Health Care Manag Sci (2013) 16:245–257
DOI 10.1007/s10729-013-9227-1

Comparing health outcomes among hospitals: the experience of the Lombardy Region

Paolo Berta · Chiara Seghieri · Giorgio Vittadini

Received: 21 August 2012 / Accepted: 22 February 2013 / Published online: 27 March 2013

Abstract In recent years, governments and other stakeholders have increasingly used administrative data for measuring healthcare outcomes and building rankings of health care providers. However, the accuracy of such data sources has often been questioned. Starting in 2002, the Lombardy (Italy) regional administration began monitoring hospital care effectiveness on administrative databases using seven outcome measures related to mortality and readmissions. The present study describes the use of benchmarking results of risk-standardized mortality from Lombardy regional hospitals. The data usage is part of a general program of continuous improvement directed to health care service and organizational learning, rather than at penalizing or rewarding hospitals. In particular, hierarchical regression analyses - taking into account mortality variation across hospitals - were conducted separately for each of the most relevant clinical disciplines. Overall mortality was used as the outcome variable and the mix of the hospitals' output was taken into account by means of Diagnosis Related Group data, while also adjusting for both patient and hospital characteristics. Yearly adjusted mortality rates for each hospital were translated into a reporting tool that indicates to healthcare managers at a glance, in a user-friendly and non-threatening format, underachieving and over-performing hospitals. Even considering that benchmarking on risk-adjusted outcomes tend to elicit contrasting public opinions and diverging policymaking, we show that repeated outcome measurements and the development and dissemination of organizational best practices have promoted in Lombardy region implementation of outcome measures in healthcare management and stimulated interest and involvement of healthcare stakeholders.

Keywords Healthcare · Effectiveness · Outcomes · Performance evaluation systems · Multilevel models · DRGs

1 Introduction

Over the past decades, the use of performance assessments in health and social sciences has increased substantially to meet patients' needs, provide effective healthcare services, and to promote quality-improvement initiatives [1]. Objective measures of performance are used at several levels across countries. For instance, the U.K. [2, 3], the U.S. [4, 5], Australia [6], Canada, and institutions such as the World Health Organization [7, 8], and the Organization for Economic Co-operation and Development [9] are actively developing performance indicators for relevant aspects of care, such as effectiveness, efficiency, appropriateness, responsiveness, and equity. These frameworks have been demonstrated to facilitate accountability, modify the behavior of professionals and organizations, and support healthcare management [10–12]. Within this context, healthcare outcomes have often been considered as part of measurement and benchmarking frameworks directed at holding hospitals accountable for the quality of care they deliver [13–17]. Although hospital ranking on outcomes poses many methodological problems, such as case-mix adjustment and estimation of random variation [18–25],

P. Berta · G. Vittadini
Department of Quantitative Methods, CRISP - University of Milan Bicocca, Milan, Italy

C. Seghieri (✉)
Laboratorio Management e Sanità Istituto di management, Scuola Superiore Sant'Anna, Pisa, Piazza Martiri della Libertà, Italy
e-mail: c.seghieri@sssup.it

league tables comparing hospitals' actual patient death rate to statistical predictions are today reported publicly in countries including the U.S., the U.K., Canada, and Denmark (AHRQ, Leapfrog Group, Centers for Medicare and Medicaid Services and Care Quality Commission[1]). Within the Italian healthcare system, the need for performance measurement has grown in urgency since the early 1990s when the government approved the first reform of the National Health Service (Legislative Decrees 502/1992 and 517/1993) [26]. Beginning in the early 1990s, national reforms started transferring several key administrative and organizational responsibilities from the central government to the 21 Italian regional administrations to make regions more sensitive to controlling expenditures and promoting efficiency, quality, and patient satisfaction [27–29]. In this context, a first national pilot performance evaluation system was developed in 2010 on behalf of the Ministry of Health in order to monitor performance across and within regions in terms of quality, efficiency, and appropriateness in the following three domains (settings) of care: hospitals; primary care, including pharmaceutical care; and public health and preventive healthcare [30]. One year later, the Outcome Evaluation National Program (PNE [31]) was introduced at the national level to compare Italian hospitals on outcome measures related to mortality, readmissions, and complications after selected clinical interventions [32]. At the regional level, only a few Italian regional administrations have adopted systematic evaluation programs to evaluate the performance of their regional healthcare systems, and some of these regions have also included clinical outcomes among other performance measures [33, 34]. Among these, starting from 2002 the Lombardy regional administration began comparing the effectiveness of hospitals using outcome measures derived from the Hospital Discharge Chart (HDC) database. This study reports on the methodological aspects and managerial implications the benchmarking analysis of between-hospital risk-standardized outcome measures promoted by the Lombardy. In particular, we describe results of both a univariate and bivariate hierarchical regression model that considers between-hospital variation and also takes into account outcome variation across Diagnosis- Related Group[2](DRGs). We also show how these results have provided healthcare managers with informative insights into hospital quality and have helped to identify services burdened with quality issues. This DRG-based approach to quality of care is original and allows the evaluation to be stratified on specific clinical areas, thus providing valid instrument in support of healthcare management. The paper is structured as follows: Section 2 summarizes the statistical methodological methods of the evaluation system; Section 3 introduces the Lombardy region healthcare system; Section 4 presents methods and data; Section 5 shows the main results of the analysis; Section 6 introduces an application of a bivariate multilevel model as a possible extension of the analysis; and, finally, Section 7 presents a summary of conclusions.

2 Measuring relative effectiveness of healthcare services

In the past few years, there has been growing research on relative effectiveness in healthcare, which, in general, intends to compare the effectiveness of two or more healthcare services, treatments, or interventions available for a given medical condition for a particular set of patients [35, 36]. When considering relative effectiveness as the result of comparative outcome analysis across providers, the designation of different types of outcome measures is of particular relevance. There are in fact many similar definitions of health outcomes: generally, a health outcome is defined as the "technical result of a diagnostic procedure or specific treatment episode" [37], or as a "result, often long term, on the state of patient well-being, generated by the delivery of health service" [38]. However, a clinical outcome such as hospital mortality as the final outcome of treatment in a hospital is considered a crucial measure of the quality of care provided. No other characteristic of healthcare, including process structure, is more closely linked to the mission of health institutions than their activities to prevent or to delay death [13, 15]. However, compared with other kinds of performance measures concerning, for instance, accessibility, appropriateness, and efficiency, the relative effectiveness in terms of outcomes across providers is more complex from both statistical and management perspectives [22, 23]. For instance, when comparing mortality data across providers, special attention should be given to issues such as random variation due to small numbers; variation among providers in case mix and severity of the patients; challenges in defining the right denominators; and data quality issues. In particular, the role of case-mix variation and the development of risk-adjustment models to allow comparison of outcomes among healthcare providers have received a great deal of methodological attention, and extensive literature on this topic has accumulated in recent years [39–41]. That is why there are still doubts

[1] www.ahrq.gov; www.leapfroggroup.org; www.cms.gov; www.cqc.org.uk

[2] DRGs are a classification system that groups hospital patients with similar clinical conditions into 524 diagnostic categories. Clinical conditions are defined by both the patient's principal diagnosis–the main problem requiring care–and other secondary diagnoses. The DRG version utilized in this paper is the 19_{th}

regarding the use of mortality rates in the comparative evaluation of quality of care. Even when elaborated statistical models and risk-adjustment techniques are adopted, unexplained differences in mortality rates have been demonstrated to persist [18, 19, 38, 42]. This is even more critical when mortality is derived from administrative data rather than from a clinical registry. Compared to clinical data, administrative records are less accurate in recording diagnosis and interventions, may lack of clinical information, and do not allow for distinguishing whether complications during hospitalization are attributable to the treatment/medical procedures or depend on conditions present at admission [43]. However, even though the accuracy of administrative data has been questioned, they have been increasingly and still are used by healthcare agencies and other stakeholders to measure hospital quality and create reports to rank institutions or providers [10, 44]. In fact, administrative data which are typically computerized are easily accessible, relatively inexpensive to use, and allow for the collection of information on a large number of individuals or theoretically on the entire population of concern. Although medical records are widely considered as the best source for monitoring adverse events and other clinical information, access to this kind of data is often restricted, and obtaining them may be time consuming. Use of administrative data, on the other hand, is a valuable easily accessible alternative [45–48]. In Lombardy, coding accuracy of the administrative database at the hospital level has been constantly monitored by regional offices so that over the years it has achieved standards providing accurate and reliable clinical description of a patient's care. It has been also suggested that "agencies should facilitate the development and dissemination of a database for best practice and improvement based on the results for primary and secondary research" [18]. In this direction, the use of administrative data that go beyond the scope of health care billing may be extremely useful for disseminating the culture of data reliability and validity. Methodologically, different statistical methods have been proposed for risk adjustment of reported outcome values to account for case-mix differences across healthcare providers, so that the outcomes can be legitimately compared despite differences in risk factors. One of the most straightforward approaches to risk adjust an outcome to compare providers is to estimate an expected value for each provider's outcome based on the relationship between the outcome and its risk factors. Among statistical models, linear and logistic regression models have been extensively adopted by various authors and benchmark agencies to estimate the relationship between an outcome and a set of risk factors [49–51]. However, these standard regression models, especially in the social sciences when the population has a hierarchical structure (i.e. patients in hospitals), might not be adequate to estimate the extent of associations of explanatory variables with the outcome of interest. When applying a standard regression model to hierarchical data, analyses can be carried out either at the individual or aggregate/group level. Regardless of the level that is chosen, the resulting analysis may be flawed for the following reasons: if the analysis is carried out at the individual level and the context in which the process may occur is ignored, key group-level effects may be ignored as well–a problem that is often referred to as the "atomistic fallacy" [52]. On the other hand, if a single-level analysis is applied at the group level by assuming that the results also apply at the individual level, the analysis may be flawed because of problems in making individual-level inferences from group-level analyses. This phenomenon is known as the "ecological fallacy" [52, 53]. In the past few decades, as an alternative to standard regression analysis, a quite extensive literature has proposed the use of multilevel models (also referred to as random-effects models or hierarchical linear models) for studying relationships between outcomes and contextual variables in complex hierarchical structures, considering simultaneously both individual and aggregate levels of analysis and distinguishing between such sources of variation [10, 38, 54–59]. Unlike standard regression models, which assume that the observations are uncorrelated, multilevel models control for the existence of a possible intra-hospital correlation, which may make patients within a hospital more alike in terms of experienced outcome than patients coming from different hospitals, everything else being equal. They indeed allow for comparisons between healthcare providers by adjusting for factors concerning both the case mix of the patients–i.e., the variability of their clinical and social demographic aspects–and factors related to the providers, such as resources and facilities that all together could affect the outcomes of interest. Multilevel models also provide a possible solution to small samples thanks to the adoption of the "shrinkage" estimation which contributes to reducing the chance that small hospitals' performance will fluctuate wildly from year to year or that they will be wrongly classified as either a worse or a better performer [41, 59, 60]. More specifically, for a given patient i within a healthcare provider j, the probability p_{ij} of the occurrence of a dichotomous outcome y_{ij} (i.e., mortality that assumes value 1 if the patient died, 0 otherwise), modeled as a multilevel logistic model can be expressed as

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_{00} + \Sigma_k \beta_{kij} x_{kij} + \Sigma_m \gamma_{mj} z_{mj} + u_j + \varepsilon_{ij} \tag{1}$$

where $i = 1 \dots I$ Patients, $j = 1 \dots J$ Hospitals, $k = 1 \dots K$ Patient Level Covariates and $m = 1 \dots M$ Hospital Level Covariates.

In this equation u_j is the random coefficient for residuals at the hospital level and can be interpreted as the relative effectiveness of hospitals with respect to outcome y_{ij} adjusted for fixed coefficients related to both patient and hospital characteristics (x_{ij}, z_j). More specifically, these u_j estimates showing the specific managerial contribution of the j_{th} health structure to the risk-of-warning event and their 95 % confidence intervals (ICs) identify hospitals with ICs under or over the regional mean of the risk-of-warning event.

3 Characteristics of the healthcare system in Lombardy

The Italian National Healthcare System (NHS) provides universal healthcare coverage throughout the Italian State as a single payer and entitles all citizens, regardless of their social status, to equal access to essential healthcare services. A recent strong policy of devolution has transferred several key administrative and organizational responsibilities and tasks from the central government to the administrations of the 21 Italian regions, which now have significant autonomy on the revenue side and in organizing services designed to meet the needs of their respective populations. Among the 21 regions, Lombardy is one of the top-ranked for socio-demographic indicators. Lombardy has a population of 10 million residents (equal to 16 % of the total Italian population) with a density of 404 inhabitants per km^2 [61], ranks for its economic indicators among the most competitive areas in Europe, and has experienced extended and dynamic entrepreneurship growth. The Lombardy healthcare system comprises approximately 200 hospitals generating 2 million discharges annually; 16 billion Euros are devoted to healthcare expenditures (73 % of the regional budget) every year. A regional reform in 1997 radically transformed the healthcare system in Lombardy into a quasi-open-market healthcare system in which citizens can freely choose the provider, regardless of the ownership (private for profit, private not for profit, or public). In contrast to the rest of the Italian regions, in which each Local Health Authority (LHA) is financed by its region under a global budget with a weighted capitation system and in which the DRG-based hospital-financing system is applied only to teaching hospitals, the healthcare system in Lombardy is entirely built on a prospective payment system based on DRGs, and the reimbursement is for all the providers within the regional accreditation system. Following the 1997 reform, the Lombardy region administration adopted the set of standards defined by the Joint Commission International to evaluate the performance of healthcare organizations in terms of processes and results. This reform also established that the Lombardy administration is responsible for monitoring the effectiveness of the care provided by health providers belonging to the regional accreditation system. As a consequence, the Lombardy Regional Healthcare Directorate, in collaboration with the Interuniversity Research Centre on Public Services (CRISP), developed starting in 2002 a set of performance measures to use alongside the JCAHO criteria to systematically evaluate the performance of healthcare providers in terms of the quality of care provided. This set of measures, which are in line with international evidence on the relative effectiveness of hospitals [5], comprises the following seven outcome measures: (1) intra-hospital mortality, (2) mortality within 30 days after discharge, (3) overall mortality (intra-hospital plus within-30-day mortality), (4) voluntary hospital discharges, (5) readmission to an operating room, (6) inter-hospital transfer of patients, and (7) readmission for the same Major Diagnostic Categories (MDC).

4 Method and data

Since 2002, in Lombardy region, multilevel models applied to regional administrative data have been used by regional healthcare administrators to compare regional hospitals in terms of selected outcomes under the hypothesis that benchmarking analysis contributes to quality improvement and helps overcoming self-referral patterns. Moreover, starting in 2008, regional managers decided to monitor outcomes not only at the hospital level but also for the different DRGs related to each clinical discipline identified through the discharging ward. More exactly, multilevel logistic regression models were conducted separately for each of the most relevant clinical disciplines using overall mortality as outcome variable and taking into account the mix of the hospitals' production (the different DRGs) while also adjusting for both patient and hospital characteristics. These models therefore had three levels: patients discharged in 2009 from any hospital in Lombardy were considered as nested in the hospital, while the DRG variability was controlled by considering the DRG as a pseudo-level [62, 63]. More in detail, to better identify critical areas of the entire ranges of the clinical activities, two random-intercept multilevel models were estimated for each of the selected clinical disciplines: in Model 1, the intercept was considered as being random at both the DRG and hospital levels, while in Model 2 the intercept was considered as being random at the DRG level but fixed at the hospital level. In Model 1, we controlled for overall hospital and DRG effects, and the estimates of these effects indicated where the hospital performance was better or worse than average after adjusting for the relative regional rates for the different

DRGs. Model 2 allowed for the ranking of the hospital-DRG combination with respect to the hospital average mortality. In terms of mathematical formulas, considering a level-1 outcome y_{ijk} taking on a value of 1 with conditional probability p_{ijk}, the two models can be written as follows:

Model 1:

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \alpha_{000} + \Sigma_g \beta_g X_{gijk} + \Sigma_m \lambda_m Z_{m00k} + \gamma_j DRG_j + v_{0jk} + u_{00k} + \varepsilon_{ijk} \quad (2)$$

where:

- $i = 1 \ldots I$ Patients, $j = 1 \ldots J\, DRGs$ and $k = 1 \ldots K$ Hospitals
- $g = 1 \ldots G$ Patients Level Covariaters, $m = 1 \ldots M$ Hospitals Level Covariaters
- γ_j is a fixed coefficient associated with a DRG-specific dummy variable
- v_{0jk} is a random residual associated with the j-th DRG within k-th hospital
- u_{00k} is a random residual associated with the k-th hospital

Model 2:

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \alpha_{0j0} + \Sigma_g \beta_g X_{gijk} + \Sigma_m \lambda_m Z_{m00k} + \gamma_j DRG_{jk} + v_{0jk} + \varepsilon_{ijk} \quad (3)$$

where:

- $i = 1 \ldots I$ Patients, $j = 1 \ldots J\, DRGs$ and $k = 1 \ldots K$ Hospitals
- $g = 1 \ldots G$ Patients Level Covariaters, $m = 1 \ldots M$ Hospitals Level Covariaters
- α_{0j0} is a random intercept associated with the j-th DRG
- γ_j is a fixed coefficient associated with a DRG-specific dummy variable
- v_{0jk} is a random residual associated with the j-th DRG and the k-th hospital within DRG

These two analyses also allowed two types of hospital profiling–a “regional profiling” and a “within-hospital profiling” –on the basis of the estimated DRG odds ratio and the associated interval confidence. That is, for each clinical discipline separately:

- Model 1: the evaluation of a single hospital effectiveness among different DRGs highlights potential areas of improvement (in this case, the reference is the average risk for the given hospital). This is what we call “within-hospital profiling”.
- Model 2: it shows the best and worst practice areas in the set of hospitals with reference to the average risk for the given DRG (the regional average mortality for that same DRG). This is what we call “regional profiling”.

The database was abstracted from the administrative regional healthcare information system, and collected information on patients admitted to 150 hospitals (those hospitals which are accredited with the regional healthcare system and also provide acute care) in the Lombardy region in the year 2009. In 2009, the discharges were 1.900.000, of which 77 % were ordinary and 23 % were in day hospital or day surgery. Moreover, hospitalizations of residents outside the Lombardy region accounted for 10 % of all admissions. The hospital discharge data contains basic demographic information (age, gender), information on hospitalization (length of stay, special-care unit use, transfers within the same hospital or through other facilities, and within-hospital mortality), and a total of 6 diagnosis codes and procedures defined according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). In addition, the linkage of hospitalization data with the patients’ administrative health registry enables reporting of mortality within longer timeframes (30 days). Linkage with other regional databases allows for collecting information on hospital structural characteristics (number of beds, number of operating rooms, etc.). Only ordinary hospitalizations for patients aged more than 2 years were retained in the sample. The analysis was limited to the following clinical disciplines that covered a total of 62 % of the hospital activity: surgery, cardiology, cardio-surgery, medicine, neurology, neurosurgery, oncology, orthopedics, and urology. Selection criteria related to DRGs were discussed with healthcare professionals for each discipline; DRGs that occurred less than 30 times and that were provided from less than three hospitals were excluded. In addition, high-risk (more than 50 % deaths) and low-risk (less than 1 % deaths) DRGs were excluded. The response variable was 30-day mortality, indicating whether or not the patient died within 30 days of hospital discharge. This outcome is obtained by matching two different administrative data sources: the hospitalization data for intra hospital mortality and the healthcare register of all residents in Lombardy for mortality after discharge. Selected variables at both the patient and hospital levels were chosen as major determinants of patient mortality during iterative discussions with regional representatives and physicians. In particular, at the patient level we controlled for the patient’s age (AGE, expressed in years); gender (SEX, a dummy variable equal to 1 if the patient is male); coexisting conditions expressed by the Elixhauser index (COMORB; [64]); the presence of selected comorbidities at admissions such as cardiovascular diseases (CARDIO, expressed as a dummy variable) and cancer (ONCO, expressed as a dummy variable), which were identified by ICD-9-CM diagnosis codes in the

Table 1 Patient's characteristics

Variable	Min	Mean	SD	Max
Age	2	71.76	14.49	108
Gender (Male = 1)	–	0.50	–	–
Cancer disease	0.00	0.10	0.30	1.00
Comorbidity index	0.00	0.55	0.82	5.00
Admission in emergency	0	0.15	0.36	1.00
Length of stay	0.00	9.72	8.15	307
Cardiovascular disease	0.00	0.36	0.48	1.00

principal diagnosis; and the admission trough the emergency department (EMERG, expressed as a dummy variable). Moreover, the following variables were introduced to control for hospital characteristics: a dummy variable indicating the ownership of hospital (public, private for-profit, or private not-for-profit), a dummy variable indicating if the hospital is a teaching or not-teaching hospital, the number of beds and the bed-load factor (expressed as the ratio of total patients to available beds), the number of operating rooms (N_OR), and the hospital mean of the patient-level variables. Table 1 reports the principal patients' characteristics: the mean age of the patients is 72 years and 50.4 % are males. 10 % of them are hospitalized with a principal diagnosis of cancer and 36 % of a cardiovascular disease. 15 % of the patients were hospitalized through the emergency department and the Elixahuser index indicates a mean of 0.52 comorbidity with a maximum of 5.00 and a standard deviation of 0.82. All of the analysis in this article was done using SAS software version 9.2 (SAS Institute, Cary, NC, USA).

5 Principal results

In this section, results for the two principal clinical disciplines, medicine and surgery,[3] are presented. The crude in-hospital death rate for medical and surgical discharges was 16.63 and 4.46 per 100 discharges, respectively. The mean number of discharges per hospital was 660 (range: 554–767) for medical wards and 224 (range: 168–280) for surgical wards. With regard to medical discharges, results of both Models 1 and 2 show that mortality was significantly affected by the patients' age, gender and emergency admission (Table 2).[4] This indicates that older patients and those patients with emergency admission had higher risk of dying within 30 days. For medical disciplines, patients with

[3]Results regarding the other disciplines are available upon request from the authors.

[4]Due to space limitation only statistically significant patient and hospital covariates and a selected number of the estimates at DRG level are displayed in Table 2.

Table 2 Results for medicine ward

Effect	Estimate	StdErr	Probt
Regional profiling			
Intercept	−5.7171	0.297	<.0001
Gender(M vs F)	−0.2386	0.0214	<.0001
Age	0.0472	0.0011	<.0001
EMERG	0.404	0.0542	<.0001
CARDIO	−2.8806	0.1321	<.0001
COMORB	−0.1458	0.0121	<.0001
MEAN_CARDIO	0.4841	0.1438	0.0008
TEACH VS NOTEACH	−0.0952	D 0.0497	0.0561
DRG_015	1.2335	0.3451	0.0004
DRG_416	1.3432	0.2863	<.0001
DRG_395	−0.4353	0.2879	0.1308
DRG_321	−1.3402	0.4466	0.0028
DRG_320	−0.4433	0.3542	0.2110
DRG_316	0.3664	0.2932	0.2117
DRG_294	−1.2054	0.3172	0.0002
DRG_297	−0.7331	0.4471	0.1013
DRG_208	−1.9963	0.5566	0.0003
DRG_207	−0.8466	0.4933	0.0864
DRG_203	1.4753	0.2901	<.0001
DRG_183	−1.8577	0.3272	<.0001
DRG_141	−1.3351	0.4312	0.002
DRG_142	−2.3337	0.451	<.0001
...	...	...	...
...	...	...	...
...	...	...	...
DRG_423	0		
Within-hospital profiling			
Intercept	−7.2811	0.7711	<.0001
Gender(M vs F)	−0.2389	0.02146	<.0001
Age	0.0471	0.0011	<.0001
EMERG	0.3998	0.0541	<.0001
CARDIO	−2.8734	0.1323	<.0001
COMORB	−0.1488	0.0123	<.0001
MEAN_CARDIO	0.0225	0.0097	0.0228
PUBLIC	0.2594	0.1064	0.0164
PRIVATE	0.1519	0.1241	0.2233
NOTFORPROFIT	0		
DRG_015	1.0806	0.3329	0.0012
DRG_416	1.2218	0.2724	<.0001
DRG_395	−0.5710	0.2744	0.0377
DRG_321	−1.5062	0.4327	0.0005
DRG_320	−0.5633	0.3394	0.0973
DRG_316	0.2576	0.2790	0.3561
DRG_294	−1.3320	0.3044	<.0001
DRG_297	−0.8087	0.4350	0.0633
DRG_208	−2.0764	0.5466	0.0002

Table 2 (continued)

Effect	Estimate	StdErr	Probt
DRG_207	−0.9930	0.4785	0.0382
DRG_203	1.3926	0.2757	<.0001
DRG_183	−1.9587	0.3148	<.0001
DRG_141	−1.4617	0.4182	0.0005
...	...	...	...
...	...	...	...
...	...	...	...
DRG_423	0		

multiple comorbidities have a lower risk of dying within 30 days than patients with fewer reported comorbidities. This finding is in contrast with the results in the surgical disciplines. The contrasting results might have resulted from patients with a surgical diagnosis being more critically injured and at greater risk of dying given the same comorbidities than patients diagnosed with a medical condition. Also, medical departments, as opposed to the surgical departments, might reserve more attention and sensibility to the care of comorbidities like hepatic or renal failure, or cardiovascular complications which are typical in internal medicine. Males with respect to females and patients with a cardiovascular diagnosis have a lower probability of dying. By considering the hospital characteristics in medical wards, we obtain two different results with respect to the regional and within-hospital profiling. In the regional profiling, ownership did not significantly affect the risk of dying, while teaching hospitals were significantly associated with a lower risk of dying than non-teaching hospitals. In the within-hospital profiling, public hospitals had a higher risk of mortality compared to the not for profit. Teaching hospital status was not significantly related to overall mortality.

For surgery wards (Table 3), elderly patients with multiple illnesses (Comorbidity Index) and patients admitted through an emergency department had a high likelihood of dying. Also, in contrast to the results for the medical wards, a cardiovascular diagnosis was associated with higher risk of mortality compared to an oncological diagnosis. This is, however, not surprising when considering that oncological patients are usually hospitalized in surgical wards as a first step in their care path, but then are immediately transferred to medical or oncological facilities. Furthermore, being hospitalized in hospitals with a high number of surgical theaters reduces the probability of dying. For all the models the area under the ROC curve was used to assess the discriminative ability of each model in predicting mortality. This area (alternatively named c-index) varies from 0.5 to 1, with larger values denoting better model performance. The c-index ranged between 0.78 and 0.83 across medicine and

Table 3 Results for surgery ward

Effect	Estimate	StdErr	Probt
Regional profiling			
Intercept	−6.0295	0.5338	<.0001
Age	0.0641	0.003243	<.0001
EMERG	0.5271	0.1202	<.0001
CARDIO	1.2407	0.3565	0.0005
ONCO	−0.4048	0.09408	<.0001
COMORB	0.276	0.04542	<.0001
MEAN_EMERG	2.5213	1.0447	0.0164
N_OR	−0.0221	0.007395	0.0029
TEACH VS NOTEACH	0.1359	1.82	0.0698
DRG_149	−3.4406	0.4157	<.0001
DRG_148	−1.7035	0.397	<.0001
DRG_181	−2.7768	0.4203	<.0001
DRG_154	−1.818	0.4313	<.0001
DRG_075	−3.1914	0.5113	<.0001
DRG_208	−3.8867	0.4693	<.0001
DRG_493	−3.4111	0.705	<.0001
DRG_189	−3.702	0.4835	<.0001
DRG_120	−4.5912	0.732	<.0001
DRG_174	−2.353	0.4485	<.0001
...	...	...	...
...	...	...	...
...	...	...	...
DRG_172	0		
Within-hospital profiling			
Intercept	−5.8475	0.5324	<.0001
Age	0.0637	0.003267	<.0001
EMERG	0.5344	0.1211	<.0001
CARDIO	1.2156	0.3571	0.0007
ONCO	−0.4009	0.09486	<.0001
COMORB	0.273	0.04586	<.0001
MEAN_EMERG	2.4312	1.2313	0.0517
N_OR	−0.0109	0.006397	0.0924
DRG_149	−3.3147	0.3891	<.0001
DRG_148	−1.5818	0.3683	<.0001
DRG_181	−2.6753	0.3922	<.0001
DRG_154	−1.6767	0.402	<.0001
DRG_075	−3.0817	0.4928	<.0001
DRG_208	−3.8031	0.4432	<.0001
DRG_493	−3.3038	0.6933	<.0001
DRG_189	−3.5674	0.4552	<.0001
DRG_120	−4.4247	0.7218	<.0001
DRG_174	−2.3065	0.417	<.0001
...	...	...	...
...	...	...	...
...	...	...	...
DRG_172	0		

surgical models denoting a good performance in predicting the outcome. In Table 4 the Intraclass Correlation Coefficient (ICC), defined as the proportion of variance that is accounted for by the group level [65], is reported. For both medicine and surgery wards the ICC of the empty (with only random intecept effects) and complete models (with also hospital and patient characteristics) are statistically significant, denoting the validity of the multilevel approach. Moreover when patient and hospital characteristics were added to the model (complete model), the ICC significantly decreased.

Results on "regional profiling" and "within-hospital profiling", obtained by the estimated DRG odds-ratio and the associated interval confidences, were translated into a reporting tool that indicated to healthcare managers, at a glance, underachieving and over-performing hospitals in terms of 30-day mortality. Table 4 reports, for example, DRG results for the discipline X of hospital Y. For each DRG (table rows), two kinds of information, one for the internal and one for the regional benchmarking, are reported. With regard to the regional benchmarking (first column), a green traffic light indicates that mortality for that DRG is significantly lower than the regional average for the same DRG; in contrast, a red traffic light indicates that mortality is significantly higher, while a yellow traffic light stands for mortality non-significantly different from the regional average. With regard to the within-hospital benchmarking (second column), a green traffic light indicates that mortality for that DRG is significantly lower than the overall hospital mortality, a red traffic light indicates that mortality is significantly higher, while a yellow traffic light stands for mortality non-significantly different from that of the overall hospital. This performance table was specifically designed to provide a visual and easy-to-read layout of performance results across all the estimated DRGs for principal clinical disciplines of all hospitals in Lombardy, thus enabling managers to quickly ascertain whether or not the discipline is performing up to both regional and hospital standards. Table 5 presents an output to be used by the managers as a diagnostic tool: both traffic lights of one DRG in the same colour indicate a best practice (green light) or a critical performance (red light). A red traffic light in the "regional profiling" and a yellow traffic light in "within-hospital profiling" stands for a critical performance. In this case, the performance for that DRG is not significantly different in the "within-hospital profiling", but mortality in the "regional profiling" for that DRG and hospital is significantly higher than mortality for the same DRG delivered by the other hospitals.

This hospital outcome profiling for both surgery and medical disciplines confirmed, at a glance, an overall good

Table 4 Intraclass correlation coefficient

Model	ICC	Pr >Z
Medicine: Regional profiling		
Empty	0.26915	<.0001
Complete	0.05752	<.0001
Medicine: Within hospital		
Empty	0.26913	<.0001
Complete	0.05500	<.0001
Surgery: Regional profiling		
Empty	0.28696	<.0001
Complete	0.08290	<.0001
Medicine: Within hospital		
Empty	0.28704	<.0001
Complete	0.08575	<.0001

Table 5 Information on risk-adjusted 30-day mortality by DRG: discipline X, hospital Y

DRG	Regional profiling	Within-hospital profiling
DRG _148		
DRG _154		
DRG _181		
DRG _149		
DRG _075		
DRG _208		
DRG _189		
DRG _493		
DRG _120		
DRG _207		
DRG _174		
DRG _478		
DRG _479		
DRG _191		
DRG _204		
DRG _131		
DRG _203		

Table 6 Results for surgery ward

Effect	Estimate	StdErr	Probt
Regional profiling			
Readmission	0.1722	0.0501	0.0007
Mortality	0.0333	0.0497	0.5024
DRG_Age	0.0028	0.0060	<.0001
Mean_Emerg	0.0908	0.0459	0.0491
Mean_Onco	−0.0289	0.0154	0.0616
DRG_148	−0.1358	0.0231	<.0001
DRG_149	−0.2073	0.0233	<.0001
DRG_154	−0.1575	0.0250	<.0001
DRG_181	−0.1875	0.0235	<.0001
DRG_075	−0.1855	0.0259	<.0001
DRG_208	−0.1824	0.0242	<.0001
DRG_189	−0.1883	0.0248	<.0001
DRG_493	−0.2048	0.0290	<.0001
DRG_120	−0.2079	0.0305	<.0001
DRG_174	−0.1740	0.0253	<.0001
DRG_207	−0.1566	0.0271	<.0001
DRG_478	−0.1759	0.0288	<.0001
DRG_479	−0.2147	0.0253	<.0001
DRG_191	−0.1614	0.0262	<.0001
DRG_131	−0.2006	0.0301	<.0001
DRG_204	−0.1782	0.0260	<.0001
DRG_203	−0.1080	0.0253	<.0001
DRG_155	−0.1985	0.0266	<.0001
DRG_172	0		
Within-hospital profiling			
Readmission	0.1960	0.0502	0.0002
Mortality	0.0579	0.0497	0.2474
DRG_Age	0.0025	0.0006	<.0001
DRG_Emerg	0.0749	0.0418	0.0737
Mean_Onco	−0.0329	0.0159	0.0414
DRG_148	−0.1424	0.0233	<.0001
DRG_149	−0.2111	0.0231	<.0001
DRG_154	−0.1597	0.0247	<.0001
DRG_181	−0.1867	0.0233	<.0001
DRG_075	−0.1894	0.0257	<.0001
DRG_208	−0.1829	0.0239	<.0001
DRG_189	−0.1918	0.0245	<.0001
DRG_493	−0.2130	0.0291	<.0001
DRG_120	−0.2098	0.0300	<.0001
DRG_174	−0.1803	0.0254	<.0001
DRG_207	−0.1574	0.0266	<.0001
DRG_478	−0.1766	0.0285	<.0001
DRG_479	−0.2150	0.0250	<.0001
DRG_191	−0.1652	0.0258	<.0001
DRG_131	−0.1980	0.0297	<.0001
DRG_204	−0.2386	0.0433	<.0001
DRG_203	−0.1070	0.0249	<.0001
DRG_155	−0.1983	0.0262	<.0001
DRG_172	0		

global performance, in line with other performance measures of care [66], thus suggesting that Lombardy region healthcare system is one of best-performing in Italy. However, the findings also revealed the need to improve patient outcomes at some hospitals or in given areas of care. These analyses were performed yearly by CRISP researchers and then presented to hospital managers in meetings with regional managers. Once hospital CEOs received these analyses, the hospitals, together with physician partners and with leadership and support from the regional managers, were asked to collaborate with physicians and other clinicians of their organizations to better address the causes of the adverse outcomes. This method has also been shared and tested through a collaboration with the Tuscany region healthcare administration and the National Agency for Regional Healthcare Services (AGENAS). With regard to Tuscany region, a multi-dimensional Performance Evaluation System (PES) was first implemented in 2005 to measure the quality of services provided by the Tuscan healthcare system. The Tuscan PES has evolved over years and now consists of 50 performance composites and more than 130 simple indicators [30]. In 2010, the Tuscany Region decided to also include measures of outcomes i.e., risk-adjusted mortality at the hospital level - in the PES. Thanks to collaborations with the Lombardy administration, multilevel methods were applied to Tuscan administrative data and both methodology and results were shared in meetings with regional administrators and CEOs of health authorities in order to initiate information sharing with all healthcare stakeholder before including these new indicators in the Tuscan PES. At the same time, the method was tested on the national administrative data provided by AGENEAS. The study analyzed the performance for 9 regions and more than 3.800.000 HDCs for 568 hospitals.

6 The bivariate multilevel model

As already mentioned, Lombardy Region reports annual data on the quality of care provided by hospitals as measured by a set of quality indicators. All the measures have been endorsed by and shared with healthcare professionals. Among these, both mortality and readmissions within

Table 7 Covariance parameter estimates

Subject	Group	Estimate	Pr >Z
Regional benchmark			
–	HOSP(DRG)	0.0016	0.5000
Residual	Readmission	0.0159	0.4996
Residual	Mortality	0.0002	0.5000
Within hospital			
–	HOSP	0.0009	0.1153
–	DRG(HOSP)	0.0015	0.4998
Residual	Readmission	0.0159	0.4983
Residual	Mortality	0.0001	0.5000

30 days are widely used in the literature as quality-of-care indicators. Within this context, a possible development of the hospital benchmarking analysis is to show how performance varies simultaneously across the two quality indicators, mortality and readmissions, by fitting a bivariate multilevel model to Lombardy hospitalization data [67]. The multivariate model is an extension of a univariate model where the two outcomes are simultaneously modeled and regressed on covariates and the correlation between outcomes at all levels are estimated [65]. One equation per outcome is considered and the responses are treated as defining the lowest level of the hierarchy, being nested, in this case, within patients [68, 69]. In the present analysis, this means fitting a four level multivariate model (outcomes,patients, DRGs, hospitals) which, in turn, makes the estimation procedures in SAS computationally complex. As a consequence, patient data were aggregated at the DRG level to handle this complex data and three level (outcomes-DRGs-hospitals) bivariate models were estimated separately for the major clinical disciplines. This new approach, although causing a loss of information at the patient level, allows for modeling, all at the same time, the two outcome measures and it provides healthcare managers with additional information about the possible correlations between these measures. For instance, results from surgical wards data are shown in Tables 6 and 7. In particular, Table 7 shows that, differently from the univariate models, variation at both DRG and hospital level are estimated to be no significant. This result suggests that the univariate model might be preferable to the bivariate model since it allows to disantangle differences in the outcome among different DRGs within the same hospital and among hospital for the same DRG.

7 Conclusion

The results addressed the fundamental objectives of the project: to provide hospitals with periodic feedback reports on their performance, at the clinical discipline level, with respect to adjusted mortality rates. Although studies on this topic tend to elicit diverging opinions regarding quality-of-care indicators and performance on risk-adjusted outcomes (particularly mortality rates), the choice of hospital benchmarking on the basis of adjusted outcomes in Lombardy region, offered in a user-friendly and non-threatening format, presents a promising alternative for helping stakeholders and health structures to detect trends and outliers. The principal objective behind the constant use of these outcome measures was to create a "culture of evaluation" as part of a general program of continuous improvement and organizational learning, rather than creating instruments to publicly penalize or reward hospitals [18, 19]. A major success has been the attention this initiative has received at all levels of healthcare organizations. Because it used regional administrative data, healthcare employees were more likely to accept the results rather than thinking they had been "manufactured" to make a point. At the same time, hospitals became more aware of how the organization was performing compared with its peers. Then, they were asked to identify and debate the structural, institutional, and human factors that could explain good or poor outcomes, and they were requested to come up with suggestions for improvement. By focusing on DRGs, areas that needed improvement were prioritized, and there was the possibility of sharing best practices among benchmarking partners. Over the years, repeated outcome measurements and the development and dissemination of organizational best practices have promoted acceptance of the outcome measures within the performance-evaluation system of Lombardy throughout healthcare organizations and have stimulated the interest and involvement of professionals. Nevertheless, the analysis of each DRG separately (ex-ante stratification) (1) helped to keep the risk of comparing non-comparable structures to a minimum since it works as a powerful standardizing or risk-adjustment mechanism that allows the evaluation of quality outcomes across different hospitals [70], and (2) helped healthcare managers to develop meaningful comparisons of relative effectiveness. To date, only a few studies to the best of our knowledge have examined whether hospital DRG case-mix risk can be used in a multilevel analysis to reveal the performance of health providers across the entire spectrum of hospital conditions as an alternative to mortality results for selected conditions such as heart attack, stroke, and pneumonia [71]. Therefore, this type of performance results should be interpreted with caution. The principal objective of the DRG classification system is in fact to provide a means for relating the type of patients a hospital treats with the costs incurred by the hospital by defining homogeneous groups of patients such as those requiring similar facilities, similar levels of organization, and similar diagnostic procedures. Although DRGs are primarily employed for prospective payment to hospitals, in recent years they have also been used by governments and healthcare providers for other purposes, including adjusting comparisons of quality measures between hospitals with no financial purpose [72, 73]. Moreover, the DRG determination depends on ICD–9–CM coding process reliability, the sequence of codes, whether complications and/or comorbidities exist, and other factors [43]. For these reasons, some researchers questioned the use of DRGs to reflect clinical attribute and then adjust for severity. However in the last years coding guidelines and regulations for controlling the quality of coding, has contributed to significantly improve the quality of medical data available in administrative information systems [74]. In Lombardy region several studies have been conducted regarding reliability of administrative data for provider benchmarking [75] and attempts have been made by Lombardy administrators to legislate fraudulent practices like upcoding (a practice which consists in classifying a patient in a DRG that produces a higher reimbursement). With regard to upcoding, for the most complex DRGs, Lombardy region set a minimum length of stay beyond which the DRG is considered as high resource intensive and it is assigned to a higher tariff than it would have been if the LOS would be lower than this minimum. Outcomes reports should present the information clearly, and all possible biases should be reported and well explained so that even non-experts may be able to understand and compare their quality. The issue of public disclosure of outcomes has been extensively debated in the past few years, with some supporting its efficacy in driving improvements in quality and others believing that it promotes risk-averse behavior by providers by discouraging physicians from accepting high-risk patients [76–78]. The Lombardy regional administration, as a first step, decided to make the results available to healthcare providers to internally stimulate their accountability. Future steps include public release of the results to guide patients choices, pending the understanding from physicians of the value to benchmarking analysis as a tool for improving their performances. Further research is, however, needed to better refine factors at both the individual and aggregate levels that might affect mortality and could be easily measured and introduced into the risk-adjustment model. In this sense, a managerial effort is needed to promote policies regarding the improvement of available administrative resources. Despite the high quality of the data system in Lombardy, hospital discharge data do not allow us to distinguish whether complications arising during hospitalization are imputable to the treatment/medical procedures or depend on conditions present at admission [43]. Furthermore, these data do not collect information on the gravity of illness at admission, and electronic procedures to infer comorbidities or to aggregate DRGs in groups homogeneous by gravity have not been implemented in regional information systems. In summary, the present paper showed that the appropriate use of data–presented in a direct and easy-to-read layout without ranking purposes–and constant discussion of the results with all of the stakeholders helped to improve outcomes, to make providers' practices more efficient, and at the same time to encourage researchers and healthcare managers to design improvements in administrative databases.

References

1. Donabedian A (2005) Evaluating the quality of medical care. Milbank Q 83:691–729
2. Appleby J, Harrison A (2000) Health care in the UK 2000: the King's fund review of health policy. King's Fund, London
3. Department of Health (2008) The NHS in England: the operating framework for 2009/10. DoH, London
4. The Commonwealth Fund Commission on a High Performance Health System (2006) Why not the best? Results from a national scorecard on US. The Commonwealth Fund, New York
5. Agency for Healthcare Research and Quality (2012) Guide to inpatient quality indicators, US department of health and human services. Computer Science, Rockville
6. Braithwaite J, Healy J, Dwan K (2005) The governance of health safety and quality. Commonwealth of Australia, Canberra
7. World Health Organization (2000) The world health report 2000: health systems improving performance. World Health Organization, Geneva
8. Canadian Institute for Health Information (2009) Hospital report 2008: acute care. Ottawa, Ontario
9. OECD (2007) OECD health care quality indicators project. Organisation for Economic Co-operation and Development, Paris
10. Ash A, Fienberg S, Louis T, Normand SL, Stukel T, Utts J (2012) Statistical issues in assessing hospital performance. COPPS-CMS White Paper
11. Nuti S, Vainieri M, Seghieri C (2013) Assessing the effectiveness of a performance evaluation system in the public health care sector: some novel evidence from the Tuscany Region experience. J Manag Gov 17(1):59–69
12. Pinnarelli L, Nuti S, Sorge C, Davoli M, Fusco D, Agabiti N, Vainieri M, Perucci CA (2011) What drives hospital performance? The impact of comparative outcome evaluation of patients admitted for hip fracture in two Italian regions. Br Med J Qual Saf 21:127–134
13. Donabedian A (1992) The role of outcomes in quality assessment and assurance. QRB Qual Rev Bull 11:356–360
14. Joint Commission on Accreditation of Healthcare Organization (JCAHO) (1994) A guide to establishing programs for assessing outcomes in clinical settings. Joint Commission on Accreditation of Healthcare Organizations, Oakbrook Terrace
15. Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand SL (2006) An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 113(13):1683–1692
16. Vittadini G, Sanarico M, Berta P (2006) Testing procedures for multilevel models with administrative data. In: Zani S, Cerioli A, Riani M, Vichi M (eds) Data analysis, classification and the forward search. Springer-Verlag, Heidelberg, pp 329–337

17. Arah OA, Klazinga NS, Delnoij DM, ten Asbroek AH, Custers T (2003) Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. Int J Qual Health C 15(5):377–398
18. Lilford R, Mohammed MA, Spiegelhalter DJ, Thomson R (2004) Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet 364:1147–1154
19. Lilford R, Pronovost P (2010) Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. Br Med J 340:c2016
20. Weiss KB, Wagner R (2000) Performance measurement through audit, feedback, and profiling as tools for improving clinical care. Chest 118(2 Suppl):53S–58S
21. Brook RH, McGlynn EA, Shekelle PG (2000) Defining and measuring quality of care: a perspective from US researchers. Int J Qual Health 12(4):281–295
22. Iezzoni LI (2003) Risk adjustment for measuring health care outcomes, 3rd edn. Health Administration Press, Chicago
23. Normand ST, Shahian DM (2007) Statistical and clinical aspects of hospital outcomes profiling. Stat Sci 22(2):206–226
24. Shahian DM, Robert EW, Iezzoni LI, Kirle L, Normand SL (2010) Variability in the measurement of hospital-wide mortality rates. New Engl J Med 363:2530–2539
25. Brook RH, Iezzoni LI, Jencks SF, Knaus WA, Krakauer H, Lohr KN, Moskowitz MA (1987) Symposium: case-mix measurement and assessing quality of hospital care. Health Care Financ R. Spec No: 39-48
26. Lo Scalzo A, Donatini A, Orzella L, Cicchetti A, Profili S, Maresso A (2009) Italy: Health system review. In: Health systems in transition. Copenhagen: WHO regional office for Europe on behalf of the European observatory on health systems and policies
27. Formez (2007) I Sistemi Di Governance Dei Servizi Sanitari Regionali. Formez, Roma
28. Censis (2008) I modelli decisionali nella sanità Locale. Censis, Roma
29. Antonini L, Pin A (2009) The Italian road to fiscal federalism. Ital J Public Law 1:16
30. Nuti S, Seghieri C, Vainieri M, Zett S (2012) Assessment and improvement of the Italian Healthcare system: first evidences from a pilot national performance evaluation system. J Healthc Manag 53:3
31. Fusco D, Barone AP, Sorge C, D'Ovidio M, Stafoggia M, Lallo A, Davoli A, Perucci CA (2012) P.Re.Val.E.: Outcome research program for the evaluation of health care quality in Lazio, Italy. BMC Health Serv Res 12:25
32. Basiglini A, Moirano F, Perucci CA (2011) Valutazioni comparative di esito in Italia: ipotesi di utilizzazione e di impatto. Mecosan 20(78):9–35
33. Nuti S (2008) La valutazione della performance in sanità. Il Mulino, Bologna
34. Vittadini G (2010) La valutazione della qualità nel sistema sanitario analisi dell'efficacia ospedaliera in Lombardia. Guerini e Associati, Milano
35. Donabedian A (1990) La qualità dell'assistenza sanitaria. Principi e metodologie di valutazione, La Nuova Italia Scientifica, Roma
36. Pagano A, Rossi C (1999) La valutazione dei servizi sanitari. In: Gori E., Vittadini G (eds) Qualità e Valutazione nei servizi di pubblica utilità, Serie gestione d'impresa e direzione. Etas, Milano
37. Opit LJ (1993) The measurement of health service outcomes. Oxford Textbook of Health Care, 10, OLJ, London
38. Goldstein H, Spiegelhalter DJ (1996) League table and their limitations: statistical issues in comparisons of institutional performances. J Roy Stat Soc A Sta 159(3):385–443
39. Werner RM, Bradlow ET (2006) Relationship between Medicare's Hospital compare performance measures and mortality rates. JAMA-J Am Med Assoc 296(22):2694–2702
40. Jha AK, Orav EJ, Li Z, Epstein AM (2007) The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 26(4):1104–1110
41. Zaslavsky A (2001) Statistical issues in reporting quality data: small samples and Casemix variation. Int J Qual Health C 13(6):481–488
42. Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J (1988) Interpreting hospital mortality data. JAMA-J Am Med Assoc 260(24):3611–3616
43. Iezzoni LI (1997) The risks of risk adjustment. JAMA-J Am Med Assoc 278(19):1600–1607
44. Goldman LE, Chu P, Osmond D, Bindman A (2011) The accuracy of present-on-admission reporting in administrative data. Health Serv Res 46(6pt1):1946–1962
45. Austin PC, Tu JV (2006) Comparing clinical data with administrative data for producing acute myocardial infarction report cards. J Roy Stat Soc A Sta 169(1):115–126
46. Vincent C, Neale G, Woloshynowych M (2001) Adverse outcomes in British hospitals: preliminary retrospective record review. Brit Med J 322:517–519
47. Michel P, Quenon JL, de Sarasqueta AM, Scemama O (2004) Comparison of three methods for estimating rates of adverse outcomes and rates of proutcomeable adverse outcomes in acute care hospitals. Brit Med J 328:199
48. Van den Heede K, Sermeus W, Diya L, Lesaffre E, Vleugels A (2006) Adverse outcomes in Belgian acute hospitals: retrospective analysis of the national hospital discharge dataset. Int J Qual Health C 18(3):211–219
49. Dubois RW, Brook RH, Rogers WH (1987) Adjusted hospital death rates: potential screen for quality of medical care. Am J Public Health 77:1162–1167
50. Canadian Institute for Health Information (2003) Hospital report 2002, acute care technical summary
51. Agency for Healthcare Research and Quality (2003) Guide to inpatient quality indicators, US department of health and human services. Computer Science, Rockville
52. Hox JJ (1995) Applied multilevel analysis. TT-Publikaties, Amsterdam
53. Robinson WS (1950) Ecological corelation ant the behavior of individuals. Am Sociol Rev 15:351–357
54. Goldstein H (1995) Multilevel statistical models. Arnold, London
55. Thomas N, Longford NT, Rolph JE (1994) Empirical Bayes methods for estimating hospital specific mortality rates. Stat Med 13:889–903
56. Normand SL, Glickman M, Gatsonis CA (1997) Statistical methods for profiling providers of medical care: issues and application. J Am Stat Assoc 92(439):803–814
57. Rice N, Leyland A (1996) Multilevel models: applications to health data. J Health Serv Res Policy 1:154–164
58. Leyland AH, Boddy FA (1998) League tables and acute myocardial infarction. Lancet 351:555–558
59. Marshall EC, Spiegelhalter DJ (2001) Institutional performance. In: Goldstein H, Leyland AH (eds) Multilevel modelling of health statistics. Wiley, Chichester, pp 127–142
60. Iezzoni LI, Ash SA, Schwartz M, Daley J, Hughes JS, Mackiernan YD (1996) Judging hospitals by severity-adjusted mortality rates:

the influence of the severity adjustment method. Am J Public Health 86:1379–1387
61. Istat (2009) Demografia in cifre
62. Browne WJ, Subramanian SV, Jones K, Goldstein H (2005) Variance partitioning in multilevel logistic models that exhibit overdispersion. J Roy Stat Soc A Sta 168:599–613
63. Levin KA, Leyland AH (2005) Urban/rural inequalities in suicide in Scotland, 1981–1999. Soc Sci Med 60:2877–2890
64. Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. Med Care 36: 8–27
65. Snijders TAB, Bosker RJ (2011) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage Publishers, New York
66. SIVEAS (2010) http://www.salute.gov.it. Accessed 15 Sept 2012
67. Hauck K, Street A (2006) Performance assessment in the context of multiple objectives: a multivariate multilevel analysis. J Health Econ 25:1029–1048
68. Wright SP (2011) Multivariate analysis using the MIXED procedure. Paper presented at the Twenty-Third Annual Meeting of SAS Users? Group International, Nashville, TN (Paper 229). Retrieved March 14
69. Thiebaut R, Jacqmin–Gadda H, Chene G, Leport C, Commenges D (2002) Bivariate linear mixed models using SAS PROC MIXED. Comput Meth Prog Bio 69:249–256
70. The World Bank (2010) Hospital performance and health quality improvements in So Paulo (Brazil) and Maryland (USA). The World Bank En Breve, No. 156
71. Ramos Rincón JM, Garca Ruipérez D, Aliaga Matas F, Lozano Cutillas MC, Llanos Llanos R, Herrero Huerta F (2001) Specific mortality rates by DRG and main diagnosis according to CIE-9-MC at a level-II hospital. Ann Med Interna 18(10): 510–516
72. Rosenberg MA, Browne MJ (2001) The impact of the inpatient prospective payment system and diagnosis-related groups: a survey of the literature. N Am Actuar J 5:84–94
73. Librero J, Marn M, Peir S, Verdaguer Munujos A (2004) Exploring the impact of complications on length of stay in major surgery diagnosis-related groups. Int J Qual Health C 16:51–57
74. Busse R, Geissler A, Quentin W, Wiley MW (2011) Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals. McGraw-Hill, Open University Press, pp 9–21
75. Berta P, Callea G, Martini G, Vittadini G (2010) The effects of upcoding, cream skimming and readmissions on the Italian hospitals efficiency: a population-based investigation. Econ Model 27(4):812–821
76. Chassin MR, Hannan EL, DeBuono BA (1996) Benefits and hazards of reporting medical outcomes publicly. New Engl J Med 334(6):394–398
77. Werner RM, Asch DA (2005) The unintended consequences of publically reporting quality information. JAMA-J Am Med Assoc 293:1239–1244
78. Keogh B, Spiegelhalter D, Bailey A, Roxburg J, Magee P, Hilton C (2004) The legacy of Bristol: public disclosure of individual surgeons' results. Br Med J 329(7463):450–454