

## Dealing with Failures of Assumptions in Analyses of Medical Care Quality Indicators with Large Databases Using Clustering

Kenneth Pietz\*, Laura A. Petersen and LeChauncy D. Woodard  
*Michael E. DeBakey VA Medical Center and Baylor College of Medicine*

*Abstract:* The application of linear mixed models or generalized linear mixed models to large databases in which the level 2 units (hospitals) have a wide variety of characteristics is a problem frequently encountered in studies of medical quality. Accurate estimation of model parameters and standard errors requires accounting for the grouping of outcomes within hospitals. Including the hospitals as random effect in the model is a common method of doing so. However in a large, diverse population, the required assumptions are not satisfied, which can lead to inconsistent and biased parameter estimates. One solution is to use cluster analysis with clustering variables distinct from the model covariates to group the hospitals into smaller, more homogeneous groups. The analysis can then be carried out within these groups. We illustrate this analysis using an example of a study of hemoglobin A1c control among diabetic patients in a national database of United States Department of Veterans' Affairs (VA) hospitals.

*Key words:* Cluster analysis, logistic regression, random effects, SAS, NLMIXED.

### 1. Introduction

When an analysis of an indicator of health care quality is presented using data from multiple hospitals or clinics, a commonly asked question is whether the analysis was adjusted for nesting of the patients within the hospitals. The implication is that an analysis that is not adjusted in this way is not valid, because systematic variations between hospitals due to factors such as skill of medical staff, hospital policies and availability of equipment may exist and this variation needs to be taken into account.

The essential part of an analysis of an indicator of quality is a regression model with the indicator as the dependent variable. The independent variables will generally include multiple covariates. To adjust for grouping we have two

---

\*Corresponding author.

choices: the hospitals enter the analysis as fixed effects or as random effects. Each of these analytic approaches carries assumptions.

Fixed effects parameters are ordinarily estimated by maximum likelihood. If there are a large number of hospitals, a model with hospitals as fixed effect requires the model to estimate a large number of parameters. It is known that if we increase the number of hospitals and try to estimate the hospital effect and the other covariates simultaneously, then inconsistent estimates can result (McCulloch *et al.*, 2008, p. 111).

In a random effects model the hospital-level residuals are assumed to be realizations of a random variable, usually taken to be normally distributed. This may not be a reasonable assumption if the hospitals in the model have a wide variety of characteristics. For example, the database may contain large tertiary care hospitals in metropolitan areas as well as smaller rural hospitals. On the other hand, if hospital characteristics are among the covariates, a random effects model is required to avoid collinearity. Many researchers argue in favor of a random effects model, although the computation can consume a large amount of computer resources. However, econometricians tend to shun random effects models because these models assume that the covariates affect all hospitals in the same way (Kennedy, 1998, p. 227).

Rating hospitals according to their performance on quality indicators is a subject of great research interest. If this is the goal, obviously the hospitals must appear in the model in some way. We will not focus on the subject of ranking hospitals, which is discussed in many papers (Goldstein and Spiegelhalter, 1996; Normand *et al.*, 1997; Draper and Gittoes, 2004). This article addresses the situation in which the goal is to investigate how one or more covariates affect the dependent variable (i.e., quality indicator) and how this effect may vary among the hospitals.

We investigated the effect of failure of the assumption that the hospitals are drawn from a normally distributed population, on the parameter estimates in random effects models with a large diverse database. We examined the use of classical cluster analysis to divide the hospitals into smaller, more homogeneous groups as a proposed method of creating an analysis environment in which the assumptions are more likely to be satisfied. Individual analyses can then be done within each cluster. The primary focus of the study was the situation in which a large database with many hospitals is used. While multilevel analysis is discussed in many books and papers, there is a paucity of information on the special problems of very large databases.

## 2. Motivating Example

The data for this example consist of 830,328 diabetic patients who received

care at 105 hospitals in the Department of Veterans Affairs (VA) health care system in the United States. The data were obtained from two national VA databases: the Decision Support System and the National Patient Care Database for each of the hospitals. These data included hemoglobin A1c (HbA1c) level and demographic information on all of the diabetic patients who received care in fiscal year 2007 (October 1, 2006 – September 30, 2007) at the subset of the 137 VA hospitals for which the required data were available. Medication and diagnosis data at the patient level were also included. The diagnosis data were used as input to Diagnostic Cost Group (DCG) software (Verisk Health, 2011). The DCG software calculates a relative risk score (RRS), a measure of the patient’s overall illness burden. The RRS is equal to the patient’s predicted health care costs divided by the average observed health care costs of a VA beneficiary.

Our objective was to determine whether black patients had similar levels of HbA1c control compared to patients of other race groups receiving care for diabetes in the VA. Several other covariates were included as risk adjustment variables. There is an extensive body of research on risk adjustment in medical studies (Pietz *et al.*, 2004; Petersen *et al.*, 2006). The goal of the risk adjustment in this case is to adjust for factors other than race that might obscure the effect of race itself. For example, if mental health conditions among patients with diabetes led to different levels of HbA1c control and if these conditions were less prevalent in black patients, the effect of race might be obscured. The variables and their descriptive statistics are presented in Table 1.

Table 1: Cohort characteristics ( $n = 830,328$ )

Characteristics	$N$ (%)
Male	810,189 (97.6)
African American	112,980 (13.6)
Limited life expectancy	49,149 (5.9)
VA priority group	
1, 4 (high disability)	204,436 (24.6)
2, 3, 6 (low/moderate disability)	152,889 (18.4)
5 (low income)	257,896 (31.1)
7, 8 (lowest priority)	215,107 (25.9)
ICD-9-CM diagnosis code for mental health condition	147,195 (17.7)
ICD-9-CM diagnosis code for overweight or obesity	204,148 (24.6)
Insulin prescription within 100 days of index	151,782 (18.3)
Age (years), mean (SD)	65.9 (10.8)
Diagnostic Cost Group Relative Risk Score, mean (SD)	1.64 (2.7)
No. VA primary care visits in 1 year prior to index, mean (SD)	4.0 (3.5)

### 3. Analysis

The dependent variable examined was glycemic control defined as HbA1c  $\leq 9$ . This was chosen because a level of HbA1c greater than 9 is classified as a VA measure of poor quality (Office of Quality and Performance, 2010). We did the analysis three ways. First we ran a fixed effects model, ignoring the nested structure, i.e., with no hospitals in the model. This model, which will be called model 1, is included for comparison to show the effect of incorrectly ignoring the hierarchical structure. The parameter estimates in this case are known as pooled estimates. It is well known that pooled estimates are consistent but not efficient (Hardin and Hilbe, 2003, p. 33). The model equation is equation (1).

$$y_i | \mathbf{x}_i \sim \text{Bernoulli}(\pi_i), \quad \pi_i = E(y_i | \mathbf{x}_i), \quad \text{logit}(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (1)$$

where  $y_i$ ,  $i = 1, \dots, N$  is a binary variable indicating whether patient  $i$  has HbA1c  $\leq 9$  (with a value of 1 indicating “yes” and 0 indicating “no”), and  $N$  is the number of patients. The column vector  $\boldsymbol{\beta}$  and the row vector  $\mathbf{x}_i^T$  for each patient have length equal to the number of covariates (12 in this case). The vector  $\boldsymbol{\beta}$  is the vector of model parameters estimated from the data and  $\mathbf{x}_i$  is the vector of values of the covariates for patient  $i$ .

Model 2 is a fixed effects model with the facilities included as covariates, along with the risk adjustment variables described in Table 1. The model equation for this analysis is the same as equation (1) except that the row vectors of covariates  $\mathbf{x}_i^T$  and the column vector  $\boldsymbol{\beta}$  now include one covariate for each hospital except one which is left out as a reference. In model 2 the row vectors  $\mathbf{x}_i^T$  and the column vector  $\boldsymbol{\beta}$  each have length 116. For each patient  $i$ , the vector  $\mathbf{x}_i^T$  will have the value 1 for the hospital in which the patient received care and 0 for all others. Following common practice, we used the hospital with median parameter estimate as the reference so that there are approximately as many positive as negative parameter estimates.

Finally, model 3 is a mixed effects model with the hospital as a random effect. The model equation for this case is equation (2).

$$y_{ij} | \mathbf{x}_{ij}, u_j \sim \text{Bernoulli}(\pi_{ij}), \quad \pi_{ij} = E(y_{ij} | \mathbf{x}_{ij}, u_j), \quad \text{logit}(\pi_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_j, \quad (2)$$

where  $u_j \sim \mathcal{N}(0, \sigma^2)$  where  $\sigma^2$  is an unknown variance. In equation (2),  $y_{ij}$ ,  $\mathbf{x}_{ij}$ , and  $\pi_{ij}$  are the response variable, vector of covariate values, and expectation as above, respectively, for patient  $i$  in hospital  $j$ ,  $i = 1, \dots, N_j$ ,  $j = 1, \dots, M$ , where  $N_j$  is the number of patients in hospital  $j$  and  $M$  is the number of hospitals. The column vector  $\boldsymbol{\beta}$  and the row vector  $\mathbf{x}_{ij}^T$  for each  $i$  and  $j$  again have dimension equal to the number of covariates. However in this model, the hospital is not

among the covariates, so the length of  $\mathbf{x}_i^T$  and  $\boldsymbol{\beta}$  is again 12. The hospital enters the model as random effect  $u_j$ .

There are multiple methods available for doing this third analysis. We used the SAS procedure NLMIXED, which uses adaptive quadrature. SAS version 9.2 was used (2009). Adaptive quadrature is widely accepted as the most accurate method, since parameter estimates are generated by maximizing the likelihood directly, rather than by approximating the maximum likelihood (ML) estimates, as in penalized quasi-likelihood (Zhang *et al.*, 2001). Ten quadrature points were used throughout our analysis.

Fixed effects and random effects models belong to the class of subject-specific models. We believe that the research question in this case leads to a subject-specific model rather than a population-averaged model, such as a generalized estimating equations model. This distinction is discussed in (McCulloch *et al.*, 2008, Chapter 9). The research question is, what would the experience of a black patient have been if he/she were of another race? The assumptions of generalized linear models, which also apply to GLMMs, are listed on page 255 of Raudenbush and Bryk (2002). If normality assumptions are violated, parameter estimates are unbiased but estimates of variance components may be inaccurate. However a key assumption of GLMMs is that the effect of parameters is constant across level 2 units. The effect of race on the care of patients with diabetes and the relationship between race and other covariates may not be the same in all types of hospitals in all markets. One solution is to use cluster analysis based on clustering variables independent of the covariates to group the level 2 units into more homogeneous clusters. The analysis can then be done within the clusters. We applied a cluster analysis based on measures developed independently to group hospitals for comparing their performance and efficiency (Byrne *et al.*, 2009). Characteristics that the administration of the hospital cannot easily change, such as size, were chosen. The measures and their weights are listed in Table 2. We used  $k$ -means clustering for this analysis (Everitt *et al.*, 2001).

Overall model fit for the fixed-effects only models was assessed using the likelihood ratio test (LRT), Wald test, and score test for significance of covariates, and the  $c$ -statistic. Model fit for the random effects models was then assessed by chi-squared tests for significance of random effects using Akaike's information criterion (AIC) (Davis, 2002, p. 138).

#### 4. Results

The LRT, Wald, and score tests all rejected the global null hypothesis for significance of covariates at the 0.001 level for all fixed-effects-only models used in this analysis (data not shown). Table 3 shows the vector of parameter estimates

Table 2: VAMC peer groups: domains, measures, sources and weights

Domain	Measure	Source (FY 2007 unless noted)	Weight
Complexity	DRG index	VHA NPCD Inpatient Files	5
	Referral care ratio	VHA NPCD Inpatient Files	1
Size	Total hospital operating beds	VSSC Bed Control File (FY 2008)	5
	Growth in VA-funded users, FY 2005 – FY 2007	DSS National Data Extracts	1
Academic mission	Resident slots per 10,000 patients	VA Office of Academic Affiliations	5
Disease burden	Average patient relative risk score	Derived using DxCG® RiskSmart 2.2	1
Care delivery structure	Ratio of CBOC clinic stops / Total outpatient clinic stops	VHA NPCD Outpatient Files	1
	Ratio of psychiatric beds / Total hospital operating beds	VSSC Bed Control (FY 2008)	1
	Ratio of nursing home and domiciliary beds / Total facility operating beds	VSSC Bed Control (FY 2008)	1
	Polytrauma Center (Yes/No)	VHA Polytrauma website: <a href="http://polytrauma.va.gov/faq.asp#FAQ1">polytrauma.va.gov/faq.asp#FAQ1</a>	1
	SCI mandate (Yes/No)	VHA Directive 2000-022	1
Reliance on the VA	Adjusted VA reliance	VIREC Medicare files; HERC Average Cost Datasets; DSS NDE (FY 2004)	1
Community & environment	Number of hospital beds in community	Area Resource File, 2006 Release (based on 2004 AHA survey)	1
	State-level Medicaid generosity	Urban Institute & Kaiser Commission On Medicaid and the Uninsured analysis of MMIS data (2004-06); HCFE: VA Bedford MA– wage adjustment, 2002 data	1
Infrastructure	Multi-site facility	PSSG-VAST	1
	Total facility square feet per unique patient	VHA Capital Asset Inventory (OFM)	1

Acronyms: CBOC = VA Community-based outpatient clinic; FY = fiscal year; DRG = Diagnosis Related Group; NPCD = National Patient Care Database; VSSC = VISN Support Services Center; DSS NDE = DSS National Data Extracts; HERC = Health Economics Resource Center (VA Palo Alto); PSSG-VAST = Planning Systems Support Group database connected with the VA Site Tracking system; MMIS = Medicaid Management Information System; HCFE = Health Care Financing and Economics (VA Boston HCS); OFM = Office of Facilities Management; AHA = American Hospital Association; VIREC = VA Information and Resource Center.

$\beta$  for the three models for the overall data base. We focus on the parameter estimate for black patients. All three analyses indicate that black patients are less likely to have their HbA1c controlled, even after risk adjustment. The greatest difference observed was between model 1 and the other two models. This is not surprising because model 1 incorrectly ignores the nested structure. Some of

Table 3: Results of fixed and random effects models

Parameter	Model 1 No hospitals		Model 2 Hospital fixed		Model 3 Hospital random	
	Value or parameter estimate	Standard error	Value or parameter Estimate	Standard error	Value or parameter Estimate	Standard error
c-statistic	0.586	-	0.586	-	-	-
AIC	863789	-	863801	-	835927 ( $< 0.001$ )*	-
Intercept	-0.441	0.025	-0.431	0.025	-0.418	0.043
Age	0.020	$< 0.001$	0.021	$< 0.001$	0.020	$< 0.001$
Male	0.108	0.017	0.114	0.017	0.114	0.017
Black	-0.170	0.008	-0.157	0.008	-0.157	0.008
Limited life expectancy	-0.099	0.013	-0.099	0.013	-0.100	0.013
DCG Relative Risk Score	-0.012	0.001	-0.013	0.001	-0.013	0.001
Low income VA priority group	0.037	0.007	0.044	0.007	0.043	0.007
High disability VA priority groups	0.079	0.008	0.100	0.009	0.100	0.008
Low/moderate disability VA priority groups	0.095	0.008	0.108	0.009	0.108	0.009
Insulin use	-0.133	0.007	-0.147	0.007	-0.147	0.007
Mental health conditions	0.084	0.008	0.100	0.008	0.100	0.008
Obesity	0.282	0.007	0.288	0.007	0.287	0.007
Number of primary care visits	0.070	0.001	0.071	0.001	0.071	0.001

AIC = Akaike's information criterion  

*p*-value for significance of random effects

variation in outcomes between values of the remaining covariates is due to variation among hospitals, which are not included in model 1. This explains why the estimates are different although they are consistent. What is surprising is that there is almost no difference between the parameter estimates and standard errors between models 2 and 3, except for the intercept. This is important because model 3, which may be regarded as the gold standard, consumed a large amount of computer time and memory. The chi-squared test using the AIC indicated that model 3 fit the data better.

The preceding analysis does not answer the question of whether the effect of being black on HbA1c control might be different for different types of hospitals. Therefore, we performed the cluster analysis. In *k*-means clustering, the number of clusters must be specified by the analyst. The goal is to derive clusters

of homogeneous groups in which the assumptions of random effects models are satisfied. Specifying a number of clusters that is too small results in large clusters that are not really homogeneous. Specifying too large a number results in very small clusters to which random effects modeling cannot be applied. After some experimentation, we settled on 10 clusters. The individual clusters can be regarded as groups of hospitals with similar characteristics. Table 4 provides the fixed effects and random effects parameter estimates for black race within the clusters. The c-statistics in Table 4 indicate generally better model performance within the clusters than in the model using the entire database.

Table 4: Parameter estimates for black race in clusters

Cluster	Number of hospitals	Fixed effect estimate	Standard error	Model c-statistic	AIC	Random effect estimate	Standard error	AIC	Chi-squared test for random effects
1	9	-0.141	< 0.001	0.625	91804	-0.163	< 0.001	90492	< 0.001
2	4	-0.245	< 0.001	0.599	43159	-0.215	0.010	43094	< 0.001
3	9	-0.102	< 0.001	0.623	85787	-0.161	< 0.001	84875	< 0.001
4	3	-0.114	0.001	0.623	21811	-0.112	0.086	21813	NS
5	29	-0.281	< 0.001	0.657	150912	-0.223	< 0.001	148638	< 0.001
6	9	-0.086	< 0.001	0.619	151353	-0.135	< 0.001	149756	< 0.001
7	9	-0.513	< 0.001	0.659	47397	-0.175	0.005	46256	< 0.001
8	13	-0.077	< 0.001	0.634	114492	-0.094	0.001	112776	< 0.001
9	13	-0.150	< 0.001	0.613	96490	-0.229	< 0.001	95462	< 0.001
10	7	-0.194	< 0.001	0.657	43077	-0.146	0.008	41517	< 0.001

AIC – Akaike’s information criterion

NS = not significant

Table 4 reveals considerable differences in the parameter estimates for black race between the clusters, although all estimates are negative indicating that black patients are at greater risk. This analysis shows that the level of HbA1c control for black patients would be expected to be somewhat different depending on the type of hospital where he/she was a patient. The fixed effects and random effects estimates are also different for the same cluster. We would expect the random effect assumption to be more likely to be satisfied in the larger clusters. The chi-squared test using Akaike’s information criterion (AIC) showed that the random effect was significant for all clusters except cluster 4.

## 5. Discussion

There are some unique features of studies, such as our example, that utilize a very large database. Considering the hospitals in the study as a random sample of a larger population is questionable because those in the study comprise 78% of VA hospitals nationwide and were not randomly selected. We suggest that the goal should be to determine accurately what took place at the hospitals in the study in fiscal year 2007 and whether there are lessons to be learned, rather than interpreting results in terms of a hypothetical super-population.



We assume that the analyst is interested in the effect of being black on a patient's HbA1c control and whether this effect was constant across the VA. It is clear that the hospitals should appear in the model. However, with a large population of hospitals as in this case, it is almost certain that the effect of the covariates will not be constant across hospitals. It is known that when this is so, both fixed effects and random effects models can lead to inaccurate parameter estimates. Applying fixed effects models when the hospital is not the focus of the analysis can mask interactions between the hospital and the covariate of interest (Donner and Klar, 2000). However, if the hospitals are correlated with the covariates in a random effects model, the parameter estimates for the covariates can be seriously biased (McCulloch *et al.*, 2008, Section 12.3b).

Our results clearly indicate that obtaining parameter estimates that are nearly the same from both fixed effects and random effects models does not mean that there is no variation among the hospitals. If the parameter estimate for black race has nearly the same value in the fixed effect model as in the random effects model, then we cannot conclude that there is no interaction of black race with hospital. We note that including interactions of hospitals with other covariates would make the model very complicated and would lead to results difficult to interpret. We recommend using cluster analysis using measures not in the regression model to obtain groups of level 2 units that can be assumed to be exchangeable and repeating the analysis within these groups to obtain a more complete picture of the effect of the covariate. The analyst must then decide whether to apply fixed effects or random effects models within the clusters. There is considerable controversy in the statistical community over this decision. Random effects models are recommended for the larger clusters if the random effect is significant. As discussed in (McCulloch *et al.*, 2008), it is not necessary to assume that the level 2 units are a random sample of a larger population. The key point is whether the variation remaining after all covariates are included is independent and identically distributed, or whether the remaining variation is more accurately described as the sum of two components, a hospital-level component and the remaining variation. A chi-squared test using the AIC provides a reasonable way to answer this question. The assumption that the covariates have the same effect across hospitals is more likely to be viable within the clusters. A reviewer has noted that more complicated distributions for the random effects may be considered (Zhang *et al.*, 2011) if there is a theoretical basis for selecting such distributions. This is an area for future research.

In the example, the cluster analysis shows that there is considerable difference in the HbA1c control of black patients between groups of hospitals with similar characteristics. Cluster analysis provides a convenient way to extract this information from the data. The cluster analysis provides reasonable assurance that

the hospitals within clusters have similar structural characteristics. If desired, further analysis can be conducted with data for individual hospitals. Differences between hospitals within the clusters are more likely to be due to actual quality differences.

## Appendix 1: Definitions of Variables

### *Measures of complexity*

- (1) **DRG Index.** The DRG Index was calculated by multiplying the relative weight of each Diagnosis Related Group (DRG) greater than or equal to 3 by the frequency of that DRGs, and then summing these products. We excluded DRG 542 (for tracheostomy care longer than 96 hours) because it may not represent a level of complexity similar to that of the other highly weighted DRGs. Relative weights were taken from CMS, DRGV24 (released in 2007).
- (2) **Referral care ratio.** This measure is the proportion of a facility's inpatient acute admissions attributed to referral patients. A referral was defined as: 1) a patient discharged from VA facility that is not their nearest "home VA facility" based on ZIP code (Planning Systems Support Group [PSSG] ZIP code database); and 2) the patient had outpatient care at a facility other than the discharging facility within 90 days prior to the patient's admission; and 3) the patient had outpatient care at a facility other than the discharging facility within 90 days following the discharge.

### *Measures of size*

- (3) **Total hospital operating beds.** Source: VISN Service Support Center (VSSC) Bed Control Data File. We calculated the number of operating beds at the end of FY 2008. The count includes internal medicine, intermediate medicine, neurology, blind rehabilitation, psychiatric, rehabilitative medicine, surgery, and spinal cord. (Domiciliary and nursing home care beds were excluded.)
- (4) **Growth in VA-funded users, FY 2005 to FY 2007.** We calculated the increase in VA-funded users (FY 2007 minus FY 2005), expressed as a percentage of the FY 2005 VA-funded users (see below). Those identified in the ADUSH enrollment file with Priority status 1-8 were included.

$$\begin{aligned} & \text{FY 2005 - FY 2007 Growth} \\ &= \frac{(\text{FY 2007 VA-funded Users} - \text{FY 2005 VA-funded Users})}{\text{FY 2005 VA-funded Users}}. \end{aligned}$$

*Measure of academic mission*

- (5) **Ratio of medical resident slots to 10,000 unique patients.** We obtained data for filled medical resident positions during academic year 2007-2008 from the VA Office of Academic Affiliations; these data are available at the VA Allocation Resource Center (ARC) website. The number of unique patients at each facility was obtained from the FY 2007 Decision Support System (DSS) cost files; patients were counted at each facility at which they had DSS costs.

*Measure of disease burden*

- (6) **Average Patient Relative Risk Score (RRS).** This is the average of a facility's patient-level Relative Risk Scores. The individual patient's DxCG® RRS is a ratio of predicted costs to average observed costs:  
The average RRS at a facility is the sum of RRSs of all patients at that facility divided by the number of unique patients at the facility in FY 2007. A RRS value of 1.00 reflects an average illness burden of the national VA population.

We derived diagnostic cost group(DCG)-based RRSs from the FY 2007 Houston VA DCG model for users in priority groups 1-8, then calculated and scaled these RRSs across the VA to yield a population mean of 1. We used DxCG® RiskSmart 2.2 software.

*Measure of patient reliance on the VA*

- (7) **Adjusted VA Reliance.** This measure is designed to capture reliance on the VA for veterans who are eligible for Medicare (including the Medicare disabled population). It is expressed as the mean reliance on the VA among VA-Medicare dually eligible veterans of all ages, adjusted for percent of facility patients who are not Medicare-enrolled:

$$\frac{\text{VA Cost of Medicare enrollees (all ages)}}{(\text{Medicare Cost}) + (\text{VA Cost}) \text{ of Medicare enrollees}}$$

Proportion of facility's veterans (all ages) not enrolled in  
× Medicare (indexed and scaled to be expressed as a ratio  
with a mean of 1).

Here is an example using Station *X*. This station's VA reliance among Fee-for-service Medicare enrollees is 0.6158. Its proportion of veterans not Medicare-enrolled is 0.4805, which is higher than the mean in the VA. After

multiplication and scaling, the adjusted mean reliance on the VA is 0.6744 at Station  $X$ .

Medicare cost was calculated from MEDPAR and carrier files. The VA cost data are from the HERC Average Cost Datasets for FY 2004 and the DSS pharmacy file for FY 2004. Medicare HMO enrollees were excluded from the cost ratio. Note: This is the most recent period for which the Medicare-VA match data are available at VIREC.

#### *Measures of care delivery structure*

- (8) **Ratio of CBOC clinic stops/total outpatient clinic stops.** We used the VA National Patient Care Database (NPCD) to identify visits and Planning Systems Support Group database connected with the VA Site Tracking system (PSSG-VAST data crosswalk) to identify CBOCs.
- (9) **Spinal Cord Injury (SCI) mandate (yes/no).** Data were taken from the VHA Directive 2000-022 entitled "Spinal Cord Injury Center Staffing and Beds."
- (10) **Polytrauma Rehabilitation Site or Network Polytrauma Center (yes/no).** Data were obtained from the VA Blind Rehabilitative webpage identifying VA facilities with blind rehabilitation centers.
- (11) **Ratio of psychiatric beds/total hospital operating beds.** Bed count data were taken from the VSSC Bed Control Data File for September 2008, the last month of FY 2008.
- (12) **Ratio of CLC and domiciliary beds/total facility operating beds.** Data were taken from the VSSC Bed Control Data File, data from September 2008. The denominator (facility operating beds) includes total hospital operating beds (see #3 above), domiciliary, nursing home, and Psychosocial Residential Rehabilitation Treatment Program (PRRTP) beds.

#### *Measures of facility infrastructure*

- (13) **Multi-Site indicator (yes/no).** This variable indicates whether a VA facility is composed of more one site, where each site has an inpatient hospital. We obtained these data from the PSSG-VAST data crosswalk.
- (14) **Square footage per unique patient.** We obtained this variable from the VHA Capital Asset Inventory, Space and Functional Database information provided directly from the VHA Office of Facilities Management in October

2007. Summary data are available at <http://vaww.vhacowebapps.cio.med.va.gov/cis/>. Square footage per unique patient is calculated as station-level gross square footage (GSF) divided by the number of station unique patients.

### *Measures of community environment*

- (15) **Number of hospital beds in community.** From the Area Resource File (ARF, updated 2006), we extracted the number of beds in the facility's designated geographic area. In December 2005, the Office of Management and Budget revised the standards used to determine geographic divisions. To be consistent with our prior work, we used the same county aggregations that comprised geographical divisions (e.g., metropolitan statistical areas [MSAs]). For facilities not located in an MSA, we used the surrounding county.
- (16) **State-level Medicaid generosity.** This is a measure of the availability of other health care benefits for low-income veterans, and is defined as state Medicaid expenditures (from both state and federal Medicaid funds) per state resident under the poverty level, adjusted for the state average hourly wage rate for hospital workers:

$$\frac{\text{Medicaid spending on (adults + elderly + blind/disabled)*}}{\text{Number of poor adults in state}},$$

\*Adjusted for state average wages for hospital workers.

Note: Adults are age 19-64.

Data Sources: Medicaid spending and poverty data are from 2004-2006. Data were accessed at the Kaiser Family Foundation (KFF) website in October 2007:

<http://www.statehealthfacts.org/cgi-bin/healthfacts.cgi?action=rawdata>. The Urban Institute and the Kaiser Commission provide KFF with estimates based on data from Medicaid Statistical Information System (MSIS). States submit MSIS data to the Centers for Medicare and Medicaid Services.

The wage rate adjustment data were obtained from the Health Care Financing and Economics (Boston VAMC) website ([www.hcfe.org](http://www.hcfe.org)). HCFE aggregated 2002 wage rate data from CMS sources. The CMS provides wage rates at the MSA and rural non-MSA levels, and HCFE aggregates them to the state level.

## Acknowledgements

This work was supported in part by VA HSR&D SHP 08-166 (PI Laura A. Petersen, MD, MPH), Houston VA HSR&D IIR 04-349 (PI Laura A. Petersen, MD, MPH), VA HRS&D PPO 09-316 (PI LeChauncy D. Woodard, MD, MPH), NIH R01 HL079173-01 (PI Laura A. Petersen, MD, MPH), and Center of Excellence HFP90-020 (PI Laura A. Petersen, MD, MPH).

## References

- Byrne, M. M., Daw, C. N., Nelson, H. A., Urech, T. H., Pietz, K. and Petersen L. A. (2009). Method to develop health care peer groups for quality and financial comparisons across hospitals. *Health Services Research* **44**, 577-592.
- Donner, A. and Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. Oxford University Press, New York.
- Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York.
- Draper, D. and Gittoes, M. (2004). Statistical analysis of performance indicators in UK higher education. *Journal of the Royal Statistical Society, Series A* **167**, 449-474.
- Everitt, B. S., Landau, S. and Leese, M. (2001). *Cluster Analysis*, 4th edition. Arnold, London.
- Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A* **159**, 385-443.
- Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. Chapman & Hall/CRC, Boca Raton, Florida.
- Kennedy, P. (1998). *A Guide to Econometrics*, 4th edition. p. 227, MIT Press, Cambridge, Massachusetts.
- McCulloch, C. E., Searle, S. R. and Neuhaus, J. M. (2008). *Generalized, Linear and Mixed Models*, 2nd edition. Wiley, New York.
- Normand, S. L. T., Glickman, M. E. and Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association* **92**, 803-814.

- Office of Quality and Performance. (2010). Fiscal year 2011 quarter 1 clinical measures specification manual, Volume 2. Veterans Health Administration, Washington, District of Columbia.
- Petersen, L. A., Pietz, K., Woodard, L. D. and Byrne, M. (2006). Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Medical Care* **43**, 61-67.
- Pietz, K., Ashton, C. M., McDonell, M. and Wray N. P. (2004). Predicting healthcare costs in a population of veterans affairs beneficiaries using diagnosis-based risk adjustment and self-reported health status. *Medical Care* **42**, 1027-1035.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition. Thousand Oaks, Sage Publications, California.
- SAS Institute. (2009). Version 9.1.3. Cary, North Carolina.
- Verisk Health. (2011). <http://www.veriskhealth.com/> (Last accessed 4/11/11).
- Zhang, H., Lu, N., Feng, C., Thurston, S., Xia, Y., Zhu, L. and Tu, X. (2011). On fitting generalized linear mixed effects models for binary responses using different statistical packages. *Statistics in Medicine* **30**, 2562-2572.
- Zucker, D. M. (1990). An analysis of variance pitfall: the fixed effects analysis in a nested design. *Educational and Psychological Measurement* **50**, 731-738.

Received December 29, 2012; accepted June 6, 2013.

Kenneth Pietz  
Health Services Research and Development (152)  
Michael E. DeBakey VA Medical Center and Baylor College of Medicine  
2002 Holcombe Boulevard, Houston, TX 77030, USA  
kpietz@bcm.edu

Laura A. Petersen  
Health Services Research and Development (152)  
Michael E. DeBakey VA Medical Center and Baylor College of Medicine  
2002 Holcombe Boulevard, Houston, TX 77030, USA  
laurap@bcm.edu

LeChauncy D. Woodard  
Health Services Research and Development (152)  
Michael E. DeBakey VA Medical Center and Baylor College of Medicine  
2002 Holcombe Boulevard, Houston, TX 77030, USA  
lwoodard@bcm.edu.