Comparison of logistic and multilevel regression in the risk adjustment

Paolo Trerotoli, Nicola Bartolomeo, Gabriella Serio

Department of Biomedical Sciences and Human Oncology, University of Bari, Italy

Corresponding Author:
Paolo Trerotoli
Dipartimento di Scienze Biomediche ed Oncologia Umana
Piazza G. Cesare 1, 70124, Bari, Italy
E-mail: trerotoli@igiene.uniba.it

Summary

Objectives. To compare risk-adjustment methods commonly used to determine the outcome indicators of either a logistic or multilevel regression, given the following hierarchical structure of hospital data: patients (first level), specialist ward (second level), hospital (third level). The effect of the subsequent risk-adjustment was evaluated by two mortality indicator indices, as proposed by the Agency for Healthcare Research and Quality (AHRQ), Inpatient Quality Indicators number 16, (congestive heart failure, IQI16) and Inpatient Quality Indicators number 17, (stroke, IQI17), using as reference population the hospital admissions of the Apulia Region, and applying the coefficients supplied by AHRQ.

Methods. Two regression models, multilevel and logistic, were applied on hospital admissions for congestive heart failure and stroke in the Apulia during the period 2001-2005. The relevant regression coefficients have been compared using the Wald test. The level effects were evaluated by means of an intraclass correlation coefficient (ICC). The risk-adjusted mortality rates utilised for the comparison were determined on admissions in 2006.

Results. For IQI16, the Wald test gave contrasting results for the variance-covariance matrix used. The multilevel model explained a high variability among second and third level units, with ICCs of 33.31% and 12.88%, respectively. For IQI17, the Wald test showed a significant difference between the regression coefficients of the two models. The ICCs of the second and third level units were 47.41% and 6.15%, respectively. The risk-adjusted rates were different: the AHRQ score was 4.36% and the Apulian score 3.15% for IQI16; the AHRQ score being 11.2% and the Apulian score for IQI17 being 7.89%.

Conclusions. It is essential to correctly define both the variables in the adjustment model and in the reference population so that the latter is matched to the population under study. We do this in order to obtain outcome indicators that truly reflect the sought-after situation in the larger population, and to avoid model-level bias.

KEY WORDS: *risk adjustment, multilevel model, hospital mortality rate*.

Introduction

The frequent changes in the Italian National Health Service (INHS) and the decline in nationally-available resources have made it necessary to accomplish effective monitoring of Italian healthcare providers, in order to control the consequences visited upon the public welfare.

Two important projects have been performed in Italy, aimed at evaluating the performance of healthcare providers and to study the outcomes indicators (1, 2). Much research, now well-consolidated and -understood, has also been previously conducted in this field on an international setting (3, 4).

A model which is easy to apply to the INHS is the one developed in the USA by the Agency for Healthcare Research and Quality (AHRQ) that is used to determine Inpatient Quality Indicators (IQI). These first level indicators can be calculated using hospital patient discharge databases alone, and allow assessment of the:

- volume of activity;
- mortality rate for each procedure;
- mortality rate per each principal diagnosis;
- rate of use of specific procedures.

Risk adjustment procedures are applied to evaluate how an indicator such as the mortality rate can be biased by casemix on the part of healthcare providers.

These methods should be used, particularly in retrospective studies, to adjust rates for the different distribution of the individual risk factors when assuming that patients have the same level of risk. The differences observed between the unadjusted rate and the adjusted rate would thus be attributable only to the different quality of care, rather than to the effect of residual variability.

Risk adjustment procedures are based on a multivariate regression model whose covariates (age, sex, and severity of the patient's condition) can be considered risk factors for occurrence of the event (e.g. death). Their relative coefficients are estimates of a changed rate of occurrence of the event depending on variations of the risk factors.

In the first version of the freeware from the AHRQ (V.2.1), a linear regression model was used, its parameters were estimated by the ordinary least squares method and the continuous dependent variable was assumed to be normally distributed. However, the indicators under analysis are very often dichotomous, discrete variables (e.g., alive/dead), with an underlying binomial distribution, so it was considered preferable to use a logistic regression model. (AHRQ freeware Version 3.0 (2006)). Nevertheless, the logistic regression model does not accommodate well to the hierarchical nature of the data in health management databases (e.g., hospital, specialist unit); therefore a multilevel model was later concluded to be more structurally suitable.

The application of a multilevel regression model is, in fact, advisable when it is reasonable to assume that the variability of the phenomenon depends not only on individual explanatory variables (first level), but depends also on a given individual belonging to a particular group (second level unit) endowed with special, distinguishing characteristics. The latest version of the AHRQ software (V.3.1) relies on a hierarchical model to determine the coefficients to be used in risk adjustment, using the hospital as the random effect (http://www.qualityindicators.ahrq.gov/newsletter/2007-February-AHRQ-QI-Newsletter.htm#_Workgroup_on_30-Day).

The object of our study was to verify the appropriateness of the multilevel logistic model as compared to the simpler logistic model. In our hospital patient discharge database structure, a patient (first level) is automatically assigned to the specialist ward discharging the patient (second level), which in turn is assigned to the admitting hospital (third level). To evaluate the effect of the reference population on the risk adjustment procedure, a score vector was also determined using hospitals admissions data from the Puglia Region. Adjusted mortality rates obtained with AHRQ scores were compared with those found using the scores obtained for the Puglia reference population. Among the indicators proposed by the AHRQ, we additionally adopted the following two: IQI16 (mortality rate for congestive heart failure) and IQI17 (mortality rate for acute stroke).

Method

Data utilised

The risk adjusted indicators were determined for hospital discharges (inpatients) in the year 2006, in Puglia, a Region in South Italy. Patients were selected on the basis of having either the discharge diagnosis of congestive heart failure for IQI16 or stroke for IQI17, this nomenclature being taken from the corresponding ICD9-CM codes, as indicated in the AHRQ manual (Table 1).

To obtain the parameters scores vector used in the risk adjustment procedure, the regression model was applied to the Puglia hospital discharge (PHD) database for the period 2001-2005, defined as the reference population.

Statistical analysis

The risk adjustment procedure consists of the following phases:

- fitting of the regression model to the reference population;
- determination of the regression coefficients, defined as scores, for each covariate;
- application of the previously determined scores to observations of the population under study (PHD 2006);
- calculating the risk-adjusted hospital rates in the population under analysis.

Table 1. Diagnosis Codes ICD9-CM for the selection of the indicators IQI16 and IQI17.

Congestive heart failure (CHF) mortality rate (IQI 16)	Acute stroke mortality rate (IQI 17)
Numerator: Number of deaths among cases meeting the inclusion and exclusion rules for the denominator.	Numerator: Number of deaths among cases meeting the inclusion and exclusion rules for the denominator.
Denominator: All discharges, aged 18 years and older, with a principal diagnosis code of CHF. Icd-9-Cm Chf Diagnosis Codes: 39891 Rheumatic Heart Failure 40201 Mal Hypert Hrt Dis W Chf 40211 Benign Hyp Hrt Dis W Chf 40291 Hyperten Heart Dis W Chf 40401 Mal Hyper Hrt/Ren W Chf 40403 Mal Hyp Hrt/Ren W Chf&Rf 40411 Ben Hyper Hrt/Ren W Chf 40413 Ben Hyp Hrt/Ren W Chf&Rf 40491 Hyper Hrt/Ren Nos W Chf 40493 Hyp Ht/Ren Nos W Chf&Rf 4280 Congestive Heart Failure 4281 Left Heart Failure 42820 Systolic Heart Failure Nos Oct02 42821 Ac Systolic Hrt Failure Oct02- 42822 Chr Systolic Hrt Failure Oct02- 42823 Ac On Chr Syst Hrt Fail Oct02- 4289 Heart Failure Nos 42830 Diastolic Hrt Failure Nos Oct02- 42831 Ac Diastolic Hrt Failure Oct02- 42832 Chr Diastolic Hrt Fail Oct02- 42833 Ac On Chr Diast Hrt Fail Oct02- 42840 Syst/Diast Hrt Fail Nos Oct02- 42841 Ac Syst/Diastol Hrt Fail Oct02- 42842 Chr Syst/Diastl Hrt Fail Oct02- 42843 Ac/Chr Syst/Dia Hrt Fail Oct02- Excluded cases: • missing discharge data • transfer to another short-term hospital • MDC 14 (pregnancy, childbirth, and puerperium) • MDC 15 (newborns and other neonates)	Denominator: All discharges, aged 18 years and older, with a principal diagnosis code for stroke. Icd-9-Cm Stroke Diagnosis Codes: 430 Subarachnoid Haemorrhage 431 Intracerebral Haemorrhage 4320 Nontraum Extradural Hem 4321 Subdural Haemorrhage 4329 Intracranial Haemorr Nos 43301 Basi Art Occl W/ Infarct 43311 Carotd Occl W/ Infrct 43321 Vertb Art Occl W/ Infrct 43331 Mult Precer Occl W/ Infrct 43381 Precer Occl Nec W/ Infrct 43391 Precer Occl Nos W/ Infrct 43401 Cere Thrombosis W/ Infrct 43411 Cere Embolism W/ Infrct 43491 Cereb Occl Nos W/ Infrct 436 Cva* * Only for discharges before September 30, 2004. Does not apply to discharges on or after October 1, 2004.

The first model used in the process of defining the scores was the multiple logistic regression models:

$$logit\left(\hat{p}_i\right) = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} \qquad [1]$$

where *i* indicates the i^{th} individual and *k* indicates the generic covariate.

The variables X_k, as utilised in our regression model, are the same as those of the AHRQ: sex, age subdivided into 14 groups, and sex by age interaction. These variables are apart from those that more heavily weigh and monitor the severity of the patient's condition.

The AHRQ software utilises the APR-DRG system in adjusting the rates according to the severity of the patient, however these severity indices were not available in the PHD database.

The value of expected probability is given by:

$$\hat{p}_i = \frac{\exp\left(\alpha + \sum_{k=1}^{K} \beta_K X_{ik}\right)}{1 + \exp\left(\alpha + \sum_{k=1}^{K} \beta_K X_{ik}\right)} \qquad 0 \le \hat{p}_i \le 1 \qquad [2]$$

And the "adjusted" rate is:

*Adjusted Rate = (Observed Rate / Expected Rate) * Average Population Rate* [3]

We compared the results obtained with simple logistic regression (LR) with those of a multilevel model (ML), assuming a nested data structure of three variables and levels. The patients are the first level unit (*i*), the specialist ward that the patient was discharged from is the second level unit (*j*), and the hospital is the third level unit (*h*).
The multilevel regression model used is then:

$$logit\left(\hat{p}_{ijh}\right) = \alpha + \sum_{k=1}^{K} \beta_k X_{kijh} + v_h + u_{jh} + r_{ijh} \quad [4]$$

The residual r_{ijh} is specific to patient i in the specialist ward *j* of the hospital *h*. The residual u_{jh} is specific to the second level unit *j* in the third level unit *h* and represents the difference between the general intercept and the specific intercept of the speciality *j* in the hospital *h*. Finally, v_h is the specific residual of the third level unit *h*, which represents the difference between the general intercept and the specific intercept of the hospital. Once the parameters α and β have been estimated, it is possible to determine the probability of the expected rate and to calculate the adjusted rate by [3]. The score vector obtained with the reference population was compared to that found with the AHRQ. Both vectors were applied to the population under analysis (PHD 2006), with the aim of evaluating the influence that the reference population has on the risk-adjusted rates.

Method for results comparison

The estimated parameters for each regression model (LR and ML) follow an approximately normal distribution. To ascertain if there were significant statistical differences between the two models, the Wald test was applied:

$$W_{Log} = (\hat{\beta}_{Log} - \hat{\beta}_{Mul})^T V^{-1}_{\beta\,Log} (\hat{\beta}_{Log} - \hat{\beta}_{Mul})$$
$$W_{Mult} = (\hat{\beta}_{Log} - \hat{\beta}_{Mul})^T V^{-1}_{\beta\,Mul} (\hat{\beta}_{Log} - \hat{\beta}_{Mul}) \quad [5]$$

in which $\hat{\beta}_{Log}$ and $\hat{\beta}_{Mul}$ represent the estimated parameters of the LR and ML models, respectively; and $V^{-1}_{\beta\,Log}$ and $V^{-1}_{\beta\,Mul}$ represent the inverse of the variance-covariance matrix of the LR and ML models' regression parameters, respectively. Under the null hypothesis that no statistically significant differences exist between the two models, the Wald test statistic is distributed according to a Chi-square with ν degrees of freedom, where ν represents the number of explanatory variables used in the statistical model. The intraclass correlation coefficient (ICC) was used to estimate the mortality variability between the groups as identified by the hierarchical model. This coefficient represents the proportion of variability explained by the presence of clusters in the observed populations (5, 6). The ICC approximation proposed by Snijders and Bosker was used; this, in the case of three levels, allows two ICCs to be calculated:

$$ICC_{2°level} = \tau_2 / (\tau_2 + \tau_3 + \pi^2/3)$$
$$ICC_{3°level} = \tau_3 / (\tau_2 + \tau_3 + \pi^2/3)$$

where τ_2 and τ_3 are the estimated variances of the random effect of the second and third level units on the average, and π is the number 3.142.
The comparison between the LR and the ML models was also carried out by calculating the difference between the hospital mortality rates (expected and adjusted), these rates obtained with the two models that are using the risk adjustment procedure. The mean and the variance of the difference between the mortality rates was used to determine whether future use of the ML model could explain higher proportions of variability among hospital death rates, perhaps offering more precise estimates than what can be seen with use of the LR model.

Results

There were 13,626 records with a discharge diagnosis of congestive heart failure (IQI16) and 4,959 with a discharge diagnosis of stroke (IQI17) selected from the PHD in the year 2006. There were 426 patients discharged as "deceased" (the outcome of interest) in the IQI16 category (3.13%), and 405 patients similarly discharged in the IQI17 category (8.17%). Tables 2a and 2b show the distribution, by age class and

Table 2a. IQI 16 (congestive heart failure mortality rate) - Mortality in the study population by age and gender (PHD 2006).

	Females			Males		
Age	Patients	Deaths	% deaths	Patients	Deaths	% deaths
18-24	4	0	0.00	5	0	0.00
25-29	5	0	0.00	15	0	0.00
30-34	9	0	0.00	16	0	0.00
35-39	9	0	0.00	29	1	3.45
40-44	29	0	0.00	60	0	0.00
45-49	41	1	2.44	120	4	3.33
50-54	97	1	1.03	151	1	0.66
55-59	189	1	0.53	345	4	1.16
60-64	294	3	1.02	434	7	1.61
65-69	635	10	1.57	717	19	2.65
70-74	1 015	18	1.77	1 029	20	1.94
75-79	1 510	36	2.38	1 323	46	3.48
80-84	1 676	64	3.82	1 268	39	3.08
85+	1 646	103	6.26	955	48	5.03
Total	7 159	237	3.31	6 467	189	2.92

Table 2b. IQI 17 (acute stroke mortality rate) - Mortality of the study population by age and gender (PHD 2006).

	Females			Males		
Age	Patients	Deaths	% deaths	Patients	Deaths	% deaths
18-24	7	0	0.00	3	0	0.00
25-29	8	1	12.50	15	1	6.67
30-34	6	0	0.00	9	1	11.11
35-39	15	1	6.67	27	2	7.41
40-44	20	1	5.00	38	3	7.89
45-49	56	4	7.14	61	3	4.92
50-54	54	2	3.70	93	5	5.38
55-59	82	0	0.00	134	7	5.22
60-64	119	9	7.56	209	15	7.18
65-69	185	13	7.03	340	16	4.71
70-74	318	29	9.12	397	24	6.05
75-79	515	42	8.16	479	49	10.23
80-84	501	39	7.78	411	37	9.00
85+	570	60	10.53	287	41	14.29
Total	2 456	201	8.18	2 503	204	8.15

sex, in the PHD database of 2006 for IQI16 and IQI17, respectively.

The IQI16 indicator

A very low frequency of death due to congestive heart failure disease was found in our data sets in the lower age groups (Table 2a). In multilevel models, convergence problems can arise when the estimates of the random components of the intercept are close to zero (7). To overcome this, the covariates that appear only in classes with a low incidence, such as ages between 18-39 years and one's interaction with sex, were eliminated. The same was done in the logistic regression model (LR) to make the two models more comparable. Table 3 shows the estimated parameters of both models, the standard errors, their significance and the components of the variance associated with the second (the specialist wards) and third levels (the hospitals). In the ML model the variance components associated with the specialist ward, $\sigma^2_{Hosp*Spec} = 2.036$ (z=7.665; p<0.001), and the hospi-

Table 3a. Parameter Estimates using the Simple Logistic Model and Multilevel Logistic Model fitted to IQI-16 (congestive heart failure mortality rate).

Parameters		Simple Logistic Regression			Multilevel Logistic Regression		
		Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Reference: Males; Aged 60-64							
Intercept		-4.127	0.155	<0.001	-4.430	0.207	<0.001
sex	Females	-0.255	0.257	0.321	-0.202	0.265	0.445
age 5	40-44	-0.755	0.726	0.298	-0.893	0.741	0.227
age 6	45-49	-0.283	0.439	0.519	-0.187	0.448	0.676
age 7	50-54	-0.070	0.329	0.829	-0.177	0.343	0.604
age 8	55-59	0.125	0.246	0.610	0.106	0.256	0.676
age 9	60-64	0	-	-	0	-	-
age10	65-69	0.037	0.202	0.852	0.057	0.209	0.784
age11	70-74	0.550	0.176	0.001	0.586	0.183	0.001
age12	75-79	0.574	0.173	0.000	0.610	0.180	0.0007
age13	80-84	1.019	0.170	<0.001	0.995	0.178	<0.001
age14	85+	1.483	0.167	<0.001	1.485	0.175	<0.001
age19	40-44*Fem	0.918	1.038	0.376	0.964	1.058	0.362
age20	45-49*Fem	-0.220	0.859	0.797	-0.241	0.869	0.781
age21	50-54*Fem	0.550	0.503	0.274	0.758	0.518	0.143
age22	55-59*Fem	-0.559	0.496	0.260	-0.380	0.506	0.452
age23	60-64*Fem	0	-	-	0	-	-
age24	65-69*Fem	0.238	0.323	0.461	0.136	0.334	0.683
age25	70-74*Fem	-0.149	0.290	0.607	-0.224	0.299	0.454
age26	75-79*Fem	-0.007	0.281	0.979	-0.082	0.289	0.775
age27	80-84*Fem	-0.082	0.275	0.765	-0.100	0.284	0.723
age28	85+ *Fem	0.251	0.269	0.350	0.137	0.277	0.620
Variance components					Variance	St. error	p-value
σ^2_{Hosp}					0.787	0.245	0.001
$\sigma^2_{Hosp*Spec}$					2.036	0.265	<0.001

tal σ^2_{Hosp} = 0.787 (z=3.206; p=0.001) were both statistically significant. The Wald test statistics suggest a significant difference between the LR and the ML model (p-value < 0.001, first line Table 4) given the variance-covariance matrix being obtained from the LR model. Instead, if the Wald test is computed with the variance-covariance matrix of the ML model, there is no significant difference between the two approaches (p=0.986).

The estimated ICCs value obtained in the multilevel model are 33.31%, which is related to a variability explained by the difference between the specialist ward in the hospitals; and 12.88%, explained by the difference between the hospitals. Table 5 shows the average and the standard deviation of the differences between the hospital mortality rates (expected and adjusted) calculated with the two models. The average difference in the expected rates between the simple logistic model and the multilevel model was -0.007, versus 0.012 as in the adjusted rates. These differences are rather high compared to the average observed rate (0.031), demonstrating that the choice of model could have a significant effect on the estimates of hospital mortality rates. The graph in Figure 1 shows the differences, for each hospital, in adjusted rates as found using the multilevel model as compared to those rates found by using the LR model for two indicators (IQI16 and IQI17). With both methods, the higher the mortality rate in a hospital implies a greater difference between calculated adjusted rates. Figure 2 shows the relation between the observed (OBR) and expected rates (EXR) with LR and ML, and the 95% CI of observed values. In both models it seems that low OBRs do not fit with the calculated EXRs.

Table 3b. Parameter Estimates using the Simple Logistic Model and Multilevel Logistic Model fitted to IQI-17 (acute stroke mortality rate).

Parameters		Simple Logistic Regression			Multilevel Logistic Regression		
		Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Reference: Males; Age 60-64							
Intercept		-2.729	0.115	<0.001	-3.484	0.208	<0.001
sex	Females	-0.320	0.210	0.126	-0.293	0.259	0.258
age 1	18-24	0.510	0.484	0.292	-0.466	0.668	0.485
age 2	25-29	1.055	0.459	0.021	0.235	0.726	0.745
age 3	30-34	0.244	0.479	0.610	-0.050	0.672	0.940
age 4	35-39	0.261	0.349	0.453	-0.792	0.472	0.093
age 5	40-44	0.369	0.178	0.184	-0.001	0.366	0.997
age 6	45-49	0.389	0.226	0.084	0.021	0.295	0.940
age 7	50-54	-0.123	0.220	0.576	-0.495	0.279	0.076
age 8	55-59	0.147	0.175	0.400	-0.117	0.224	0.599
age 9	60-64	0	-	-	0	-	-
age10	65-69	-0.088	0.152	0.561	0.096	0.188	0.607
age11	70-74	-0.044	0.141	0.752	0.267	0.175	0.127
age12	75-79	0.225	0.134	0.094	0.680	0.166	<0.001
age13	80-84	0.446	0.137	0.001	1.032	0.169	<0.001
age14	85+	0.791	0.136	<0.001	1.522	0.167	<0.001
age15	18-24*Fem	-0.377	0.890	0.671	0.160	1.103	0.884
age16	25-29*Fem	0.097	0.790	0.902	0.178	1.136	0.874
age17	30-34*Fem	0.293	0.664	0.658	-0.540	0.898	0.547
age18	35-39*Fem	0.464	0.524	0.375	1.056	0.695	0.128
age19	40-44*Fem	-0.005	0.476	0.991	-0.242	0.608	0.689
age20	45-49*Fem	-0.013	0.413	0.973	-0.208	0.525	0.691
age21	50-54*Fem	0.940	0.339	0.005	0.741	0.439	0.091
age22	55-59*Fem	0.116	0.315	0.710	0.173	0.394	0.660
age23	60-64*Fem	0	-	-	0	-	-
age24	65-69*Fem	0.452	0.257	0.079	0.318	0.318	0.318
age25	70-74*Fem	0.472	0.241	0.050	0.571	0.294	0.052
age26	75-79*Fem	0.216	0.233	0.353	0.208	0.284	0.463
age27	80-84*Fem	0.358	0.231	0.121	0.418	0.281	0.136
age28	85+ *Fem	0.141	0.229	0.537	0.063	0.278	0.818
Variance components					Variance	St. error	p value
σ^2_{Hosp}					0.435	0.312	0.082
$\sigma^2_{Hosp*Spec}$					3.358	0.428	<0.001

Table 4. Wald test statistics and p-value comparing models fitted to IQI-16 (congestive heart failure mortality rate) and IQI-17 (acute stroke mortality rate).

	IQI16		IQI17	
	Wald Chi-square	Pr > ChiSq	Wald Chi-square	Pr > ChiSq
Simple vs. Multilevel	185.991	<0.001	566.693	<0.001
Multilevel vs. Simple	8.043	0.986	148.840	<0.001

The Wald test uses the estimated covariance matrix from the first model indicated in the comparison.

Table 5. Estimated differences (and standard deviation) in provider-level model predictions of expected and risk adjusted rates for IQI-16 (congestive heart failure mortality rate) and IQI-17 (acute stroke mortality rate).

	IQI16		IQI17	
	Mean	SD	Mean	SD
Observed Rate	0.031	0.033	0.081	0.062
Difference from expected rate	-0.007	0.001	-0.018	0.015
Difference from adjusted rate	0.012	0.014	0.032	0.022

Comparison between the AHRQ and the Apulian scores for IQI16.

The reference values applied with the AHRQ score vector are higher (4.36%) compared to those determined with the Puglia region score vector (3.15%). The values for each hospital, when adjusted with the vector of the PHD, are closer to the regional average value and 76.7% of studied hospitals, if adjusted with ML model, 80% of hospitals, if adjusted with LR model, fall within the limits of the confidence interval, as compared with only 35% of hospitals after initial application of the AHRQ scores. The comparison is biased because the AHRQ-adjusted rates are obtained from a score vector with APR-DRG covariates, which its studied population and Apulian score vector do not include.

The IQI17 indicator

The ML model applied to the reference population does not have a single convergence problem, so no covariates need to be excluded. The parameter values and standard errors calculated with the two models are comparable (Table 3b). The variance component associated with the specific random effects of the specialist wards in the multilevel model was $\sigma^2_{Hosp*Spec}$ =3.358 (z=7.84; p<0.001), which was higher than for the IQI16; meanwhile, the component associated with the hospitals was σ^2_{Hosp} = 0.435 (z=1.394;

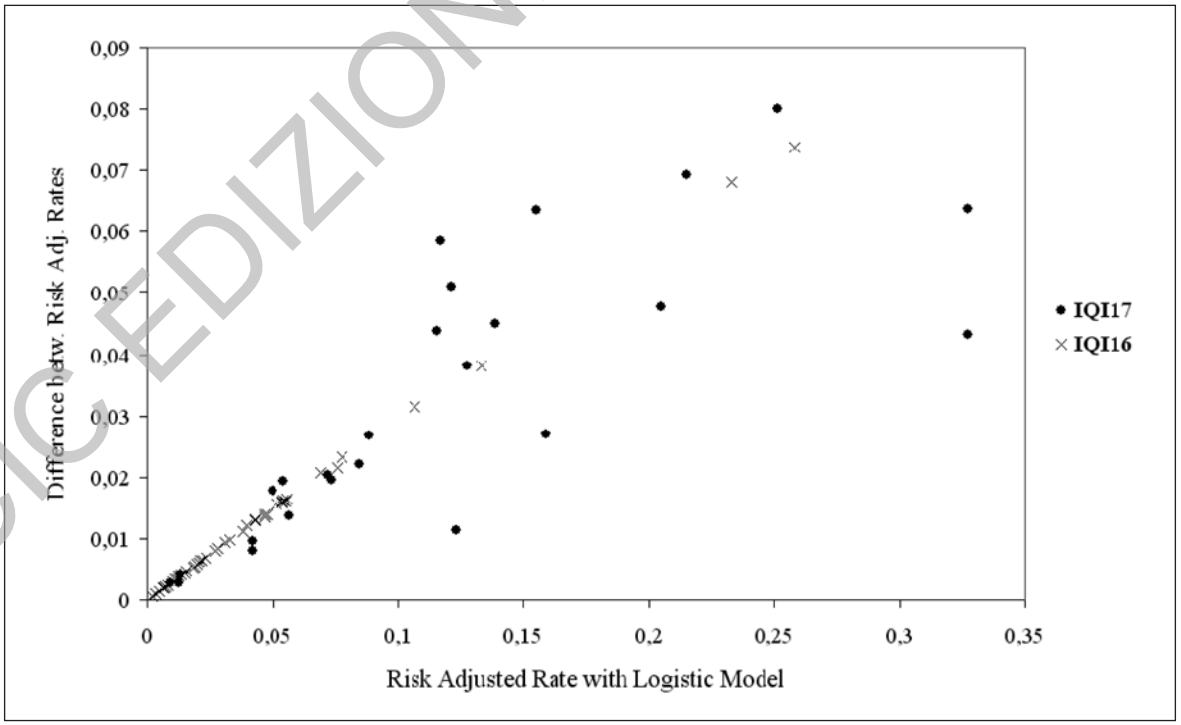

Figure 1. Relation between the adjusted rate of the simple logistic model and the difference between the adjusted rates of the two models.

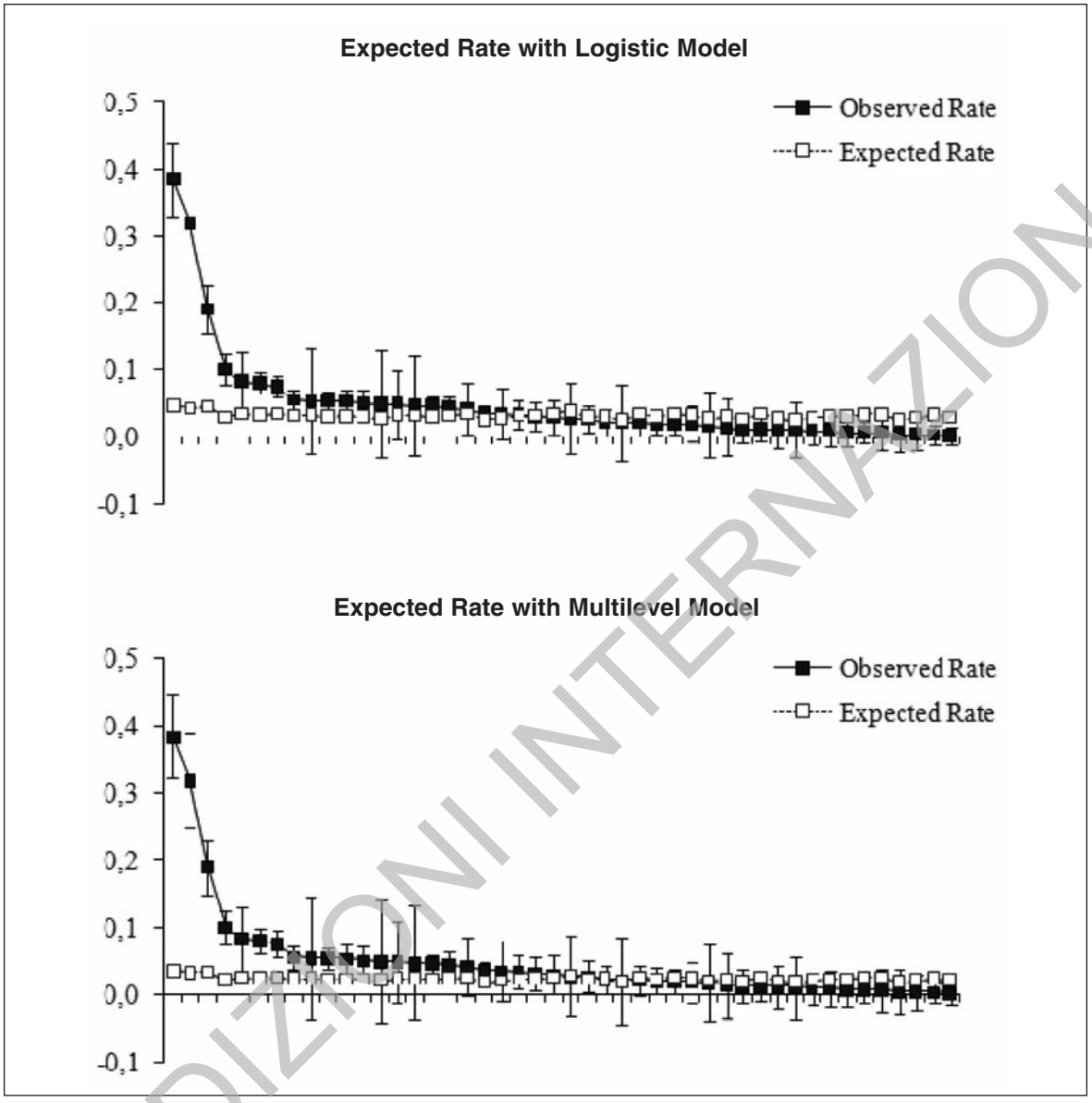

Figure 2. Indicator IQI16. Observed mortality rate with CI 95% and expected mortality rate.

p=0.082), which was lower than for the IQI16, and not statistically significant. The Wald test statistics indicate significant differences between the parameters estimated with the LR and with the ML model ($p<0.0001$) when calculated using either the variance-covariance matrix obtained from the LR model or from the ML model.

The estimate of the ICC value in the ML model was 47.41%-related to the difference between the specialities in the hospitals, and only 6.15%-related to the difference among the hospitals.

The average difference in the expected rates between the two models was -0.018, versus 0.032 for the adjusted rates (Table 5). These differences are rather high compared to the average observed rate (0.081), demonstrating once again that the choice of model could have a significant effect on the estimates of mortality rates. The graph in Figure 1 shows how the different choices in modelling have a greater effect on the higher mortality rates than on the lower ones. Figure 3 shows the relation between the observed (OBR) and expected rates (EXR) using the LR and ML models with a 95% CI of the observed values. In both models the OBRs do not fit with the EXRs for

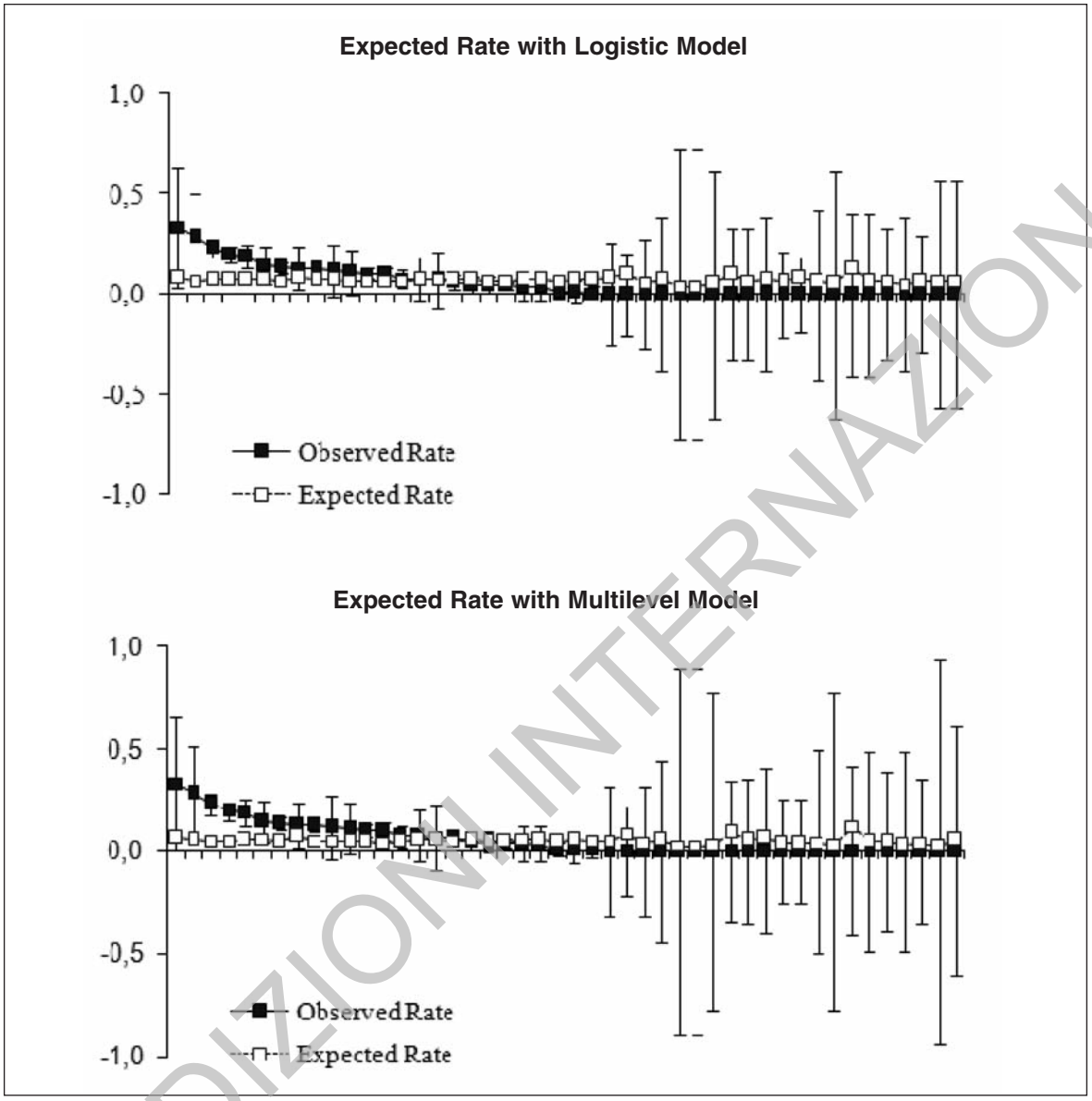

Figure 3. Indicator IQI17. Observed mortality rate with CI 95% and expected mortality rate.

either the low or high rates. The OBRs 95% CIs are wide, especially for the found high rates. This width is probably due to the small number of events and cases.

Comparison between the AHRQ and the Apulian Scores for IQI17

The reference values applied with the AHRQ score vectors are higher (11.2%) compared to those determined with the Puglia region score vectors (7.89%). The values for each hospital, adjusted with the PHD vector, are closer to the regional average value and 83% of hospitals, if adjusted with ML model, 85% of hospitals, if adjusted with LR model, fall within the limits of the confidence interval as compared with 60% after application of the AHRQ scores. The comparison is biased because the AHRQ adjusted rates are obtained from a score vector with APR-DRG covariates, which, as mentioned earlier, studies a population which the Apulian score vector does not include.

Discussion

The main aim of this study was to evaluate the effect of a multilevel logistic model versus a simple logistic model in application to the Risk Adjustment Process. The validity of using multilevel models is supported by recent research conducted by the American Heart Association, which suggests that a hierarchical regression model should be applied to data organised according to distinguishing and mutually-exclusive characteristics, as in the case of hospital information grouped and differentiated by various, described sorting criteria (8, 9). Categorizing individual patients within the first level unit, the hierarchical model treats the second level (speciality wards) and third level (hospitals) units as a random sample of elements in the respective populations, the variability of which is further subdivided into three parts: attributable to the patients characteristics, attributable to the specialist unit and attributable to differences between the hospitals. The variability of these latter two levels shows the real difference between hospitals.

In our data set, the mortality indicator for congestive heart failure was attributable to both speciality wards and hospitals, whereas the mortality indicator for stroke was conditioned only by the speciality wards. This could be due to the fact that stroke patients are usually treated in more specialised hospitals (with departments of neurology, neurosurgery, intensive care, etc.). Therefore, we concluded that there is less variability between the general consideration of hospitals and more variability to the actual healthcare performance offered by the same specialist ward (e.g., intensive care) in different hospitals. Congestive heart failure does not necessarily need to be treated in a highly specialised hospital, so the variability in the mortality rates can be affected both by the specific speciality of the discharge ward and by the type of hospital. In the study by D'Errigo et al. on the outcome following coronary artery by-pass grafts in Italy, the proportion of variability attributable to the hospital was 10.1% (10). In their work on acute myocardial infarction Austin et al. found a percentage of variability equal to 12.6% (11). In the present study, the variability attributable to the hospital was similar in the case of congestive heart failure; however, the discharging specialist ward contributed a much higher variability.

In a ML model, it is very important to identify hierarchical levels for interpreting the performance between macro- and micro-levels (12), so as to choose the covariates correctly. In our study we assumed a nested structure only for the patients and not for the covariates (13).

Applying the Wald test, we found a significant difference between the LR and ML models both using the variance-covariance matrix of the LR model and using the variance-covariance matrix of the ML model, when studying the stroke mortality rate (IQI17). The potential effect of a positive correlation between patients treated in the same hospital could cause significant differences between the vectors of the estimated parameters using the multilevel model or the logistic model. It should be remembered that the simple logistic model treats all the patients, on all levels of hierarchy or grouping, as independent observations of some random process.

As to the IQI17 (mortality for acute stroke), in view of the homogeneity of the results obtained with the two Wald tests we could conclude that unlike the LR model, the multilevel model is able to identify the variability linked to hierarchical levels with the covariates used in the present study. However, with regard to the congestive heart failure indicator (IQI16), the Wald test produced contrasting results. Perhaps the multilevel model requires a more correct definition of the covariates employed, or of the specific assignment of the levels on the intercept as well as on the covariates, as previously stated in (14).

The use of a robust variance-covariance matrix in the multilevel model did not modify the result of the comparison between the models (data not shown) (15, 16). The mortality rates of the hospitals are significantly different when using the hierarchical model in the risk adjustment process as compared to the simple logistic regression model. The use of covariates relative to the severity of the illness or further hierarchical levels could further highlight the greater effectiveness of the ML model.

In this study we have also verified how vital it is in risk adjustment process to choose a suitable reference population, confirming what was stated in the report by Mattke et al. (17). AHRQ adjusted rates could be determined even without employing severi-

ty covariates, using the score vector yielded by the AHRQ, but these scores are conditioned by other covariates included in its own estimation model. Therefore, even if it were possible, adjustment could bias the results and comparison among hospitals. Furthermore, the average regional mortality rates that resulted were higher than those rates obtained using scores determined from the PHD reference population.

The importance of choosing the covariates to be inserted in the model, as according to the characteristics of the reference population and the study population, is highly evident. In their study, Huang et al. (18) concluded that the selection of the risk adjustment variables has more influence on the positioning of the rates than on the statistics strategy applied. Regarding the congestive heart failure analysis, our study confirms this opinion.

In conclusion, it is clear that when applying risk adjustment it is necessary to correctly define not only the model to be used but also a homogeneous reference population. This allows one to obtain valid indicators of outcome and to so promote a more accurate organisational decision-making process.

References

1. AAVV, Identificazione, Sperimentazione e validazione di alcuni indicatori di processo ed esito della qualità delle attività sanitarie-Rapporto conclusivo, Monitor n.5, Supplemento, 2005, p11-157, (Italian).
2. Seccareccia F., Progetto Mattoni, Mattone 8 Misura dell'outcome-Manuale sul risk-adjustment-risultati sui 43 indicatori selezionati. http://www.mattoni.ministerosalute.it/ mattoni/documenti/5_Risultati_dello _studio_dei_43_indicatori_selezionati.pdf (Italian).
3. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. Am J Public Health 1996; 86:1379-87.
4. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. 2. Michigan: Health Administration Press Ann Arbor; 1997.
5. Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in Public Health: Issues in the Application of Multilevel Models. Annu Rev Public Health 2004; 25:53-77.
6. Snijders T, Bosker R. Discrete dependent variables. Multilevel analysis: an introduction to basic and advanced multilevel modelling. London: Sage Publication, 2000.
7. Moinnedin R, Matheson FI, Glazier RH. A simulation study of sample size for multilevel logistic regression models. BMC Medical Research Methodology. 2007; 7:34.
8. Brocco S, Alba N, Rosato R, Spolaore P, Fedelia U. The choice between different statistical approaches to risk-adjustment influenced the identification of outliers. J Clin Epidemiol 2007; 60: 858-862.
9. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes. Circulation 2006; 113:456-462.
10. D'Errigo P, Tosti ME, Fusco D, Perucci CA, Seccareccia F; Research group, Italian CABG Outcome Study. Use of hierarchical models to evaluate performance of cardiac surgery centres in the Italian CABG outcome study. BMC Medical Research Methodology 2007; 7:29.
11. Austin PC, Tu JV, Alter DA. Comparing Hierarchical Modelling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? Am Heart J 2003; 145(1):27-35.
12. Soobader M, Cubbin C, Gee GC, Rosenbaum A, Laurenson J. Levels of analysis for the study of environmental health disparities. Environmental Research 2006; 102:172-1780.
13. Abrahantes JC, Molenberghs G, Burzykowski T, Shkedy Z, Abad AA, Renard D. Choice of units of analysis and modelling strategies in multilevel hierarchical models. Computational Statistics & Data Analysis 2004; 47:537-563.
14. Li Y, Dick AW, Glance LG, Cai X, Mukamel DB. Misspecification issues in risk adjustment and construction of outcome-based quality indicators. Health Services and Outcomes Research Methodology 2007; 7(1-2):39-56.
15. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73:13-22.
16. Fay MP, Graubard BI. Small-Sample Adjustments for Wald-Type. Tests Using Sandwich Estimators. Biometrics 57:1198-1206
17. Mattke S, Kelley E, Scherer P, Hurst J, Gil Lapetra ML and the HCQI Expert Group Members. Health care quality indicators project initial indicators report. OECD Health Working Papers (2006).
18. Huang IC, Dominici F, Frangakis C, Diette GB, Damberg CL, Wu AW. Is risk-adjustor selection more important than statistical approach for provider profiling? Asthma as an example. Medical Decision Making 2005; 25(1):20-34.