: 1–9

DOI: 10.1177/0046958020919275
journals.sagepub.com/home/inq
SAGE

Assessing the Impacts of Misclassified Case-Mix Factors on Health Care Provider Profiling: Performance of Dialysis Facilities

Yi Mu, PhD[1,2], Andrew I. Chin, MD[3,4], Abhijit V. Kshirsagar, MD, MPH[5,6], and Heejung Bang, PhD[7]

Abstract

Quantitative metrics are used to develop profiles of health care institutions, including hospitals, nursing homes, and dialysis clinics. These profiles serve as measures of quality of care, which are used to compare institutions and determine reimbursement, as a part of a national effort led by the Center for Medicare and Medicaid Services in the United States. However, there is some concern about how misclassification in case-mix factors, which are typically accounted for in profiling, impacts results. We evaluated the potential effect of misclassification on profiling results, using 20 744 patients from 2740 dialysis facilities in the US Renal Data System. In this case study, we compared 30-day readmission as the profiling outcome measure, using comorbidity data from either the Center for Medicare and Medicaid Services Medical Evidence Report (error-prone) or Medicare claims (more accurate). Although the regression coefficient of the error-prone covariate demonstrated notable bias in simulation, the outcome measure—standardized readmission ratio—and profiling results were quite robust; for example, correlation coefficient of 0.99 in standardized readmission ratio estimates. Thus, we conclude that misclassification on case-mix did not meaningfully impact overall profiling results. We also identified both extreme degree of case-mix factor misclassification and magnitude of between-provider variability as 2 factors that can potentially exert enough influence on profile status to move a clinic from one performance category to another (eg, normal to worse performer).

Keywords

CMS-2728, measurement error, medical evidence form, misclassification, USRDS, medicare claims, profiling

What do we already know about this topic?
Reliability of comorbidity data as case-mix factors adjusted in health policy models has been questioned and its impact of misclassification on profiling has been studied outside dialysis.
How does your research contribute to the field?
Misclassification on case-mix using different data sources did not meaningfully impact profiling results in dialysis practice.
What are your research's implications toward theory, practice, or policy?
Center for Medicare and Medicaid Services (CMS) may continue to use the current sources of comorbidity data in profiling purposes, but still need to monitor extreme degree of case-mix factor misclassification and magnitude of between-provider variability that can potentially influence profile status in end-stage renal disease (ESRD).

[1]Actelion Pharmaceuticals US, Inc., South San Francisco, CA, USA
[2]A Janssen Pharmaceutical Company of Johnson & Johnson
[3]Division of Nephrology, University of California, Davis School of Medicine, Sacramento, USA
[4]Division of Nephrology, Sacramento VA Medical Center—VA Northern California Health Care System, Mather Field, USA
[5]UNC Kidney Center, Chapel Hill, USA
[6]Division of Nephrology and Hypertension, University of North Carolina, Chapel Hill, USA
[7]Department of Public Health Sciences, University of California, Davis, USA

Received 28 December 2018; revised 6 February 2020; revised manuscript accepted 23 March 2020

Corresponding Author:
Heejung Bang, Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, One Shields Avenue, Med Sci 1C, Davis, CA 95616, USA.
Email: hbang@ucdavis.edu

Introduction

With the availability of increasingly large amounts of patient outcome data and the growing interest in measuring quality of patient care delivered by health care providers, quantitative metrics have been developed to profile hospitals, dialysis clinics, and even individual providers. Much is at stake for individual facilities as well as organizations, whose profiles are used to compare against national averages or norms in the United States, and may result in reduced reimbursement for services for sub-par performance, increased inspection by regulators, and continuous surveillance for quality assurance.[1,2] Therefore, there is a growing interest on ensuring the validity of the metric, ascertainment of patient characteristics and comorbidities, and statistical methods from which these profiles are developed.[3-5] One major concern is the impact of misclassification of case-mix factors, typically used as adjustment variables, on the outcome of interest.

In the United States, the majority of end-stage renal disease (ESRD) patients on dialysis are covered by the Center for Medicare and Medicaid Services (CMS), a federal health insurance program. For this population, Medicare claims and the Medical Evidence report (the CMS-2728 form) represent the 2 primary data sources for comorbidity determination that are presently used in health care policy and research in ESRD. Two main uses in practice are its use in Quality Incentive Program (QIP) and epidemiology research via its availability in US Renal Data System (USRDS). The comorbidity information available on the CMS-2728 form, a data form that is unique to the ESRD population, is a list of known patient comorbidities at incidence of dialysis. These data, *not meant for direct reimbursement claims*, are entered at the dialysis facility by the physician, nurse, or administrative staff based on hospital and ambulatory care medical records. Center for Medicare and Medicaid Services (via University of Michigan—Kidney Epidemiology and Cost Center [UM-KECC]) methodologies for profiling the USRDS dialysis facilities are based on the previous year's claims data. Comorbidity assessment from claims data, captured from diagnostic (ICD) and procedure codes (CPT), is generally considered more reliable than assessment based on information available on the CMS-2728 form, which is required to be completed *once* at incidence of dialysis.[6,7] However, CMS-2728 data are still used for health care policy development because they are much easier to access and process, compared to the resources required to create claims-based models.

However, there has been a concern for many years regarding the accuracy of data in CMS-2728.[8,9] Earlier studies have attempted to validate comorbid conditions reported on CMS-2728 versus clinical data or claims data; the results showed sensitivity <0.6, specificity >0.9, agreement and Kappa statistics <0.5.[7,8,10-12] On the other hand, case-mix adjustment based on administrative claims data (compared to more reliable medical records) is generally considered suitable for profiling hospital performance.[13] In other words, using information garnered from claims data in case-mix profile development models appear to have acceptable quality. With this background, we decided to assess the impact of misclassification in case-mix factors on profiling in dialysis.

In this article, we compared dialysis profiling results using comorbidity data from Medicare claims versus CMS-2728 with 30-day standardized readmission ratio (SRR) as the outcome metric. In addition, we conducted simulation studies to examine the potential effect of misclassification on the estimation of regression coefficients in the statistical models used in the development of profiling strategies as well as profiling itself. We sought to check if real data analysis and simulation study provide consistent results and messages.

Methods

Underlying Models

CMS has employed a hierarchical logistic regression exchangeable model for profiling health care providers.[14,15] Given binary outcome Y_{ij} for the jth patient and discharge in the ith provider ($i = 1, \ldots, n; j = 1, \ldots, n_i$), and case-mix factors $[X, Z]$, the model can be written in a simple form:

$$\text{logit}\left(p_{ij}\right) = \log\left(p_{ij} / \left(1 - p_{ij}\right)\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \gamma_i, \quad (1)$$

where $p_{ij} = P(Y_{ij} = 1 \mid X_{ij}, Z_{ij}, \gamma_i)$, the provider-specific intercepts or random effect are $\gamma_i \sim i.i.d.N(0, \sigma^2)$, and X and Z are accurately measured covariates. The CMS model adopted in practice may be written with error-prone W in place of generally unmeasured X:

$$\text{logit}\left(p_{ij}\right) = \beta_0^{ME} + \beta_1^{ME} W_{ij} + \beta_2^{ME} Z_{ij} + \gamma_i^{ME}, \quad (2)$$

where the superscript ME denotes measurement error and indicates parameters to be estimated with observed covariate W. When X is categorical, for example, X is a binary variable such as true baseline comorbidity status (1 = *yes*, 0 = *no*), ME is often called misclassification, and the relationship between X and W may be quantified via sensitivity (SN) and specificity (SP)[16,17]:

$$\text{SN} = P(W_{ij} = 1 | X_{ij} = 1),$$

$$\text{SP} = P(W_{ij} = 0 | X_{ij} = 0).$$

We assume that W only depends on X, not Y, that is, ME is non-differential.[18]

Profiling Schemes

SRR for the ith provider can be defined as:

$$\widehat{SRR_i} = E_i^* / E_i = \frac{\sum_{j=1}^{n_i} p_{ij}^*}{\sum_{j=1}^{n_i} p_{ij}} = \frac{\sum_{j=1}^{n_i} H\left(\gamma_i + \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij}\right)}{\sum_{j=1}^{n_i} H\left(\beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij}\right)}$$

where $H(v) = (1+\exp(-v))^{-1}$ is the logistic function. Here, E_i denotes the "expected" outcome rate based on fixed effect parameters $(\beta_0, \beta_1, \beta_2)$, and E_i^* denotes the "predicted" outcome rate based on both fixed effect parameters $(\beta_0, \beta_1, \beta_2)$ and provider-specific random effect (γ_i) in model (1). Let E_i^{ME} and E_i^{*ME} be the corresponding estimations based on model (2) with X replaced by W. Bootstrap algorithm for profiling providers was proposed by CMS.[1,19] We obtained 95% confidence interval (CI) of the SRR for the ith provider: profiling as "worse" (ie, under performance) if lower 2.5% limit >1; "better" if upper 2.5% limit <1; and "normal" otherwise. For our simulation later, a provider was assigned to true "worse" if $\gamma_i > 2.5\%$ upper limit of theoretical CI given random intercepts $\gamma_i \sim N(0, \sigma^2)$; true "better" if $\gamma_i < 2.5\%$ lower limit of theoretical CI; otherwise true "normal."[20]

To assess the profiling performance, we focused on the 2 evaluation criteria—profiling sensitivity and specificity. Of note, the identification of truly "worse" providers could be of particular importance as they could face financial penalty in the form of reduced reimbursement for services rendered.

Sensitivity (SN) for profiling as worse providers is

$$\mathrm{SN}_{worse} = \mathrm{P}(\text{SRR profiling as worse} \mid \text{True worse}), \text{ and}$$

Specificity (SP) for profiling as worse providers is

$$\mathrm{SP}_{worse} = \mathrm{P}(\text{SRR profiling as better or normal} \mid \text{True better or normal})$$

where "Normal" performance implies "No reduction in payment," when quality linked to payment.[21]

USRDS Example

In this section, we conducted a case study using 30-day unplanned hospital readmission (namely, SRR) as the profile outcome. We analyzed SRR and the subsequent effects on dialysis facility rating scores using either Medicare claims or CMS-2728 (see Supplemental Table S3), the 2 commonly used sources of comorbidity data in nephrology. We wanted to determine if case-mix adjustment using different data sources would alter the final dialysis facility rating.

CMS utilizes 2-stage model: the first stage of the model is a double random effect logistic regression model where both dialysis facilities and hospitals are modeled as random effects; the second stage is a mixed effect logistic regression model to calculate SRR when profiling dialysis facilities, in which dialysis facilities are modeled as fixed effects and hospitals are modeled as random effects with its standard deviation estimated from the first stage. For each index hospitalization, past year comorbidity based on Medicare claims were grouped as the Hierarchical Condition Categories, see Supplemental Table S4 for the list.

In this analysis, we assessed misclassification under a simplified model only including dialysis facilities as random effects. This random intercept logistic regression model was used by CMS for hospital-wide readmission measure, and we followed the set of guidelines provided by CMS for data processing.[19,22] The algorithm to assign index discharges and unplanned post-index readmission within 30-day from index discharge was derived from the hospital-wide all-cause unplanned readmission measure, and we modeled the case-mix-adjusted 30-day SRR. For case-mix, we adjusted the following factors: age, sex, body mass index, primary cause and years of ESRD, duration of index hospitalization, and a total of 11 comorbidities (alcohol dependence; drug dependence; tobacco use; diabetes; cancer; chronic obstructive pulmonary disease; and cardiovascular diseases including atherosclerotic heart disease, congestive heart failure [CHF], cerebrovascular disease, peripheral vascular disease, and other cardiac).[8,10]

The dialysis facility profile that used claims data prior year to dialysis initiation was regarded as the reference standard.[10] We compared it against the 2 alternative approaches using comorbid conditions captured from CMS-2728: (1) using CHF as recorded on CMS-2728, while all other conditions from claims, and (2) using all of 11 comorbidities from CMS-2728. We chose 11 comorbidities as in previous studies on concordance of data in CMS-2728 and claims.[8,10] These 11 comorbidities on CMS-2728 can be compared with those with ICD-9 codes. The other variables such as "institutionalization" does not have ICD-9 codes. Also, CHF is among the important risk factors in kidney disease (https://nccd.cdc.gov/CKD/Calculators.aspx) and its prevalence is not only relatively high but also differs substantially between the 2 data sources (57% based on claims and 39% based on 2728 form, to be shown below). We selected CHF to examine the impact of misclassification as an illustrative purpose. Also, the list of the final risk adjusters could differ year to year, as reflected in different years' manuals.[23] Data analyses were carried with SAS® 9.4, following the technical notes from the CMS guidelines.[19]

Among 90 373 elderly patients 67 years old or older captured from the USRDS starting dialysis during July 1, 2006, to June 30, 2009, we extracted hospitalization information during January 1, 2010, to June 30, 2012. After excluding small facilities with 10 or less index discharges, there were 63 142 index discharges corresponding to 20 744 patients discharged from 2740 dialysis facilities. The overall 30-day unplanned all-cause readmission rate was about 29%, similar to 30% national readmission rate in the 2014 Dialysis Report.[22] The number of index discharges per facility showed the mean and median of 23 and 20 with standard deviation of 12.

Table 1 shows that after using CHF information recorded on the CMS-2728 in place of the claims data, the estimated odds ratio for each predictor did not change or only

Table 1. USRDS Case Study: Model Fits with Hierarchical Logistic Regression.

		Model A			Model B			Model C		
Variable	Level	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*
Age at hospitalization	[75, 85)	0.93	0.90-0.97	.001	0.93	0.90-0.97	.001	0.94	0.90-0.98	.002
Ref: [67, 75)	≥85	0.93	0.87-0.98	.009	0.93	0.87-0.98	.009	0.93	0.88-0.99	.02
Time on ESRD (year)	1-2	1	0.90-1.10	.935	1	0.90-1.10	.928	1	0.91-1.10	.987
Ref: <1	2-3	0.99	0.90-1.09	.794	0.99	0.90-1.09	.783	0.99	0.90-1.09	.863
	3-6	0.95	0.87-1.05	.327	0.95	0.87-1.05	.315	0.96	0.87-1.05	.358
Length of stay (day)	5	1.05	1.00-1.11	.063	1.05	1.00-1.11	.068	1.05	0.99-1.11	.076
Ref: <5	6	1.2	1.13-1.28	<.0001	1.2	1.13-1.28	<.0001	1.21	1.13-1.28	<.0001
	> 6	1.33	1.28-1.39	<.0001	1.33	1.28-1.39	<.0001	1.33	1.28-1.39	<.0001
Gender	Male	0.88	0.85-0.91	<.0001	0.87	0.84-0.91	<.0001	0.87	0.84-0.91	<.0001
BMI category	[20, 25)	1.01	0.94-1.09	.754	1.01	0.94-1.10	.709	1.02	0.94-1.10	.696
Ref: <20	[25, 30)	0.99	0.92-1.07	.885	0.99	0.92-1.07	.891	0.99	0.92-1.07	.881
	[30, 35)	0.92	0.85-1.00	.061	0.93	0.85-1.01	.067	0.93	0.85-1.01	.068
	≥35	0.88	0.81-0.96	.004	0.88	0.81-0.96	.004	0.88	0.81-0.96	.005
Diabetes as primary ESRD cause	Y	1.01	0.97-1.06	.581	1.01	0.97-1.05	.677	0.99	0.94-1.05	.769
Alcohol dependence	Y	1.17	0.97-1.42	.106	1.19	0.99-1.45	.07	0.87	0.67-1.14	.32
AHD	Y	1.1	1.05-1.14	<.0001	1.11	1.06-1.15	<.0001	1.04	0.99-1.08	.115
Cancer	Y	0.98	0.93-1.03	.444	0.98	0.93-1.03	.478	0.94	0.89-1.01	.078
CHF	Y	1.12	1.07-1.16	<.0001	1.1	1.06-1.15	<.0001	1.11	1.07-1.16	<.0001
COPD	Y	1.14	1.09-1.19	<.0001	1.15	1.10-1.20	<.0001	1.19	1.12-1.26	<.0001
CBVD	Y	1.07	1.02-1.11	.004	1.07	1.02-1.12	.003	1.07	1.01-1.13	.026
Diabetes	Y	0.98	0.94-1.03	.463	0.99	0.95-1.04	.693	1.03	0.97-1.09	.375
Drug dependence	Y	1.23	0.99-1.54	.064	1.23	0.99-1.54	.064	2	1.25-3.21	.004
Other cardiac	Y	1.01	0.97-1.05	.785	1.02	0.98-1.06	.433	1.08	1.03-1.13	.001
PVD	Y	1.04	1.00-1.08	.051	1.05	1.01-1.09	.023	0.98	0.93-1.03	.467
Tobacco user	Y	1.1	1.03-1.18	.003	1.11	1.04-1.18	.002	1.21	1.10-1.33	<.0001

Note. Models: A = 11 types of comorbidity conditions based on past year claims prior to dialysis initiation. B = Replace CHF from CMS 2728 form. C = Replace all 11 types of comorbid conditions based on CMS-2728 form. USRDS = US Renal Data System; OR = odds ratio; CI = confidence interval; ESRD = end-stage renal disease; BMI = body mass index; AHD = atherosclerotic heart disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; CBVD = cerebrovascular disease; PVD = peripheral vascular disease.

minimally changed in the multiple regression. However, there were 3 facilities whose profile status did change; 2 were upgraded and 1 downgraded in their performance ratings, as seen in Table 2. We further computed the prevalence of CHF, SN, and SP among 2740 facilities and reported the results in Supplemental Table S1. The prevalence dropped from 56.6% using claims data to 38.9% when using CMS-2728. However, the prevalence of CHF among the 2 upgraded facilities remained similar; worse to normal: 86.8% (claims) versus 84.2% (CMS-2728), and normal to better: 64.3% (claims) versus 67.9% (CMS-2728). In contrast, the prevalence of CHF dropped from 100% (claims) to 0% (CMS-2728) in the facility downgraded from normal to worse. This may imply that extreme under-reporting (eg, no recording of a key factor) can make a difference in the end result.

Next, SRR estimates and profiling status were compared when all of the 11 comorbid conditions were obtained from claims data versus CMS-2728. Figure 1 demonstrates that the bootstrapped means of SRR obtained with the 2 data sources were highly correlated $(\rho = 0.99)$. The median value

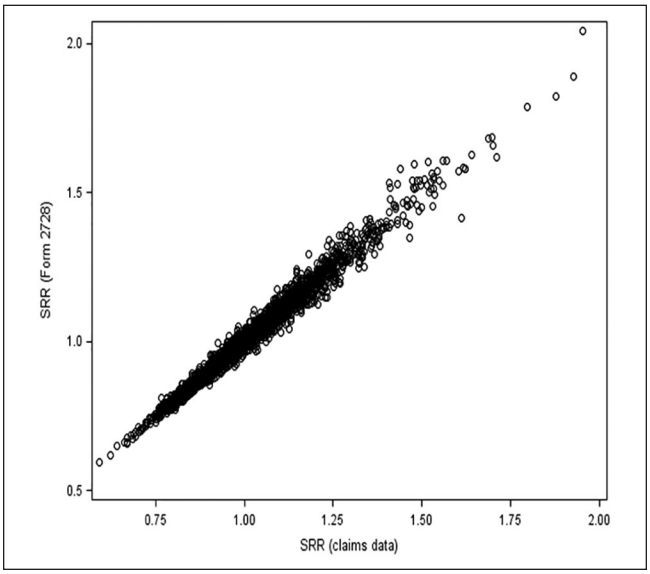

Figure 1. Standardized readmission ratio (SRR) derived from claims data versus CMS-2728 data using bootstrap.
Note. CMS = Center for Medicare and Medicaid Services.

Table 2. USRDS Case Study: Profiling.

	Profile (model B)			Profile (model C)			
Profile (model A)	Better	Normal	Worse	Better	Normal	Worse	Total
Better	3 (0.1%)	0 (0%)	0 (0%)	3 (0.1%)	0 (0%)	0 (0%)	3 (0.1%)
Normal	1 (<0.1%)	2663 (97%)	1 (<0.1%)	0 (0%)	2661 (97%)	4 (0.1%)	2665 (97%)
Worse	0 (0%)	1 (<0.1%)	71 (2.6%)	0 (0%)	8 (0.3%)	64 (2.3%)	72 (2.6%)
Total	4 (0.1%)	2664 (97%)	72 (2.6%)	3 (0.1%)	2669 (97%)	68 (2.5%)	2740 (100%)

Note. Models A = Comorbidity based on past year claims prior to dialysis initiation. B = Replace CHF from CMS 2728 form. C = Replace all 11 types of comorbidity conditions based on CMS-2728 form. USRDS = US Renal Data System.

Table 3. Effect of Misclassification on the Estimation of Fixed Effect Coefficients.

			$\beta_0 = -0.847$ (intercept)				$\beta_1 = 0.5$ (for X)				$\beta_2 = -0.5$ (for Z)			
σ^2	SN	SP	Mean	Var	MSE	CP	Mean	Var	MSE	CP	Mean	Var	MSE	CP
0.2^2	1	1	−0.843	0.002	0.002	0.95	0.499	0.002	0.002	0.95	−0.498	0.002	0.002	0.95
	0.9	0.9	−0.790	0.002	0.005	0.74	0.399	0.002	0.012	0.36	−0.495	0.002	0.002	0.94
	0.5	0.9	−0.656	0.002	0.038	0.00	0.231	0.002	0.075	0.00	−0.493	0.002	0.002	0.94
	0.1	0.9	−0.585	0.001	0.070	0.00	−0.002	0.005	0.257	0.00	−0.491	0.002	0.002	0.94
	0.9	0.5	−0.756	0.003	0.011	0.56	0.241	0.003	0.069	0.00	−0.493	0.002	0.002	0.94
	0.9	0.1	−0.589	0.005	0.072	0.06	0.004	0.005	0.252	0.00	−0.491	0.002	0.002	0.94
	0.5	0.5	−0.585	0.002	0.070	0.00	0.000	0.002	0.252	0.00	−0.491	0.002	0.002	0.94
1	1	1	−0.834	0.011	0.011	0.96	0.496	0.002	0.002	0.94	−0.496	0.002	0.002	0.96
	0.9	0.9	−0.781	0.011	0.015	0.92	0.395	0.002	0.013	0.39	−0.494	0.002	0.002	0.96
	0.5	0.9	−0.649	0.010	0.050	0.52	0.231	0.002	0.075	0.00	−0.491	0.002	0.002	0.96
	0.1	0.9	−0.578	0.010	0.083	0.25	−0.003	0.006	0.259	0.00	−0.490	0.002	0.002	0.96
	0.9	0.5	−0.747	0.011	0.021	0.87	0.239	0.003	0.071	0.00	−0.491	0.002	0.002	0.96
	0.9	0.1	−0.581	0.014	0.085	0.41	0.003	0.006	0.253	0.00	−0.490	0.002	0.002	0.96
	0.5	0.5	−0.578	0.011	0.083	0.27	−0.002	0.002	0.254	0.00	−0.490	0.002	0.002	0.96

Note. SN = 1 and SP = 1 represents no misclassification. Results are based on 1000 simulations. Data are generated from equation (1). SN = sensitivity; SP = specificity; Var = Variance; MSE = mean squared error; CP = coverage probability.

of the relative differences was −0.06 percentage points, with its range in −12.2% to 9.7%. With the reference to claims-based comorbidity adjustment, 8 (out of 72) worse providers were upgraded to normal and 4 normal providers were downgraded to worse when the same model was derived with CMS-2728 ($SN_{worse} = 0.89$ and $SP_{worse} \sim 1.0$); see Table 2.

Simulation Study

We further designed a set of simulation studies to address the 2 objectives: (1) to investigate the effect of misclassification on estimations of fixed coefficients and random intercepts and (2) to compare profiling behavior/performance under different misclassification settings.

Guided by the original CMS model developers, we chose $\beta_0 = \log(3/7)$ to approximate the national readmission rate among dialysis facilities (~30%).[24] X and Z were generated independently from Bernoulli distribution with probability 0.5, with the associated coefficients, $\beta_1 = 0.5$ and $\beta_2 = -0.5$, respectively. The simulations were carried out under a fixed number of providers (100) and a fixed volume $(n_i = 100)$.

The unobserved (X) and observed (W) covariates $[X, W_1, W_2, ..., W_7]$ were generated from multivariate Bernoulli distribution, where $[W_1, W_2, ..., W_7]$ were set to be conditionally independent on X, with varied SN/SP in the 7 misclassification scenarios. We also examined between-provider variability, for example, low and high $(\sigma^2 = 0.2^2$ and $1)$ in equation (1), informed by previous studies.[20,24] Simulations were conducted using R version 3.3.3, including *lme4* and *bindata* packages.[25,26]

The first experiment using 1000 simulations examined the effect of misclassification on regression parameters. From Table 3, when SN or SP for variable X decreased, the estimates for fixed effect parameters (β_0, β_1) tended to be attenuated toward the null—a well-known phenomenon in the ME literature.[17,18,27] Given that the empirical variability of the estimates of (β_0, β_1) was stable across settings under varied SN/SP, the increment in absolute bias in (β_0, β_1) led to the increment in mean squared error. In contrast, regarding the precisely measured Z, which is independent of X, neither bias nor variance in the regression coefficient, β_2, was meaningfully affected by the presence

Table 4. Effect of Misclassification on the Estimation of Coverage Probability for Random Intercepts Based on True Profiling Status.

		$\sigma^2 = 0.2^2$			$\sigma^2 = 1$		
SN	SP	Better	Normal	Worse	Better	Normal	Worse
1	1	0.53	0.96	0.57	0.83	0.93	0.89
0.9	0.9	0.52	0.95	0.56	0.83	0.93	0.88
0.5	0.9	0.51	0.95	0.55	0.82	0.93	0.87
0.1	0.9	0.50	0.95	0.56	0.82	0.93	0.87
0.9	0.5	0.51	0.95	0.56	0.82	0.93	0.87
0.9	0.1	0.50	0.95	0.56	0.81	0.93	0.87
0.5	0.5	0.50	0.95	0.56	0.82	0.93	0.87

Note. 1000 simulations are used. SN = sensitivity; SP = specificity.

Table 5. Effect of Misclassification on Profiling.

			Low variability ($\sigma^2 = 0.2^2$)					High variability ($\sigma^2 = 1$)				
Misclassification		SRR	True profiling			SRR profiling		True profiling			SRR profiling	
SN	SP	profiling	Better	Normal	Worse	SN	SP	Better	Normal	Worse	SN	SP
1	1	Better	0.27	0.72	0	0.11	0.99	2.45	27.77	0	1.00	0.72
		Normal	2.18	93.22	1.81	0.98	0.19	0	37.71	0	0.40	1.00
		Worse	0	1.15	0.65	0.26	0.99	0	29.61	2.46	1.00	0.70
0.9	0.9	Better	0.28	0.67	0	0.11	0.99	2.45	27.67	0	1.00	0.72
		Normal	2.17	93.3	1.83	0.98	0.19	0	37.71	0	0.40	1.00
		Worse	0	1.12	0.63	0.26	0.99	0	29.71	2.46	1.00	0.70
0.5	0.9	Better	0.28	0.67	0	0.11	0.99	2.45	27.63	0	1.00	0.72
		Normal	2.17	93.31	1.84	0.98	0.18	0	37.84	0	0.40	1.00
		Worse	0	1.11	0.62	0.25	0.99	0	29.62	2.46	1.00	0.70
0.1	0.9	Better	0.27	0.65	0	0.11	0.99	2.45	27.61	0	1.00	0.72
		Normal	2.18	93.35	1.84	0.98	0.18	0	37.84	0	0.40	1.00
		Worse	0	1.09	0.62	0.25	0.99	0	29.64	2.46	1.00	0.70
0.9	0.5	Better	0.28	0.66	0	0.11	0.99	2.45	27.66	0	1.00	0.72
		Normal	2.17	93.35	1.83	0.98	0.19	0	37.82	0	0.40	1.00
		Worse	0	1.08	0.63	0.26	0.99	0	29.61	2.46	1.00	0.70
0.9	0.1	Better	0.28	0.67	0	0.11	0.99	2.45	27.65	0	1.00	0.72
		Normal	2.17	93.31	1.83	0.98	0.19	0	37.81	0	0.40	1.00
		Worse	0	1.11	0.63	0.26	0.99	0	29.63	2.46	1.00	0.70
0.5	0.5	Better	0.27	0.66	0	0.11	0.99	2.45	27.63	0	1.00	0.72
		Normal	2.18	93.31	1.83	0.98	0.18	0	37.82	0	0.40	1.00
		Worse	0	1.12	0.63	0.26	0.99	0	29.64	2.46	1.00	0.70

Note. 100 simulations are used. SRR = standardized readmission ratio; SN = sensitivity; SP = specificity.

of misclassification, with the coverage probability (CP) maintained close to desired 95%.

Table 4 summarizes CP based on whether the 95% Wald's CI contains the true value of random intercept for the *i*th provider, grouped by true profiling status. When SN or SP of *X* decreased, CP for random intercepts was stable across the 3 types of true profiling status. However, when σ^2 increased from 0.2^2 to 1, CP for true worse and true better providers increased markedly, while CP for normal decreased minutely, which implies that higher variability improved sensitivity among better or worse.

The second experiment, using 100 simulations, investigated the effect of misclassification on profiling under the same set of simulation parameters as in the first experiment. Simulation findings indicate that profiling results appeared to be robust. The case of $\gamma_i \sim N(0, 0.2^2)$ showed low sensitivity for both true worse (eg, SN 0.26) and true better (SN 0.11) providers, but higher sensitivity for true normal (SN 0.98). To compare, the case of $\gamma_i \sim N(0,1^2)$ showed highest sensitivity for both true worse (SN 1) and true better (SN 1) providers, but lower sensitivity for true normal providers (SN 0.4); see Table 5. We also observed that, under high

between-provider variability, a substantial number of normal performers (~30%) were declared to be worse performers. Such downgrading of clinic ratings may subject those clinics to unjust penalties.[20]

Discussion

In this era of "pay for performance" and initiatives to enhance patient choice in choosing health care, it is important to understand how case-mix adjustments using various data sources can affect the results of profiling health care providers.[1] For patients on dialysis with Medicare coverage and for research purposes, there are 2 major data sources for comorbidity ascertainment in the USRDS: Medical Claims and CMS-2728 Medical Evidence form (incident dialysis comorbidity information). In health care policy, CMS-2728 is used to capture the comorbidities in the development of the standardized mortality ratio (SMR) and standardized hospitalization ratio (SHR), which are the 2 components of the "Dialysis Facility Compare Star Rating" (https://www.medicare.gov/dialysisfacilitycompare/), a program aimed to provide consumers with information when choosing outpatient dialysis services.[28-30] However, the SRR in the ESRD QIP, another program implemented by CMS, used prior year claims data for comorbidity adjustment. Thus, the method for case-mix adjustment in dialysis clinic profiling differs even within the same cohort of ESRD patients and the same operating agency and may change over different years. The QIP has been used for both payment reduction for facilities that underperform and a publicly available online rating on the CMS "Dialysis Facility Compare" Web site to inform consumers.[31,32]

In this study based on both real and simulated data, we found that commonly encountered, moderate miscoding in covariates or case-mix may have limited influence on profiling. This phenomenon might be partly explained by similarity in profiling versus prediction, where there is no strong need for the modeling of ME to play an important role in prediction problems. In contrast, misclassification generally affects the regression coefficients (measure of association) in the statistical model, well explained by mathematical theory; that is, regression dilution.[18]

Between-provider variance can play an important role in the profiling results.[20,33] Simulation results without misclassification in predictor in Table 5 agree with those from a previous study.[20] For true worse or true better providers, simulations suggest low SN (0.11 for true better, 0.26 for true worse)/high SP (0.99) under smaller variance versus high SN (1.0)/not high SP (0.7) under larger variance. For true normal providers, simulations suggest high SN (0.98)/low SP (0.19) under smaller variance versus low SN (0.4)/highest SP (1.0) under larger variance. Given that true worse/better providers were based on upper/lower 2.5% under our simulations (unlike 20% better in Ding et al[33]), profiling based on random intercept model can be more useful under smaller between-provider variance if the goal is to flag out a small percent of outliers, that is, to avoid misclassification of a large number of true normal providers. On the other hand, the case of larger variance showed improved coverage probability overall for the random intercept indicating each provider, and high sensitivity and specificity (in sum as summary measure, which is called the Youden Index). From our USRDS data example, the variance of the random intercepts for facilities on the logit scale was estimated to be $\hat{\sigma}^2 \cong 0.14$. A total of 72 out of 2740 facilities were flagged as "worse," and only 3 facilities were flagged as "better," as presented in Table 2.

In addition, we found that regardless of adjusting comorbidities from either the CMS-2728 form or claims data, SRR estimates from the 2 approaches agreed closely ($\rho = 0.99$; median relative difference = −0.06). In prior studies, investigators also observed that hospital readmission rates developed using different data sources and adjusters were similar.[13,34,35] Along the line, it has also been reported that relative profiling approaches for pay-for-performance were more robust to missing data than absolute profiling approaches,[36] where missing data can be viewed as an extreme, special case of misclassification.

Also, in studies using simulations to evaluate the impact of under-coding of cardiac disease severity on hospital profiles or report cards, investigators found that the outlier status of most hospitals was robust to under-coding. However, miscoding of very influential predictors of mortality, such as shock or renal failure, could lead to a change in the 30-day mortality rate profile.[37]

In our real data analysis example, the prevalence of individual comorbid conditions was lower when taken from the CMS-2728 form (Supplemental Table S2), but similar profiling results were observed with the same statistical model using either data source. However, it was also revealed that profiling status can change in the extreme facilities when misclassification severely varied across providers; see Supplemental Tables S1 and S2. When we replaced 1 covariate (CHF) ascertained from CMS-2728 form, 3 out of 2740 (0.1%) facilities changed the profiling status (facility #1 to #3, Supplemental Table S2). When we replaced all 11 types of comorbidity conditions with different data sources, 12 out of 2740 facilities (0.4%) changed profiling status (facility #2 to #13, Supplemental Table S2). A total of 4 facilities (facility #3 to #6) can newly face penalty when CMS-2728 form (less reliable data source) was used. In CMS dry run of SRR for dialysis facilities, CHF was removed from past-year comorbidity due to its presence in many ESRD patients and modifiability.[23] Our real data analysis (Supplemental Table S2, facility #1, #2, and #3) may suggest a potential flaw in current dialysis facility QIP when using SMR as outcome. Standardized mortality ratio was adjusted for comorbidities from patient's CMS-2728 form, for example, CHF.[23]

There is already existing literature on the agreement between different data sources for comorbidities (eg, CMS 2728 vs claims) in ESRD and on the impact of using different data sources on profiling models outside ESRD. Thus,

we consider our work as the combination of these two, accompanied by "statistical" evaluation (eg, mean squared error, coverage probability and sensitivity/specificity) and the first study of its kind in ESRD. Readers may find our findings are generally supported by theory, empirical real-world data analysis, and statistical simulation (where truth is known), and in agreement with previous related findings. Other unanswered questions include whether the duration of time between dialysis initiation and the CMS-2728 form completion date affects misclassification, and if facilities with dialysis patients of greater vintage (prevalent time on dialysis) may also face more misclassification. The process of data input onto the CMS2728 is extremely variable and done to various degrees of accuracy. It is supposed to be done within 45 days of first dialysis treatment for ESRD, at the dialysis outpatient clinic, not in the hospital. Notably, there is no penalty if the completion and submission of the form to the local dialysis network are delayed. The local dialysis network will generate a form listing the incomplete 2728 submissions. There are no published data frequency of incomplete submissions at 45 days. These could serve as good future research questions.[7,8,10,38,39]

The limitations of our study should be noted. First, in the simulation study, we only considered simple scenarios with limited configurations; for example, misclassifications and size constant across providers, non-differential ME, and 2 covariates. Although simple settings can better elucidate mechanisms and facilitate interpretations, future investigations are warranted under more complicated settings. Second, there are different profiling models besides the CMS model/method that we selected. For example, random versus fixed effects, 2-stage, Cox and piecewise Poisson model, and observed or predicted value (vs expected value in standardized ratios) have been used and results with different policy implications have been obser ved.[1,10,12,24,40] These contradictions can be investigated for further elucidation and possible resolution in future. Third, we did not have a gold standard for comorbidity determination so claims data served as the reference standard, which is currently utilized by CMS for profiling hospitals based on 30-day readmission ratios.[13,19]

Based on simulation and real data example, we conclude that misclassification on covariates can affect regression coefficients in the models used for profiling, but less on profiling itself. However, extreme scenarios (such as in completely missing or omitted data in an important covariate) and between-provider variability can influence and make a difference in the final profile status.

Authors' Note

The interpretation and reporting of the data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the US government.

Acknowledgments

We thank Dr Lorien Dalrymple for her early contribution to the conception of the work.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: H.B. and Y.M. were partly supported by Dialysis Clinic, Inc. H.B. was additionally supported by the National Institutes of Health through grant UL1 TR001860. The interpretation and reporting of the data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the US government.

IRB Approval

The University of California has determined that studies using USRDS data do not constitute human subject research.

ORCID iD

Heejung Bang https://orcid.org/0000-0002-5755-9951

Supplemental Material

Supplemental material for this article is available online.

References

1. Ash AS, Fienberg SF, Louis TA, Normand SLT, Stukel TA, Utts J. Statistical issues in assessing hospital performance. https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/statistical-issues-in-assessing-hospital-performance.Pdf. Accessed December 2019.
2. Normand SLT, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci*. 2007;22:206-226.
3. Martsolf GR, Barrett ML, Weiss AJ, et al. Impact of race/ethnicity and socioeconomic status on risk-adjusted readmission rates: implications for the Hospital Readmissions Reduction Program. *Inquiry*. 2016;53:1-9.
4. Chin DL, Bang H, Manickam RN, Romano PS. Rethinking thirty-day hospital readmissions: shorter intervals might be better indicators of quality of care. *Health Aff (Millwood)*. 2016;35(10):1867-1875.
5. Manickam RN, Mu Y, Kshirsagar A, Bang H. Area-level poverty and excess hospital readmission ratios. *Am J Med*. 2017;130(4):e153-e155.
6. Delva O. Saved and missing CMS-2728 forms could affect ESRD patients' Medicare enrollment benefits. *J Nephrol Soc Work*. 2017;41(2):22-25.
7. Longenecker JC, Coresh J, Klag MJ, et al. Validation of comorbid conditions on the end-stage renal disease medical evidence report: the CHOICE study. *J Am Soc Nephrol*. 2000;11(3):520-529.
8. Mu Y, Chin AI, Kshirsagar AV, Bang H. Data concordance between ESRD Medical Evidence Report and Medicare claims: is there any improvement? *PeerJ*. 2018;6:e5284.

9. Center for Medicare and Medicaid Services (CMS). https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNEdWebGuide/Downloads/MLN-Compliance-Webinar.pdf. Published 2014. Accessed December 2019.
10. Krishnan M, Weinhandl ED, Jackson S, Gilbertson DT, Lacson E Jr. Comorbidity ascertainment from the ESRD Medical Evidence Report and Medicare claims around dialysis initiation: a comparison using US Renal Data System data. *Am J Kidney Dis*. 2015;66(5):802-812.
11. Kim J, Desai M, Chertow G, Winkelmayer W. Validation of reported predialysis nephrology care of older patients initiating dialysis. *J Am Soc Nephrol*. 2012;23(6):1078-1085.
12. Liu J, Krishnan M, Zhou J, Nieman KM, Peng Y, Gilbertson DT. Data completeness as an unmeasured confounder in dialysis facility performance comparison with 1-year follow-up. *Clin Nephrol*. 2016;86(11):262-269.
13. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2011;4(2):243-252.
14. Normand SLT, Glickman ME, Gatsonis C. Statistical models for profiling providers of medical care: issues and applications. *J Am Stat Assoc*. 1997;92(439):803-814.
15. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley; 2011.
16. Slate EH, Bandyopadhyay D. An investigation of the MC-SIMEX method with application to measurement error in periodontal outcomes. *Stat Med*. 2009;28(28):3523-3538.
17. Bang H, Chiu Y-L, Kaufman JS, Patel MD, Heiss G, Rose KM. Bias correction methods for misclassified covariates in the Cox model: comparison of five correction methods by simulation and data analysis. *J Stat Theory Pract*. 2013;7:381-400.
18. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2006.
19. Center for Medicare and Medicaid Services (CMS). Hospital-wide (all-condition) 30-day risk-standardized readmission measure. Produced by Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/downloads/MMSHospital-WideAll-ConditionReadmissionRate.pdf. Published 2011. Accessed December 2019.
20. Yang X, Peng B, Chen R, et al. Statistical profiling methods with hierarchical logistic regression for healthcare providers. *J Appl Stat*. 2014;41:46-59.
21. ESRD QIP—total performance scores—payment year 2019. https://data.medicare.gov/Dialysis-Facility-Compare/ESRD-QIP-Total-Performance-Scores-Payment-Year-201/fqah-awe3. Published 2019. Accessed December 2019.
22. University of Michigan Kidney Epidemiology and Cost Center (UM-KECC). Measure methodology report for the proposed SRR measure. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/MeasureMethodologyReportfortheProposedSRRMeasure.pdf. Published 2014. Accessed August 2017.
23. Center for Medicare and Medicaid Services (CMS). CMS ESRD measures manual for the 2019 performance period. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/ESRD-Manual-v40-.pdf. Published 2019. Accessed December 2019.
24. He K, Kalbfleisch JD, Li Y, Li Y. Evaluating hospital readmission rates in dialysis facilities: adjusting for hospital effects. *Lifetime Data Anal*. 2013;19(4):490-512.
25. Bates B, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1-48.
26. Leisch F, Weingessel A, Hornik K. On the generation of correlated artificial binary data. In: Working Paper Series, SFB "Adaptive Information Systems and Modelling in Economics and Management Science." https://epub.wu.ac.at/286/1/document.pdf. Published 1998. Accessed April, 2020.
27. Wang N, Carroll RJ, Lin X, Gutierrez RG. Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *J Am Stat Assoc*. 1998;93:249-261.
28. https://www.medicare.gov/dialysisfacilitycompare/#data/about-data. Published 2017. Accessed December 2019.
29. Kshirsagar AV, Manickam RN, Flythe JE, Chin AI, Bang H. Area-level poverty, race/ethnicity & dialysis star ratings. *PLoS ONE*. 2017;12(10):e0186651.
30. Kshirsagar A, Alishahi Tabriz A, Bang H, Lee S-YD. Patient satisfaction is associated with dialysis facility quality and star ratings. *Am J Med Qual*. 2019;34(3):243-250.
31. Center for Medicare and Medicaid Services (CMS). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/ESRDQIPSummaryPaymentYears2014-2018.pdf. Published 2018. Accessed December 2019.
32. Center for Medicare and Medicaid Services (CMS). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/ESRDQIPPY2017finaltechnicalmeasurespecifications-.pdf. Published 2018. Accessed December 2019.
33. Ding VY, Hubbard RA, Rutter CM, Simon GE. Assessing the accuracy of profiling methods for identifying top providers: performance of mental health care providers. *Health Serv Outcomes Res Methodol*. 2013;13(1):1-17.
34. Bernheim SM, Parzynski CS, Horwitz L, et al. Accounting for patients' socioeconomic status does not change hospital readmission rates. *Health Aff*. 2016;35(8):1461-1470.
35. Martsolf GR, Barrett ML, Weiss AJ, et al. Impact of race/ethnicity and socioeconomic status on risk-adjusted hospital readmission rates following hip and knee arthroplasty. *J Bone Joint Surg Am*. 2016;98(16):1385-1391.
36. Ryan MA, Bao Y. Profiling provider outcome quality for pay-for-performance in the presence of missing data: a simulation approach. *Health Serv Res*. 2013;48(2, pt 2):810-825.
37. Austin PC, Tu JV, Alter DA, Naylor CD. The impact of under coding of cardiac severity and comorbid diseases on the accuracy of hospital report cards. *Med Care*. 2005;43(8):801-809.
38. Layton JB, Hogan SL, Jennette CE, et al. Discrepancy between medical evidence form 2728 and renal biopsy for glomerular diseases. *Clin J Am Soc Nephrol*. 2010;5(11):2046-2052.
39. Malas MS, Wish J, Moorthi R, et al. A comparison between physicians and computer algorithms for form CMS-2728 data reporting. *Hemodial Int*. 2017;21(1):117-124.
40. University of Michigan Kidney Epidemiology and Cost Center (UM-KECC). Technical notes on the Standardized Mortality Ratio (SMR) for the Dialysis Facility Reports. https://dialysisdata.org/sites/default/files/content/Methodology/SMRDocumentation.pdf. Published 2018. Accessed December 2019.