Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models

Francesca Ieva · Anna Maria Paganoni

Received: 28 May 2013 / Accepted: 16 December 2013 Published online: 10 January 2014

Abstract In this work we propose the use of a graphical diagnostic tool (the funnel plot) to detect outliers among hospitals that treat patients affected by Acute Myocardial Infarction (AMI). We consider an application to data on AMI hospitalizations recorded in the administrative databases of our regional district. The outcome of interest is the in-hospital mortality, a variable indicating if the patient has been discharged dead or alive. We then compare the results obtained by graphical diagnostic tools with those arising from fitting parametric mixed effects models to the same data.

Keywords Funnel plots · Generalized linear mixed models · In-hospital survival · Provider profiling · Acute Myocardial Infarction

Mathematics Subject Classifications (2010) MSC 62P10 · MSC 90B50

1 Introduction

It has become common to adopt a hierarchical model structure (i.e., a multilevel model that takes into account the grouped nature of data) when comparing the performance of healthcare providers. It is not immediately clear, however, how unusual providers, that is, any with particularly high or low rates, can be identified based on such a model. Outlier detection in provider profiling is a growing interest topic in decisional processes related to healthcare regulation [2, 3, 11, 16, 18]. In this context we consider outliers the institutions that significantly differ from other hospitals under/over performing in the quality of treatments provided to patients affected by the same pathology. Once these hospitals have been detected, people in charge of health care planning may set up monitoring and investigation strategies to understand the causes of these occurrences.

In fact, measuring and understanding process variability and its impact on principal outcomes are necessary for a significant improvement of healthcare systems. We focus our readings on the patterns of care undertaken by patients affected by Acute Myocardial Infarction (AMI) and hospitalized in a structure of the Northern Italy regional district named Regione Lombardia. The nature of such data is hierarchical (patients within hospitals). The main outcome of interest is the in-hospital mortality, a binary variable that indicates if the patient has been discharged alive or not. We fit suitable hierarchical logistic parametric models to these data with a twofold aim: (i) to estimate the hospitals Standardized Mortality Ratios to be graphically compared and (ii) to compute estimates of the random effects, after adjusting for significant covariates. The joint use of funnel plots and diagnostic graphical tools on the random effects estimates, leads to an effective way for detecting outliers in terms of hospital performance. These graphical tools provide an effective way to summarize information to be shared with and communicate to clinicians. Moreover, they enhance possibility to convey key conclusions for supporting decision makers in the healthcare regulation process. In particular we mainly refer to [13, 22]and[23], among others, for the use of funnel plots in healthcare regulation and

F. Ieva (✉) · A. M. Paganoni
MOX - Modellistica e Calcolo Scientifico,
Dipartimento di Matematica, Politecnico di Milano,
via Bonardi 9, 20133 Milan, Italy
e-mail: francesca.ieva@polimi.it

A. M. Paganoni
e-mail: anna.paganoni@polimi.it

to [9] for the use of graphical methods and diagnostic analysis on random effect estimates. The usual league tables or caterpillar plots have been criticized as leading to a spurious focus on rank ordering. The funnel plots, not only overcome these criticisms, but they are also very simple to construct and understand. Besides, this graphical tool is an efficient way to flag the out of control structures. Then the study of the value of the related overdisperision allows for a more quantitative estimate of the outlyingness.

The article is structured as follows: first we present the data we deal with (Section 2), then we present funnel plots to detect outliers and extreme structures with respect to Standardized Mortality Ratio (Section 3). Mixed effects models and the study of random effects estimates are detailed in Section 4. In Section 5 we discuss the achieved results. All the analyses have been performed with R program [17].

2 Data and extraction criteria

Administrative health care databases play today a central role in the evaluation of healthcare systems, because of their widespread diffusion and the real-time, low-cost information they provide (see [10, 25], among others). There is an increasing agreement among epidemiologists on the validity of disease and intervention registries based on administrative databases (see, for example, [4, 5] and [24] and references therein). Therefore more and more frequently administrative data are used to address epidemiological issues in observational studies. The most critical issue when using administrative databases for observational studies is represented by the selection criteria of the statistical units. In fact several different criteria may be used, and they will result in different images of prevalence or incidence of diseases. In the case of interest, we focus on in-hospital mortality after an Acute Myocardial Infarction (AMI). Concerning this pathology, since every hospital admission ends in a record collected in the administrative datawarehouse, the database of SDO (*Scheda di Dimissione Ospedaliera*, i.e., hospital discharge paper) has been used in order to identify AMI episodes and related subsequent hospitalizations. In fact, the SDO database contains data for each hospitalization that a patient experiences along time, providing information both on patient features (in terms of sex, age) and on her/his hospitalization details (date of admission and discharge, diagnoses and procedures, type of admission, type of discharge, vital status at discharge, hospital of admission/discharge). A detailed list (in italian) of fields contained into the SDO is provided in [20].

The case study presented here concerns data arising from a project named "Exploitation, integration and study of current and future health databases in Lombardia for Acute Myocardial Infarction", funded by Ministry of Health and Regione Lombardia, the region in the northern part of Italy whose capital is Milan. The principal aim of the project was to exploit clinical registries and administrative databases for evaluating the performance of the cardiological network in treating AMI patients and to carry out an effective healthcare planning based on real evidence and needs.

In order to identify AMI patients within the administrative database we considered admissions ended between 2000 and 2010 and classified within the Major Diagnostic Categories (MDC) 01 - *Nervous System*, 04 - *Respiratory System* and 05 - *Circulatory System*. Among these records, admissions for acute myocardial infarction have been identified as those presenting an AMI code in the first diagnosis field among the six available in the SDO. The list of ICD-9-CM codes referred to AMI has been created following Agency for Healthcare Research and Quality - AHRQ [1] and consists in the following ones: 41000; 41001; 41010; 41011; 41020; 41021; 41030; 41031; 41040; 41041; 41050; 41051; 41060; 41061; 41070; 41071; 41080; 41081; 41090; 41091. In particular AMI cases are the ones pointed out by Impatient Quality Indicator 15 that automatically exclude patients transferred by another hospital and patients under 18. We do not consider also hospitals managing a number of cases less than 20. The dataset consists then of 46079 patients treated in 96 hospitals of Regione Lombardia.

3 Funnel plots for Standardized Mortality Ratio

The first statistical method we consider for outliers detection is based on funnel plots. The funnel plots have been originally introduced in meta-analysis studies, with the primary aim of being a visual aid to detection of bias or systematic heterogeneity (see, for example [8, 14]). More recently they have been proposed as a graphical aid for institutional comparisons, especially because they overcome some criticism of the more traditional caterpillar plots [15, 21]. The aim of funnel plots is to provide a simple and effective graphical tool for identifying unusual performances, adjusting for the possibly overdispersion the phenomenon under study is characterized by.

Following [13] we consider three different approaches to identify unusual performance of providers using the administrative data previously introduced: (a) a funnel plot based on the common mean model, (b) a funnel plot to identify outliers in the random effects distribution and (c) a funnel plot to identify extremes in the random effects distribution. In the construction of a funnel plot we need, an indicator of performance X, a target θ_0 which represents the desired expectation, so that $\mathbf{E}[X] = \theta_0$ for institution "in-control" and a precision parameter ρ that controls the accuracy of the measured indicator. In general ρ is inversally proportional to the variance, i.e., $\rho \propto \frac{1}{Var[X]}$. Given a series of J observations x_j, and the associated precisions ρ_js a funnel plot is a scatterplot of x_j versus ρ_j, with superimposed control limit lines, function of ρ, computed according to the chosen model for accounting overdisperion. In our case we chose as performance indicator of each hospital the Standardized Mortality Ratio (SSR) defined as

$$SSR_j = \frac{\sum_{i=1}^{n_j} y_{ij}^{obs}}{\sum_{i=1}^{n_j} \hat{p}_{ij}} = \frac{O_j}{E_j}, \; j = 1, ..., J. \tag{1}$$

where y_{ij}^{obs} is the observed outcome of in-hospital mortality for patient i treated in the hospital j, n_j is the number of patients treated in hospital j and $\hat{p}_{ij}$ is the corresponding estimated probability of death.

SSR_j relates the actual mortality at the jth hospital to the expected mortality in the same hospital, adjusted for different patient severity resumed in the covariates of the logistic model. The estimated number of deaths E_j, can be obtained fitting the following logistic regression model:

$$logit(\mathbb{E}[Y_i]) = logit(p_i) = \beta_0 + \sum_{k=1}^{2} \beta_k x_{ik}. \tag{2}$$

Y_i is the Bernoulli variable representing the in-hospital mortality of patient i and $(\beta_0, \beta_1, \beta_2)$ are the parameters corresponding to the fixed covariates: $\mathbf{x}_i = (x_{1i}, x_{2i}) = (\text{Age}_i, \text{Risk Index}_i)$. We considered also the sex of the patient, but it has been discarded because highly correlated with the age. Since the distribution of age of female is stochastically higher than the males one (non parametric Wilcoxon comparison test: p-value $< 2 * 10^{-16}$), adjusting for sex adds little after standardising for age and risk index. In the selected model age and risk index have a very high statistical significance (p-value $< 2*10^{-16}$). The Risk index is evaluated with the APR-DRG (All Patient Refined Diagnosis Related Groups) mortality score see [6, 19].[1] Finally, as a precision parameter to be displayed on the abscissa we adopted the expected number of deaths of each hospital.

Starting from the first case (a) which doesn't consider any overdispersion due to the grouped nature of data, we draw a funnel plot for Standardized Mortality Ratio reported in Fig. 1. In this case limits are given by

$$\theta_0 \pm z_p \sqrt{\frac{\theta_0}{E}} \tag{3}$$

where θ_0 is the fixed target (in our case θ_0 is set to 1 meaning that the target level is a number of observed deaths which is equal to the expected one), z_p are the quantiles of a standard gaussian distribution, and E is the volume of expected cases.

[1]This is a risk index developed by 3M Health Information Systems and it combines the ICD-9-CM codes of Primary diagnosis, Secondary diagnoses and Procedures. Once the clinical model for severity of illness and risk of mortality was developed for each base APR-DRG, it was evaluated with historical data, extensively reviewed and regularly updated.

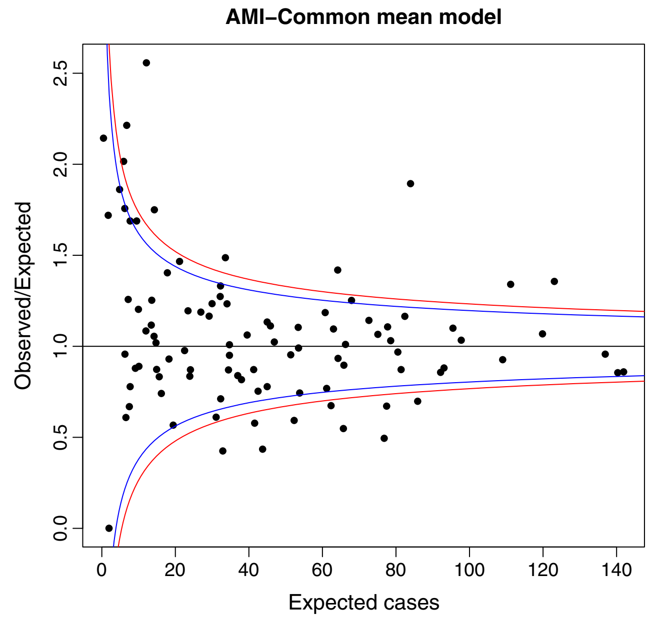

Fig. 1 In-hospital Standardized Mortality Ratio from 96 hospitals in Regione Lombardia. Unadjusted funnel plot with band limits at 95 % (*blue lines*) and 98 % (*red lines*). The *horizontal solid black line* is the target limit

In case (b) an additive adjustment factor is introduced in the band limits of the funnel plot (see Fig. 2). The limits are then given by

$$\theta_0 \pm z_p \sqrt{\frac{\theta_0}{E} + \hat{\tau^2}} \tag{4}$$

where, following [7], if we call $\sigma_j^2 = \frac{\theta_0}{E_j}$, $w_j = 1/\sigma_j^2$ for every hospital $j = 1, ..., J$, $\hat{\tau^2}$ is given by

$$\hat{\tau^2} = \frac{J\hat{\phi}^W - (J-1)}{\sum_j w_j - \sum_j w_j^2 / \sum_j w_j}$$

and $\hat{\phi}^W$ is estimated using the q-Winsorised Z-scores z_j^W

$$\hat{\phi}^W = \frac{1}{J} \sum_{j=1}^{J} \left(z_j^W\right)^2.$$

The Z-scores z_j are obtained by standarizing the observed outcomes of interest, choosing the target θ_0 as the mean, and the estimated variability $\sqrt{\sigma_j^2}$ as standard deviation, i.e.,

$$z_j = \frac{SSR_j - \theta_0}{\sqrt{\text{Var}[X_j]}} = \frac{SSR_j - \theta_0}{\sqrt{\frac{\theta_0}{E_j}}}.$$

The q-Winsorised Z-scores are obtained setting the lowest $100q$ Z-scores to the quantile of order q of $z_1, ..., z_J$ and, analogously the highest $100q$ Z-scores to the quantile of order $(1-q)$ of $z_1, ..., z_J$. If there is no true

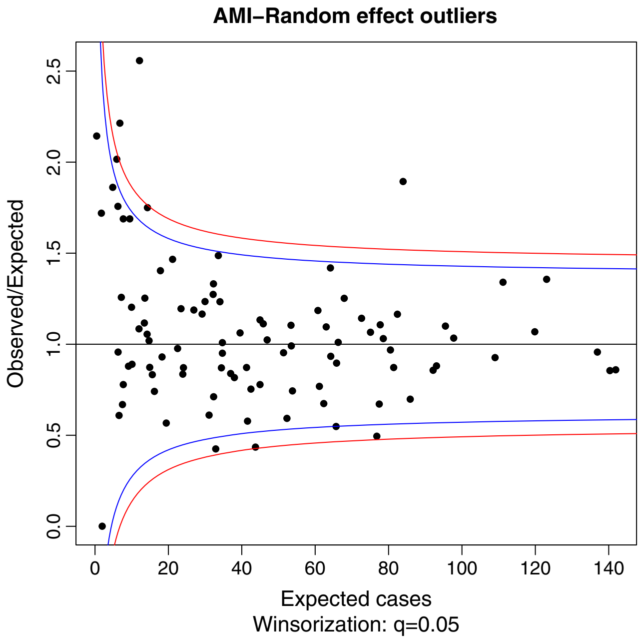

Fig. 2 In-hospital Standardized Mortality Ratio from 96 hospitals in Regione Lombardia. Additive adjusted funnel plot with band limits at 95 % (*blue lines*) and 98 % (*red lines*). The *horizontal solid black line* is the target limit

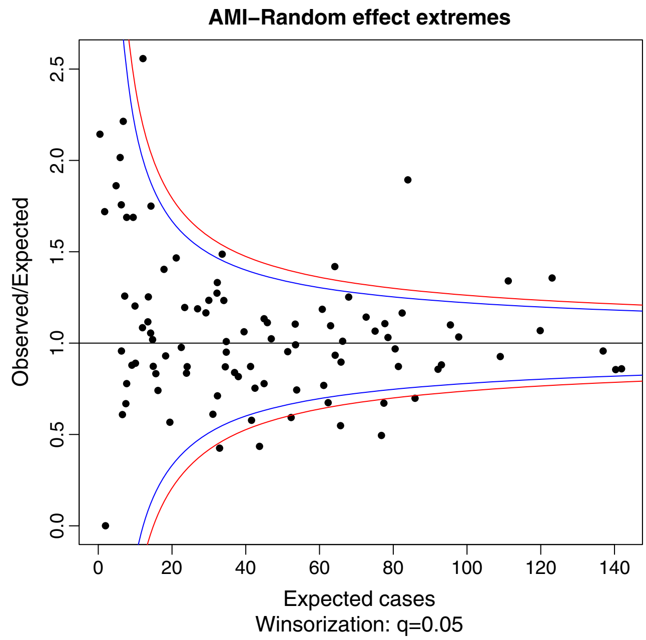

Fig. 3 In-hospital Standardized Mortality Ratio from 96 hospitals in Regione Lombardia. Multiplicative adjusted funnel plot with band limits at 95 % (*blue lines*) and 98 % (*red lines*). The *horizontal solid black line* is the target limit

overdispersion $J\hat{\phi}^W$ has approximately a χ^2 distribution with J degrees of freedom. In this case $\mathbf{E}[\hat{\phi}^W] = 1$ and $\text{Var}[\hat{\phi}^W] = \frac{2}{J}$. So data require a statistically significant adjustment for overdispersion when $\hat{\phi}^W > 1 + 2\sqrt{2/J}$. In our application $\hat{\phi}^W = 2.6$ and this value is strictly greater than $1 + 2\sqrt{2/J} = 1.28$. This fact proves, as suggested in [22], a statistical evidence for overdispersion.

In case (c) a multiplicative adjustment factor is introduced in the band limits of the funnel plot (see Fig. 3). The limits are then given by

$$\theta_0 \pm z_p \sqrt{\frac{1}{\hat{\tau^2}} \frac{\theta_0}{E} \left(\frac{\theta_0}{E} + \hat{\tau^2} \right)}. \tag{5}$$

The plot is reported in Fig. 3. As stressed in [13] since the band limits in (b) are more conservative the detected out of control structures are more likely to be real outliers. For this reason the case (b) is the most appropriate adjustment for the identification of outliers to random effect distribution. In fact there is a slight difference between *outliers* and *extreme cases*. While an extreme observation may be accommodated within the assumed model, and it's just a rare event occurring without questioning the underlying model, an outlier essentially lies beyond the range allowed by the model, and probably it's an event coming from a different model.

The additive correction proposed in Eq. 4 sets the random effect distribution in the model for "in-control" units, so the control limits detect the true outliers observations. Moreover the multiplicative adjustment proposed in Eq. 5 will tend to identify too many providers as unusual.

4 Outlier detection in random intercept estimates

Although funnel plots are very easy to construct and understand, they are unable to quantify the deviation from the incontrol state. That is why we propose to use them jointly with a second statistical method for outliers detection, based on the identification of extremes values of the random effect point estimates arising from a suitable mixed effect model (see Eq. 6) fitted to the data. Again, the hospital of admission is considered as a grouping factor. The study of the random effects distribution for provider profiling has been used both in frequentist [11] and Bayesian context [12] with the main aim of clustering hospitals and detecting similar behaviors. In this paper we take advantage of the study of random effect distribution in order to (i) understand the agreement with the graphical procedures presented in Section 3 and (ii) to point out a method for quantifying the degree of outlyingness and to make inference on it.

So we fit a parametric mixed effect model to explain the in-hospital mortality, with significant covariates selected in model (2) as fixed effects and a random intercept. In particular

$$logit(\mathbb{E}[Y_{ij}|b_j]) = logit(p_{ij}) = \beta_0 + \sum_{k=1}^{2} \beta_k x_{ijk} + b_j. \quad (6)$$

Y_{ij} is the Bernoulli variable representing the in-hospital mortality of patient i treated in hospital j. $(\beta_0, \beta_1, \beta_2)$ is the parameters vector corresponding to the fixed part. $b_j \sim \mathcal{N}\left(0, \sigma_b^2\right)$ is the Normal random effect superimposed to the grouping factor (i.e., the hospital where a patient is admitted to).

In our application the point estimates of parameters (± standard deviation) are $\hat{\beta}_0 = -7.8108 \pm 0.1379$, $\hat{\beta}_1 = 0.0408 \pm 0.0017$ and $\hat{\beta}_2 = 1.1257 \pm 0.0214$, respectively.

The estimate of the standard deviation of the random effect is equal to $\hat{\sigma_b} = 0.37$.

The points over the limits of the whiskers in the boxplot (see Fig. 4) of the point estimates of the random intercept in model (6) are a subset of the hospitals which locate out of the 95 % limits of all three funnel plots in Eqs. 3–5. In Fig. 4 we draw also the quantiles of order 2.5 % and 97.5 % (blue lines) and the quantiles of order 1 % and 99 % (red lines), in order to highlight the agreement with Fig. 2. In Fig. 5 the Normal Q-Q plot of the point estimates of the random intercept in model (6) is shown. In fact the Normality assumption on the random effect b_j in model (6) is questioned and weakened by the detected outliers in Fig. 4). Moreover the detected outlier hospitals are

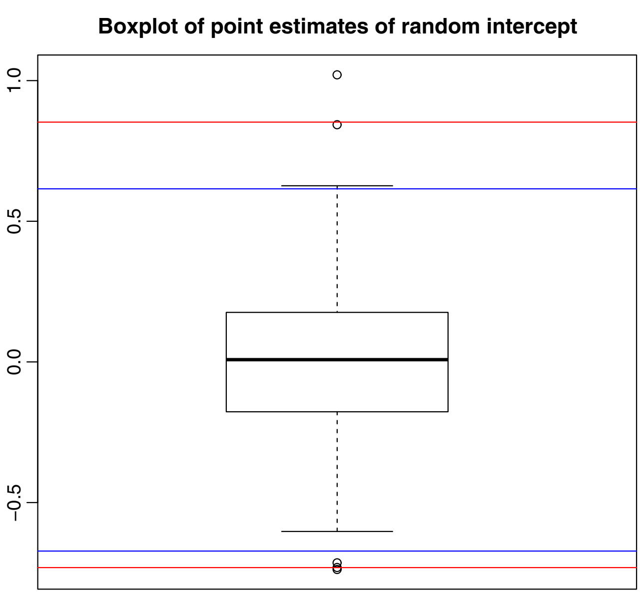

Fig. 4 *Boxplot* of the point estimates of the random intercept in model (6). The *horizontal blue lines* correspond to the quantiles of order 2.5 % and 97.5 %. The *horizontal red lines* correspond to the quantiles of order 1 % and 99 %

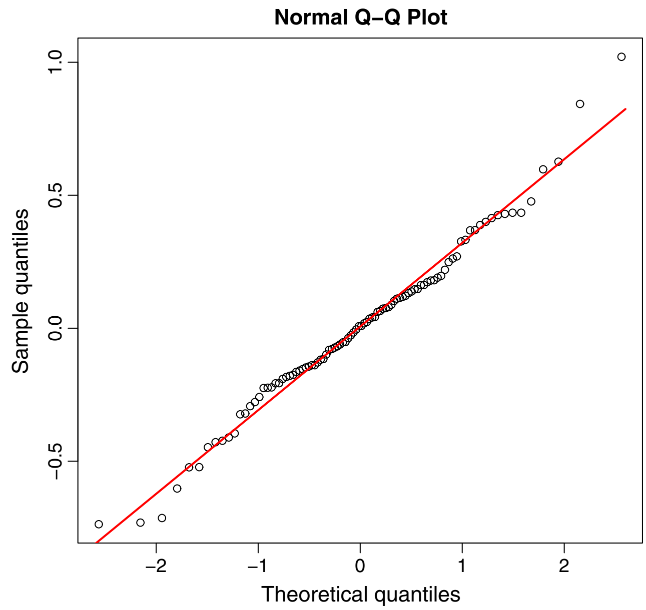

Fig. 5 Normal Q-Q plot of the point estimates of the random intercept in model (6)

the ones located out of the limits in the funnel plot shown in Fig. 2 that, as said before is a suggested tool to point out unusual performance.

Of course, the monitoring of in-hospital mortality rate is just an example of regulatory application that Funnel Plot and GLMM can be jointly used for. Another outcome,

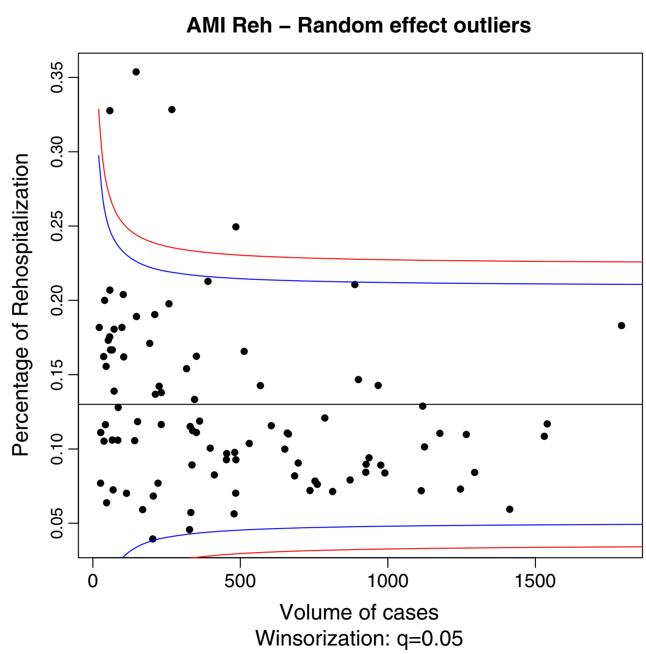

Fig. 6 Rehospitalization Rate from 96 hospitals in Regione Lombardia. Additive adjusted funnel plot with band limits at 95 % (*blue lines*) and 98 % (*red lines*). The *horizontal solid black line* is the target limit

arising from the same context and used as indicator for monitoring hospitals practice is the rate of readmissions. On the same data we observed a mean rate equal to 0.13. We set the "in-control" mean parameter θ_0 at this value. We considered the additive adjustment to draw the confi-dence limits of the funnel plot (see Fig. 6). The detected outliers are hospitals apparently under-performing in terms of this outcome. We fitted also a generalized random effect model like Eq. 6. Y_{ij} is the Bernoulli variable representing the in-hospital readmission of patient i treated in hospital j. $(\beta_0, \beta_1, \beta_2)$ is the parameters vector corresponding to the fixed part where the significant covariates are Age and Risk index. $b_j \sim \mathcal{N}\left(0, \sigma_b^2\right)$ is the Normal random effect super-imposed to the grouping factor (i.e., the hospital where a patient is admitted to). The estimate of the standard deviation of the random effect is equal to $\hat{\sigma_b} = 0.43$. The outlier detected by the study of the random effect point estimates are in agreement with the ones depicted in the funnel plot in Fig. 6.

5 Conclusions

The identification of unusual performance of healthcare institutions is a hot topic of the last years. It is crucial, especially in supporting decisions of people in charge with healthcare governance. We propose the joint use of graphical instruments (like funnel plots) and random effect models to detect, visualize and study out-performing institutions.

In fact the use of hierarchical model allows for a more proper modelling of the overdisperion highlighted by the funnel plots. Moreover, they enable researchers to quantify the outlyingness of a performance, providing a quantitative support to decision makers. In fact, the funnel plots indicate the presence of overdispersion in data, and by computing suitable confidence limits they allow to detect unusual institutions. On the other hand, the study of tails and extremes in estimated random effect distribution confirm the obtained results and provide also an estimate of the quantitave effect of these structures in outcome prediction.

The joint use of these two techniques is a simple and effective way for profiling providers within the context of healthcare regulation. It enables real-time notification of out-of-control performances to be further investigated.

Acknowledgments This work is within the Strategic Program "Exploitation, integration and study of current and future health databases in Lombardia for Acute Myocardial Infarction". The authors wish to thank Regione Lombardia - Healthcare division for having funded and sustained the project, Lombardia Informatica S.p.A. for having provided data and all the physicians who collaborated to STEMI Archive planning and data collection.

References

1. AHRQ QualityIndicators (2007) Guide to inpatient quality indicators: quality of care in hospitals - volume, mortality, and utilization. Department of Health and Human Services Agency for Healthcare Research and Quality, Version 3.1
2. Austin PC, Naylor CD, Tu JV (2001) A comparison of a Bayesian vs. a frequentist method profiling hospital performance. J Eval Clin Pract 7:35–45
3. Austin PC, Alter DA, Tu JV (2003) The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. Med Dec Making 23:526–539
4. Barbieri P, Grieco N, Ieva F, Paganoni AM, Secchi P (2010) Exploitation, integration and statistical analysis of Public Health Database and STEMI archive in Lombardia Region. Complex data modeling and computationally intensive statistical methods Series "Contribution to Statistics". Springer, pp 41–56
5. Campbell MJ, Jacques RM, Fotheringham J, Maheswaran R, Nicholl J (2012) Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. Br Med J 344:e1001
6. De Marco MF, Lorenzoni L, Addari P, Nante N (2002) Evaluation of the capacity of the APR-DRG classification system to predict hospital mortality. Epidemiol Prev 26(4):183–190
7. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7:177–188
8. Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. Br Med J 315(7109):629–624
9. Farrel PJ, Groshen S, MacGibbon B, Tomberlin TJ (2010) Outlier detection for a hierarchical Bayes model in a study of hospital variation in surgical procedures. Stat Methods Med Res 19:601–619
10. Garbe S, Primer E (2007) Administrative health databases in observational studies of drug effects-advantages and disadvantages. Nat Clin Pract Rheumatol 3:725–732 [review]
11. Grieco N, Ieva F, Paganoni AM (2012) Performance assessment using mixed effects models: a case study on coronary patient care. IMA J Manag Math 23(2):117–131
12. Guglielmi A, Ieva F, Paganoni AM, Ruggeri F, Soriano J (2013) Semiparametric Bayesian modeling for the classification of patients with high observed survival probabilities. J R Stat Soc-Ser C. (in press)
13. Jones HE, Spiegelhalter DJ (2011) The identification of "unusual" health-care providers from a hierarchical model. Am Stat 65(3):154–163
14. Light RJ, Pillemer DB (1984) Summing up: the science of reviewing research. Harvard University Press, Cambridge
15. Mohammed MA, Deeks JJ (2008) In the context of performance monitoring, the caterpillar plot should be mothballed in favor of the funnel plot. Ann Thorac Surg 86:348
16. Normand SLT, Glickman ME, Gatsonis CA (1997) Statistical methods for profiling providers of medical care: issues and applications. J Am Stat Assoc 92:803–814
17. R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna [Online]. http://www.R-project.org
18. Racz J, Sedransk J (2010) Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. J Am Stat Assoc 105(489):48–58
19. Romani PS, Chan BK (2000) Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? Health Serv Res 34(7):1469–1489
20. Ministero della Sanità (1991) Decreto 28 Dicembre, [online] http://www.trovanorme.salute.gov.it/dettaglioAtto?id=12967&articolo=3

21. Spiegelhalter DJ (2002) Funnel plots for institutional comparisons (letter). Qual Saf Health Care 11:390–391
22. Spiegelhalter DJ (2005) Funnel plots for comparing institutional performance. Stat Med 24:1185–1202
23. Spiegelhalter DJ, Sherlw-Johnson C, Bardsley M, Blunt I, Wood C, Grigg O (2012) Statistical methods for healthcare regulation: rating, screening and surveillance. J R Stat Soc - Ser A 175:1–47
24. Wirehn AB, Karlsson HM, Cartensen JM et al (2007) Estimating Disease Prevalence using a population-based administrative healthcare database. Scand J Public Health 35:424–431
25. Van Walraven C, Austin P (2012) Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol 65:126–131