Analysis of Hospital Mortality Data: The Role of DRG's

Mohamed M. Shoukri[1]*, Sara N. Algahtani[2], Abdelmoneim M. Eldali[3], Manal R. AlMarzouqi[3], Saleh M. Al-Ageel[3]

[1]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada
[2]King Fahd Medical City, Riyadh, KSA
[3]Department of Biostatistics, Epidemiology, and Scientific Computing, King Faisal Specialist Hospital and Research Center, Riyadh, KSA
Email: *Shoukri.mohamed@gmail.com, S_algahtani@live.com, adalee@kfshrc.edu.sa, malmarzouqi@kfshrc.edu.sa, sageel@kfshrc.edu.sa

Abstract

Background: Factors associated with hospital mortality are usually identified and their effects are quantified through statistical modeling. To guide the choice of the best statistical model, we first quantify the predictive ability of each model and then use the CIHI index to see if the hospital policy needs any change. *Objectives*: The main purpose of this study compared three statistical models in the evaluation of the association between hospital mortality and two risk factors, namely subject's age at admission and the length of stay, adjusting for the effect of Diagnostic Related Groups (DRG). *Methods*: We use several SAS procedures to quantify the effect of DRG on the variability in hospital mortality. These procedures are the Logistic Regression model (ignoring the DRG effect), the Generalized Estimating Equation (GEE) that takes into account the within DRG clustering effect (but the within cluster correlation is treated as nuisance parameter), and the Generalized Linear Mixed Model (GLIMMIX). We showed that the GLIMMIX is superior to other models as it properly accounts for the clustering effect of "Diagnostic Related Groups" denoted by DRG. *Results*: The GLM procedure showed that the proportional contribution of DRG is 16%. All three models showed significant and increasing trend in mortality (P < 0.0001) with respect to the two risk factors (age at admission, and hospital length of stay). It was also clear that the CIHI index was not different under the three models. We re-estimated the models parameters after dichotomizing the risk factors at the optimal cut-off points, using the ROC curve. The parameters estimates and their significance did not change.

1. Introduction

The ability to gauge hospital performance using patient outcome data depends upon many factors. In principle, the outcome needs reflect features that are directly affected by the quality of hospital care, to name but a few; mortality, readmission rates patients and employee satisfaction. Beyond this, however, there are a number of important data and statistical considerations:

1) Data must be available and used to adjust for differences in patient health at admission across different hospitals (case-mix differences). These adjustments are required to ensure that variations in reported performance apply to hospitals' contributions to their patients' outcomes rather than to the intrinsic difficulty of the patients they treat. Needless to say that, performance of the adjustments depends on the type and quality of available data.

2) In distinct contrast to the previous point, reported performance should not adjust away differences related to the quality of the hospital. For example, if "presence of a special dialysis care unit" is systematically associated with better survival following organ failure, a hospital's reported performance should capture the benefit provided by that unit and as a consequence such hospital-level characteristic should not influence the risk adjustment.

3) The reported performance measure should be little affected by the variability associated with rates based on the small numbers of cases.

In this report we address technical statistical issues associated with the KFSHRC hospital mortality data. The salient point is that there is no consensus to guide our choice of an appropriate statistical model. However, we shall use the most scientific statistical models to analyze our data. To enhance the traditional modeling techniques we include use of more flexible models incorporating Diagnosis Related Groups (DRG) adjustment [1]; stressing the use of statistical distributions that do not belong to the well-known Gaussian family used in the hierarchical, random effects models; evaluation of the effectiveness of current outlier detection methods; and consideration of producing an ensemble of hospital-specific Standardized Mortality Ratios (HSMR) that accurately estimates the true, underlying distribution of ratios [2]. Discussion with clinicians and other quality experts concluded that risk adjustments should not reflect hospital characteristics, but their use in reducing confounding of the case-mix/ risk relation. Statistical models are available for each of these operations. The ability to develop and implement such models is now available since the adoption of the SAS software and the acquisition of its important components.

In Section 2 we define what is meant by DGR, and in Section 3 we describe the

data that were made available to us, with mortality as the primary outcome at the King Faisal Specialist Hospital (KFSHRC). In Section 4 we compare the models, and in Section 5 we discuss the quantitative merits of these models, followed by recommendations.

2. The Importance of Incorporating DRG's within the Proposed Models

The Diagnostic Related Groups (DRGs) were first developed at Yale University in 1975. The main objective was to group patients with similar treatments and conditions for comparative studies. DRGs were designed to be homogeneous units of hospital activity to which binding prices could be attached. A central theme in the advocacy of DRGs was that this reimbursement system would, by constraining the hospitals, oblige their administrators to alter the behavior of the physicians and surgeons comprising their medical staffs. Hospitals were forced to leave the nearly risk-free world of cost reimbursement, and face the uncertain financial consequences associated with the provision of health care. DRGs were designed to provide practice pattern information that administrators could use to influence individual physician behavior.

DRGs were designed to be homogeneous units of hospital activity to which binding prices could be attached. A central theme in the advocacy of DRGs was that this reimbursement system would, by constraining the hospitals, oblige their administrators to alter the behavior of the physicians and surgeons comprising their medical staffs. Hospitals were forced to leave the nearly risk-free world of cost reimbursement, and face the uncertain financial consequences associated with the provision of health care. DRGs were designed to provide practice pattern information that administrators could use to influence individual physician behavior.

In 2007, author Rick Mayes [3] described DRGs as:

...the single most influential postwar innovation in medical financing: Medicare's prospective payment system (PPS). Inexorably rising medical inflation and deep economic deterioration forced policymakers in the late 1970s to pursue radical reform of Medicare to keep the program from insolvency.

In the USA the most significant change in health policy since Medicare and Medicaid's passage in 1965 went virtually unnoticed by the general public [4]. Nevertheless, the change was nothing short of revolutionary. For the first time, the federal government gained the upper hand in its financial relationship with the hospital industry. Medicare's new prospective payment system with DRGs triggered a shift in the balance of political and economic power between the providers of medical care (hospitals and physicians) and those who paid for it—power that providers had successfully accumulated for more than half a century. From statistical view point DRG's are considered artificial clusters of subjects.

64

Krumholz *et al.* [5] discussed several factors that should be considered when assessing hospital quality. These relate to differences in the chronic and clinical acuity of patients at hospital presentation, the numbers of patients treated at a hospital, the frequency of the outcome studied, the extent to which the outcome reflects a hospital quality signal, and the form of the performance metric used to assess hospital quality. However, issues related to DRG have not been considered as factors of importance. Since the outcome of interest is hospital mortality, any attempt to derive risk adjusted mortality that does not take into account the relative importance of DRG will produce biased estimates [6]. The performance measure is reported as:

$$\text{Observed \# deaths/Expected (model based) \# of deaths} \tag{1}$$

The denominator of Equation (1) results from applying a model that adjusts/standardizes for an ensemble of patient-level, pre-admission risk factors, rather than only demographic factors such as age and gender as is typical in epidemiological applications. The statistical issues arising in the estimation of the standardized death rate and the SMR are identical because the latter is simply the hospital-specific value divided by the expected number of deaths computed from the postulated risk model.

3. Methods

3.1. Study Design

Hospital discharge status, available from the hospital medical records from 2014 through 2016 were extracted. For each subject, the age at admission, length of stay and DRG membership were included in this cross sectional retrospective design. The study was reviewed and approved by the Institutional Review Board at the King Faisal Specialist Hospital and Research Center (KFSHRC).

3.2. Study Variables

3.2.1. Dependent Variable

Discharge status is the dependent variable (dead/alive). Because of the Bernoulli distribution of the outcome, the log-odds of death were calculated in the analytical cohort.

3.2.2. Independent Variables

Regression models included parameters that defined age at admission, length of stay, gender, and DRG. Because DRG is a categorical variable with excessive number of levels our modeling strategy used DRG as a clustering variable, and as a random effect variable. The fundamental aim was to adjust the standard errors of the estimated model parameters for the possible within DRG correlation. Another reason is, to preserve the stochastic process and hierarchical structure of the data, and develop an effective risk adjustment. Because patient-specific outcomes are binary (death indicator), a Bernoulli model operating at the patient level is appropriate. Risk adjustment and stabilization should adopt this

model and thus logistic regression is a suitable approach for including the effects of patient-level characteristics. With flexible modeling of covariate influences, the model would produce a valid risk adjustment and there is no reason to replace the logistic by another function.

The evaluation process must be based on an effective risk adjustment. Though one might wish to have additional information of patient attributes and clinical severity, even with currently available data we should evaluate whether a more flexible risk adjustment model will improve performance. Patient characteristics (clinical and demographic) are of the three types, measured and accounted for, measurable but not accounted for, and characteristics that are difficult or impossible to measure. Prudence dictates that risk adjustments should include pre-admission medical conditions, but whether or not to include demographic attributes is a policy decision.

3.3. Statistical Analysis

Univariate and descriptive statistics were used to profile the study covariates, including the frequency distribution of the top twelve DRG's, as shown in Table 1. Because of the binary nature of the outcome of interest (patient's status when discharged), we fitted logistic regression models to estimate change in level (intercept) and trend (slope) on log-odds of age at admission and length of stay. Each model was adjusted to account for the clustering effect of DRG.

Three statistical estimation procedures in SAS (GLM, GEE, GLIMMIX) were used to account for the correlation between responses with a DRG and heterogeneity across individuals in the study. The intra-class correlation was calculated using the one-way ANOVA using the GLM procedure in SAS. Data management and analyses were accomplished via PC-SAS (v9.4) [7], with an a-priori Type I error rate set at 0.01.

Table 1. The ICD9 diagnoses for the top most frequent DRG's in the hospital data base.

DRG	frequency	ICD9 Diagnosis
1-P67D	4270	Neonate-admission weight > 2499 g W/O significant OR procedure.
2-O60Z	2097	Vaginal delivery complications
3-F42B	1434	Diseases of the circulatory system
4-R60B	1308	Acute Leukemia W catastrophic CC
5-R61B	1273	Lymphoma & non-acute Leukemia W/O catastrophic CC
6-O01B	1187	Pregnancy complications
7-L04C	1173	Diseases of Kidney and Urinary tract
8-K64B	1127	Endocrine and metabolic diseases
9-Q60A	1064	Immunity disorders
10-E62B	1016	Diseases of the respiratory system
11-O66Z	1004	Antenatal and other obstetric admissions
12-B02C	1002	Diseases of the nervous system

The analyses produced point estimates and 95% confidence intervals of the odds ratios whether the two covariates were entered the models as continuous or as categorical variables.

4. Results

The study included 191943 discharges, of which 184,907 alive and 7046 dead. The summary statistics are outlined in Table 1 and Table 2.

Assuming that the number of DRG's in the data base is k, and the size of the ith DRG is k_i. The estimated intra-cluster correlation obtained from the one-way ANOVA using the GLM procedure in SAS (Shoukri [8]) is given in Equation (2):

$$\hat{\rho} = \frac{\text{MSBD} - \text{MSWD}}{\text{MSBD} + (k_0 - 1)\text{MSWD}} \tag{2}$$

where MSBD and MSWD are respectively the between DRG mean squares and the within DRG mean squares. Moreover:

$$k_0 = \frac{1}{n-1}\left[\sum k_i - \frac{\sum k_i^2}{\sum k_i}\right], \; i = 1, 2, \cdots, k \text{ and } n = \sum k_i$$

Summary statistics for age at admission and length of stay are presented in Table 2. Note that the standard deviation formula uses the (number of observations minus one) to produce an unbiased estimator for the corresponding population parameter (Shoukri [8]).

The main purpose of using the GLM, which requires independent responses, is to produce a point estimator of the within cluster correlation (Shoukri, [8]).

From the GLM procedure we have, $\text{MSBD} = 1.8214$, $\text{MSWD} = 0.0218$ and $k_0 = 474$. From Equation (1) the intra-cluster correlation coefficient $\hat{\rho} = 0.16$.

Note that:

The Effect of Dichotomization

Measurements of continuous variables are made in all branches of epidemiological studies aiding in the diagnosis and treatment of patients. In clinical practice it is helpful to label individuals as having r not having an attribute, such as

Table 2. Summary statistics for age at admission (AAA) and length of stay (LOS) presented by discharge status (Alive, Dead). (a) Status = Alive; (b) Status = Dead.

(a)

	N	Minimum	Maximum	Mean	Std. Deviation
AAA	184907	0	100	32.39	24.841
LOS	184894	1	358	9.10	15.352

(b)

	N	Minimum	Maximum	Mean	Std. Deviation
AAA	7046	0	100	46.31	27.196
LOS	6990	1	362	28.74	41.572

being "old" or "young" or having "long stay" depending on the number of days.

Dichotomization of continuous variables is also common in clinical research, but the statistical analysis has some serious drawbacks as there will be reduction in the precision of the estimated effect sizes. Though grouping may help data presentation, notably in tables, categorization is unnecessary for. Here we consider the impact of converting continuous data to two groups (dichotomizing), as this is the most common approach in clinical research.

Within each model we estimated for each effect the log-odds as an effect size using continuous and categorized covariates. We found that the GLIMMIX has superior advantage over the logistic regression and the GEE models. We calculated the optimal split for the AAA and LOS using the "Receiver Operating Characteristic curve" or ROC curve. Figure 1 and Table 3 shows the optimal cut off for LOS is 160 days, and the corresponding area under curve 73%. This means that the risk of death is significantly higher among patients who are hospitalized over 160 days relative to those who stay less than 160 days, (corrected P-value = 0.0001). Additionally, in Figure 2, and Table 4 we show that the optimal split for AAA is 53 years. The areas under the ROC curve corresponding to the dichotomized covariate AAA is 65%.
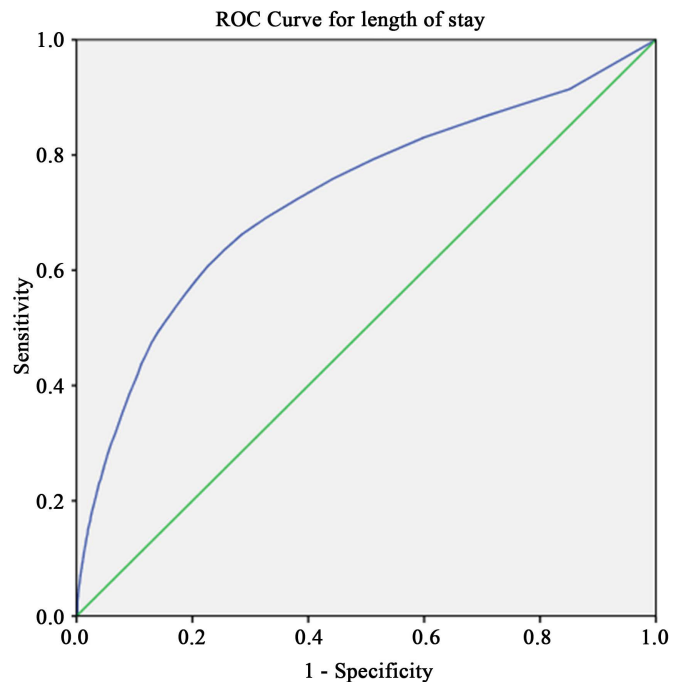

Figure 1. The ROC curve for the LOS.

Table 3. Area under the curve for the LOS optimal cut-off point.

Area	Std. Error	P-value	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.727	0.004	0.0001	0.720	0.734

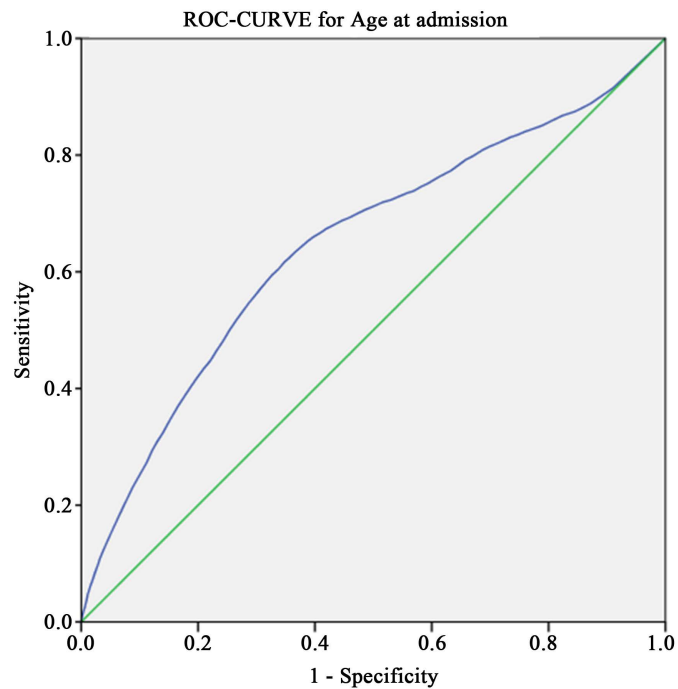Corrected P-value = 0.0001; The optimal cut-off point is LOS ≥ 160 days.

Figure 2. The ROC curve for AAA.

Table 4. Area under the curve, test result variable(s) for AAA.

Area	Std. Error	P-Value	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.647	0.004	0.0001	0.640	0.654

Corrected p-value = 0.0001; The optimal cut-off pint is age ≥ 53 is associated with death.

Dichotomizing leads to several problems. Firstly, much information is lost, and this can been seen from the increase in the estimated standard errors of the odds ratios. Moreover, the odds ratios point estimates are inflated as well and therefore are potentially biased. For example, in **Table 5** the estimated odds ratio of the dichotomized LOS is 14.53 while its value is 1.024 when measured on the continuous scale under the same model. The remark holds true if the fitted model is the GEE as shown in **Table 6**. The estimates are somewhat stable under the GLIMMIX and the results are shown in **Table 7**. One may conclude that dichotomization may increase the risk of a positive result being a false-positive.

Categorical Age:

Alternative to dichotomization we categorized age in a meaningful way such that:

Group 1: Age is less than 14 years

Group 2: Age between 15 and 30

Group 3: Age between 31 and 59

Group 4: Age above 60.

When we plotted the mortality rate, with 95% confidence limits, against the 4 age categories, as shown in **Figure 3**, there was an increasing trend in mortality
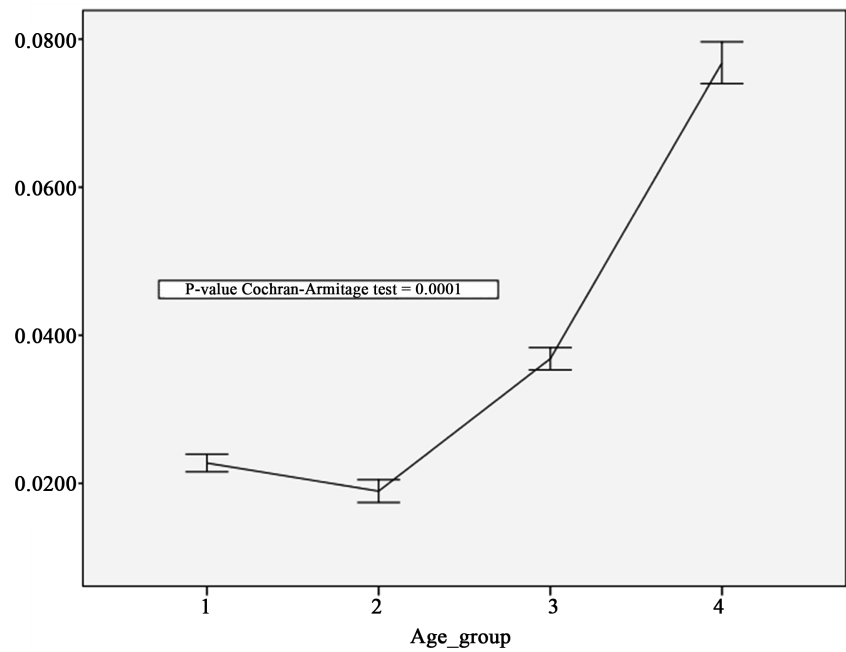
Figure 3. Death rate versus the 4 age categories.

Table 5. Estimating the odds ratios using the logistic regression models.

Effect	Continuous Covariates		Dichotomized Covariates	
	OR	95% CI	OR	95% CI
AAA	1.021	1.02 - 1.022	2.88	2.75 - 3.03
LOS	1.024	1.023 - 1.025	14.53	11.86 - 17.80
AIC = 54522			AIC = 57694	

Table 6. Estimating the odds ratios using the GEE models.

Effect	Continuous Covariates		Dichotomized Covariates	
	OR	95% CI	OR	95% CI
AAA	1.025	1.02 - 1.031	3.22	2.304 - 4.511
LOS	1.030	1.024 - 1.036	19.88	10.665 - 37.308
QIC = 13192			QIC = 14319	

Table 7. Estimating the odds ratios using the GLIMMIX models.

Effect	Continuous Covariates		Dichotomized Covariates	
	OR	95% CI	OR	95% CI
AAA	1.014	1.012 - 1.017	1.68	1.48 - 1.90
LOS	1.013	1.011 - 1.015	2.42	1.48 - 3.97
$\sigma_b^2 = 3.24 \pm 0.50$			$\sigma_b^2 = 3.56 \pm 0.54$	
Scaled deviance = 0.62			Scaled deviance = 0.65	
−2 Res log-likelihood = 477379			−2 Res log pseudo-likelihood = 478461	

as age groups moved up. The one-degree of freedom Cochran-Armitage test for trend was quite significant with p-value < 0.001.

$\sigma_b^2 = 3.56 \pm 0.54$, Scaled deviance = 0.61, and, −2 Res log pseudo-likelihood = 477030.

The GLIMMIX estimation when age is categorized into 4 groups is given in Table 8. The odds ratio estimate of LOS, which is highly correlated with age has improved and in fact is almost similar to the estimated odds ratio when the age was taken as a continuous covariate. There is almost no change in the between DRG variance component estimate, $\sigma_b^2 = 3.56$, confirming the hypothesis that the measured covariates are not correlated with the random component in the model. The scaled deviance is less than one, indicating that the model has captured the effect of the measured and the unmeasured covariates. The value of −2 Res log pseudo-likelihood, (which is equivalently defined as the AIC),indicates that model goodness of fit is also acceptable.

Under the GLIMMIX model, whose results are summarized in Table 8, the-CIHI [9] index of hospital performance when mortality is the outcome of interest is:

$$NUM = \sum O_i = 7053, \text{ and } DEN = \sum E_i = 7117$$

giving CIHI = *NUM/DEN* less than unity. This indicates that the hospital risk adjusted mortality meets the CIHI criteria for quality.

5. Discussion

Although the fitted models produced odds ratio estimates whose changes and trend were in the same direction and of the same significance, the magnitude of point estimates and length of confidence intervals varied. Clearly the logistic regression produced a smaller length of confidence intervals. This should be expected, since this model ignores the nature of the correlation structure among responses within each DRG, and hence the standard errors are under-estimated. The GEE, introduced by Liang and Zeger [10] is suitable for the analysis of clustered data. The GEE estimation produced similar magnitude of point estimates but relatively less precise confidence intervals. The GEE is supposed to produce consistent estimates even if the within DRG correlation parameter is misspecified. In our case we assigned an exchangeable correlation, as a working correlation parameter to represent the average within DRG heterogeneity. Finally the GLIMMIX produced entirely different set of estimates, and an estimate of an

Table 8. GLIMMIX: Age is categorized into 4 groups with group 4 being the reference.

Effect	OR	95% CI
Age group 1	0.901	0.816 - 0.996
Age group 2	1.805	1.684 - 1.935
Age group 3	3.589	3.353 - 3.842
LOS	1.024	1.023 - 1.025

additional parameter representing the variance component or quantification of the between DRG variation.

When applied to dichotomous response variable, the GEE results are to be interpreted at the population average (PA) level and does not account for heterogeneity across the DGRs. The GEE indicates that the risk of death depends on AAA and LOS that are measured at the individual level and not on any random effect across the clusters of DRGs. This model uses a working correlation as an instrument to account for the within DRGs variations.

However the, the fundamental feature of the GLIMMIX is the assumption of heterogeneity across DRGs in our study population. The GLIMMIX interprets the estimated model parameter (the odds ratio) as conditioned on DRG specific intercepts. These intercepts reflect a natural heterogeneity due to unmeasured covariates. For example, presence or absence of co-morbid conditions, or re-admission (yes/no) may produce different trajectory for the risk of death. Therefore, there is as expected a sizable difference in the magnitude of uncertainties between the GEE and the GLIMMIX.

6. Conclusion

In conclusion, health services and quality of care research might use any of the above models depending on the scientific question posed. The GEE, will produce estimates that will be of most interest to quality health services and policy makers who evaluate hospital performance on an average level. On the other hand, the GLIMMIX will have RDRG level interpretation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Blackford, G. (2008) Discover How to Survive Medicare Severity DRGs with a Successful Clinical Documentation Improvement Program and Present on Admission Indicators. 2008 *AHIMA Convention Proceedings*, 11-16 October 2008, Seattle, WA.

[2] Glance, L.G., Dick, A.W., Mukamel, D.B., *et al.* (2010) How Well Do Hospital Mortality Rates Reported in the New York State. CABG Reportcards Predict Subsequent Hospital Performance? *Medical Care*, **48**, 466-471. https://doi.org/10.1097/MLR.0b013e3181d568f7

[3] Mayes, R. (2007) The Origins, Development, and Passage of Medicare's Revolutionary Prospective Payment System. *Journal of the History of Medicine and Allied Sciences*, **62**, 21-55. https://doi.org/10.1093/jhmas/jrj038

[4] Donabedian, A. (1988) The Quality of Care. How Can It Be Assessed? *JAMA*, **260**, 1743-1748. https://doi.org/10.1001/jama.1988.03410120089033

[5] Krumholz, H.M. (2008) Outcomes Research: Generating Evidence for Best Practice. *Circulation*, **118**, 309-318. https://doi.org/10.1161/CIRCULATIONAHA.107.690917

[6]   Martin, G. (2008) MS-DRG Journey: How One Hospital Joined Together to Successfully Implement MS-DRGs. *Journal of AHIMA*, **79**, 70-72.

[7]   SAS Institute (2010) STAT-SAS, Version 9.4. SAS Institute, Cary, North Carolina.

[8]   Shoukri, M. (2010) Measures of Inter-Observer Agreement and Reliability. Chapman and Hall-CRC Press, Boca Raton. https://doi.org/10.1201/b10433

[9]   Canadian Institute for Health Information (2009) The CIHI Data Quality Framework.

[10]   Quality Framework (2009). http://www.cihi.ca/CIHI-ext-portal/pdf/internet/DATA

[11]   Liang, K.Y. and Zeger, S.L. (1993) Regression Analysis for Correlated Data. *Annual Review of Public Health*, **14**, 43-68.
        https://doi.org/10.1146/annurev.pu.14.050193.000355