

Process indicators and outcome measures in the treatment of acute myocardial infarction patients

Alessandra Guglielmi¹, Francesca Ieva¹,
Anna Maria Paganoni¹ and Fabrizio Ruggeri²

¹*Department of Mathematics, Politecnico di Milano, Milan, Italy*

²*CNR IMATI, Milan, Italy*

Synopsis

Studies of variations in healthcare utilization and outcome involve the analysis of multilevel, clustered data, considering in particular the estimation of a cluster-specific adjusted response, covariate effects and components of variance. Besides reporting on the extent of observed variations, these studies quantify the role of contributing factors including patients' and providers' characteristics. In addition, they may assess the relationship between healthcare process and outcomes. We consider Bayesian generalized linear mixed models to analyze MOMI² (Month MOnitoring Myocardial Infarction in Milan) data on patients admitted with ST-elevation myocardial infarction (STEMI) diagnosis in the hospitals belonging to the Milano Cardiological Network. Both clinical registries and administrative databanks were used to predict survival probabilities. We fit a logit model for the survival probability with one random effect (the hospital), under a semiparametric prior. We take advantage of the in-built clustering property of the Dirichlet process prior assumed for the random-effects parameters to obtain a classification of providers.

10.1 Introduction

Performance indicators for assessing quality in healthcare research have drawn more and more attention over recent years, since they can evaluate some aspects of the healthcare process, clinical outcomes and disease incidence. At the same time, questions about the right use of such indicators as a measure of quality of care have emerged.

Several examples, available in clinical literature (see, for instance, Hasday, Behar and Wallentin, 2002, and Saia, Marzocchi and Manari, 2009), make use of clinical registries to evaluate performance of medical institutions, because they enable people concerned with the healthcare governance to plan activities on real epidemiological evidence and needs; moreover clinical registries help in evaluating performance of structures they manage, providing knowledge about the number of cases, incidence, prevalence and survival concerning a specific disease.

In this work, clinical registries are used to model in-hospital survival of acute myocardial infarction patients, in order to classify providers' performances and to enable healthcare governance to better manage resources.

The disease we are interested in is the ST-segment Elevation acute Myocardial Infarction (STEMI): it consists of a stenotic plaque detachment, which causes a coronary thrombosis and a sudden critical reduction of blood flow in coronary vessels, leading to an inadequate feeding of myocardial muscle itself. STEMI is characterized by a very high incidence (650–700 events per month have been estimated in just the Lombardia Region, the inhabitants of which number approximately 10 million) and serious mortality (in Italy, about 8% in adults aged between 35 and 74 years). A case of STEMI is usually treated by percutaneous transluminal coronary angioplasty (PTCA): an empty and collapsed *balloon* on a guide wire, known as a balloon catheter, is passed into the narrowed or obstructed vessels and then inflated to a fixed size.

The balloon crushes the fatty deposit, so opening up the blood vessel to improved flow, and is then collapsed and withdrawn. Good results for the treatment can be evaluated for instance by observing, first, the in-hospital survival of inpatients.

For heart attacks, survival strongly depends on time saved during the process and, therefore, in this work we focus on the survival outcome. In any case, time indeed has a fundamental role in the overall STEMI healthcare process. By *Symptom Onset to Door time* we mean the time since symptom onset up to the arrival at the Emergency Room (ER); and *Door to Balloon time* (DB time) is the time from the arrival at the ER up to the surgical practice of PTCA. Clinical literature strongly stresses the connection between in-hospital survival and procedure time (Cannon, Gibson and Lambrew, 2000; Jneid, Fonarow and Cannon, 2008; MacNamara *et al.*, 2006): 90 minutes for DB time in the case of primary PTCA (i.e. PTCA without any previous pharmacological treatment) is the actual gold standard limit suggested by the American Heart Association (AHA)/American College of Cardiology (ACC) guidelines; see Antman, Hand and Armstrong, 2008.

The presence of differences in the outcomes of healthcare has been documented extensively in recent years. In order to design regulatory interventions by institutions, for instance, it is interesting to study the effects of variations in healthcare utilization on patients' outcomes, in particular examining the relationship between process indicators, which define regional or hospital practice patterns, and outcome measures, such as patients' survival or a treatment's efficacy.

Analyses of variations concerning the comparison of the performance of healthcare providers are commonly referred to as *provider profiling* (Normand, Glickman and Gatsonis, 1997; Racz and Sedransk, 2010).

The results of profiling analyses often have far-reaching implications. They are used to generate feedback for healthcare providers, to design educational and regulatory interventions by institutions and government agencies, to design marketing campaigns by hospitals and managed-care organizations, and, ultimately, used by individuals and managed-care groups to select healthcare providers.

The aim of this work is twofold: on one hand we want to quantify the magnitude of the variations in healthcare providers and to assess the role of contributing factors, including patients' and providers' characteristics, on survival outcome. Data on healthcare utilization have a 'natural' multilevel structure, usually with patients at the lower level and hospitals forming the upper-level clusters. Within this formulation, the main goal is to derive estimates of providers' effects; that is, differences between hospitals. On the other hand, we want to cluster hospitals according to their performance in patients' care.

Hierarchical regression modelling from a Bayesian non-parametric perspective provides a framework that can accomplish both these goals.

Here, this article considers a Bayesian generalized linear mixed model (Zeger and Karim, 1991) to predict the binary survival outcome by means of relevant covariates, taking into account overdispersion induced by the grouping factor, and modelling the random effects non-parametrically.

In particular, as in Kleinman and Ibrahim (1998), the random-effects parameters are a sample from a Dirichlet process prior (Ferguson, 1973), which provides a natural setting for the classification of hospitals thanks to the discreteness property of its trajectories. We illustrate the analysis on data coming from a survey on patients admitted with STEMI diagnosis in one of the structures belonging to the Milano Cardiological Network, using a logit model for the survival probability and a Dirichlet process for the distribution of the random effect.

For this analysis, patients are grouped by the hospital they have been admitted to for their infarction. A Markov chain Monte Carlo (MCMC) algorithm is necessary to compute the posterior distributions of parameters and predictive distributions of outcomes. The choice of covariates and link functions was suggested first in Ieva and Paganoni (2011), according to frequentist selection procedures and clinical know-how, and was confirmed in Guglielmi *et al.* (2012) using Bayesian tools.

Concerning modelling of provider's variability, we take advantage of in-built clustering provided by the Dirichlet process, jointly with the partitioning around medoids (PAM) algorithm (see Kaufman and Rousseeuw, 1987) to obtain classification of providers and estimation of their effects on survival outcome adjusted for case mix.

The advantages of a Bayesian non-parametric approach to this problem are more than one: the providers' profiling or patients' classification can be guided not only by statistical but also by clinical knowledge, hospitals with low exposure can be automatically included in the analysis, providers' profiling can be simply achieved through the posterior distribution of the hospital-effects parameters and in-built clustering is naturally provided by the non-parametric setting. In the following sections, dataset, performed analyses and results are discussed and future work is presented.

All the analyses have been performed with the R (version 2.10.1, R Development Core Team, 2009) program.

10.2 A semiparametric Bayesian generalized linear mixed model

We fit a generalized mixed-effects model for binary data from a Bayesian viewpoint. For patient $i = 1, \dots, n_j$ in each hospital $j = 1, \dots, J$, let Y_{ij} be a Bernoulli random variable with mean p_{ij} , which represents the probability that the patient survived after STEMI.

The p_{ijs} are modelled through a logit regression with covariates $\mathbf{x} := \mathbf{x}_{ij}$, $\mathbf{x}_{ij} \in \mathbb{R}^p$; that is,

$$Y_{ij} | p_{ij} \stackrel{\text{ind}}{\sim} \text{Be}(p_{ij}), \quad j = 1, \dots, J, \quad i = 1, \dots, n_j \quad (10.1)$$

and

$$\text{logit}(p_{ij}) = \log \frac{p_{ij}}{1 - p_{ij}} = \beta_0 + \sum_{h=1}^p \beta_h x_{ijh} + \sum_{l=1}^J b_l z_{il}, \quad (10.2)$$

where b_j represents the j th hospital effect, and $z_{ij} = 1$, $z_{il} = 0$ for $l \neq j$.

We will denote by $\boldsymbol{\beta}$ the vector of regression parameters $(\beta_0, \beta_1, \dots, \beta_p)$, which are called fixed effects, while $\mathbf{b} = (b_1, \dots, b_J)$ are the random-effects parameters.

Note that (10.1)–(10.2) is a generalized linear mixed model with $(p + 1)$ regression coefficients and one random effect (the random intercept taking into account the grouping structure of hospitals).

Traditionally, $\boldsymbol{\beta}$ and \mathbf{b} are assumed *a priori* independent, $\boldsymbol{\beta}$ is Gaussian distributed, and the random variables b_j , conditionally on σ^2 , are independent identically Gaussian distributed, with random variance σ^2 . Here, according to Kleinman and Ibrahim (1998), we assume a nonparametric prior for \mathbf{b} , namely the b_j s will be i.i.d. as a Dirichlet process, to include robustness to misspecification of the prior at this stage, since it is known that the regression parameters can be sensitive to the assumption of normality about the random effects.

Generally, nonparametric Bayesian models are assumed to avoid critical dependence on parametric assumptions, to robustify parametric models, or to perform sensitivity analysis for parametric models by embedding them in a larger encompassing nonparametric model. Priors under a nonparametric Bayesian perspective consist in probabilities on probability spaces: instead of considering models that can be indexed by a finite-dimensional parameter, we consider a prior probability q for the unknown population distribution G , which, in the case considered here, represents the probability distribution of the random-effect parameter b_j . In particular, we will assume that q is a Dirichlet prior, or, equivalently, that G is a Dirichlet process. For a more formal definition of the Dirichlet process and a review of Bayesian nonparametric inference, see Müller and Quintana (2004). Here we would like to mention two properties only.

First, the Dirichlet process is indexed by two ‘parameters’, a positive parameter α and a distribution G_0 on \mathbb{R} . This latter represents the mean trajectory of G ; that is, $E(G(A)) = G_0(A)$ for each measurable subset A of \mathbb{R} ; in this way, the parametric model $G_0(\vartheta)$ could be embedded in a larger encompassing nonparametric model G . The parameter α is a precision parameter that defines variance: when α increases, the prior q will concentrate more and more mass on the mean distribution G_0 .

The second property we refer to is the discreteness of the trajectories of G : realizations of the Dirichlet process are infinite mixtures of point masses.

This feature of G will provide a natural setting for the classification of the hospitals. In fact, if, conditionally on G , b_1, \dots, b_j are i.i.d. according to G , there is a positive probability, depending on α and the sample size J , of having coincident values among the b_j s, since the next observation b_j in the sample will be a new sampled value from G_0 with probability $\alpha/(\alpha + j - 1)$, or one of the previously sampled $(b_1, b_2, \dots, b_{j-1})$, with probability $1/(\alpha + j - 1)$; that is,

$$b_j \sim \frac{\alpha}{\alpha + j - 1} G_0 + \frac{1}{\alpha + j - 1} \sum_{i=1}^{j-1} \delta_{b_i}. \tag{10.1}$$

The joint posterior distribution of the random effects will preserve a similar clustering structure as well, so that this will prompt a natural classification among random effects, and consequently among hospitals (*a posteriori*). See the next section.

With more details, the prior we assume is

$$\begin{aligned} \beta \perp \mathbf{b}, \beta &\sim N_{p+1}(0, \Sigma_0) \\ b_1, \dots, b_j \mid G &\sim iid G, \\ G &\sim Dir(\alpha N(\mu, \sigma^2)), \quad \mu \sim N(\mu_b, S_b), \quad \frac{1}{\sigma^2} \sim gamma\left(\frac{\nu_0}{2}, \frac{t_0^{-1}}{2}\right). \end{aligned} \tag{10.3}$$

Integrating out G , this prior yields a prior marginal for b_j which, conditionally on (μ, σ^2) , is $N(\mu, \sigma^2)$, but hyperparameters are not fixed, and, to decrease sensitivity of the inferences, we assume them random, as in (10.3).

Moreover, as in the parametric case (Guglielmi *et al.*, 2012), the random-effects parameters are assumed dependent (this is a sensible assumption), and we will be able to use the whole dataset to make inferences on hospitals which have few or no patients in the study, borrowing strength across hospitals.

Of course, model (10.1)–(10.2) under prior (10.3) cannot be fit without resorting to an MCMC scheme to compute the joint posterior distribution of all parameters, which will be used to compute the Bayesian estimates of interest.

The joint posterior of the random-effect parameters can be expressed via the full conditionals of a Gibbs sampler algorithm: a ‘new’ value for b_j , given the data, the other random-effects parameters $\mathbf{b}_{-j} = (b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_J)$, and all the ‘rest’, is sampled either from a Gaussian distribution with some probability, or it is equal to one of the component b_i s of the vector \mathbf{b}_{-j} (with appropriate weight).

See Kleinman and Ibrahim (1998) for the expressions of the full conditionals of a Gibbs sampler algorithm in this case.

10.3 Hospitals’ clustering

As we mentioned before, there will be coincident values among the MCMC-sampled random-effects parameters \mathbf{b} .

Here we propose how to use the sample’s bias of the posterior distribution of \mathbf{b} , in order to detect a clustering structure between hospitals in affecting in-hospital survival.

Let us denote by $\mathbf{D}(b_1, \dots, b_J)$ a $J \times J$ symmetric matrix such that the (i, j) th element $[\mathbf{D}(b_1, \dots, b_J)]_{ij} = 1$ if $b_i \neq b_j$ and 0 otherwise.

Then we compute the matrix D of the posterior means of $[D(b_1, \dots, b_J)]_{ij}$; in short,

$$D = E[D(b_1, \dots, b_J) | y]. \quad (10.4)$$

It is easy to prove that D is a pseudo-metric which represents a mean dissimilarity measure between hospitals and can be computed via the MCMC samples of (b_1, \dots, b_J) ; a PAM algorithm is then applied to hospitals. A PAM algorithm is based on the search for k representative objects, called *medoids*, among objects of the dataset (in our case hospitals).

These medoids are computed such that the total dissimilarity of all objects to their nearest medoid is minimal. In this case our goal is to find a subset $\{m_1, \dots, m_k\} \subset \{1, \dots, J\}$ which minimizes the objective function

$$\sum_{j=1}^J \min_{t=1, \dots, k} D_{jm_t}$$

where D_{ij} is the (i, j) th element of the matrix D .

A critical point is the choice of k , the number of groups: a helpful method is the computation of the average silhouette width, and the inspection of the silhouette plot of PAM. For each hospital j , we denote by A the cluster to which it belongs and compute $a(j)$, the average dissimilarity of j to all other objects of A :

$$a(j) = \frac{1}{|A| - 1} \sum_{i \in A, i \neq j} D_{ij}.$$

Now let us consider any cluster C different from A and denote by

$$d(j, C) = \frac{1}{|C| - 1} \sum_{i \in C} D_{ij}$$

the average dissimilarity of j to all objects of C ; we define $c(j)$ as the smallest value of all $d(j, C)$ for all clusters C different from A . The *silhouette value* $s(j)$ of an object j is defined as:

$$s(j) = \frac{c(j) - a(j)}{\max\{a(j), c(j)\}}.$$

The silhouette value $s(j)$ of each object j , the entire silhouette plot, that is, the plot of all $s(j)$, and the average of all silhouette values are qualitative indexes to judge and compare the results of different PAM procedures (see Struyf, Hubert and Rousseeuw, 1997 for more details).

10.4 Applications to AMI patients

The dataset we are interested in is about patients admitted with STEMI diagnosis in one of the hospitals belonging to the Milano Cardiological Network.

For these units, information concerning mode of admission (on his/her own or by three different types of 118 rescue units), demographic features (sex, age), clinical appearance

(presenting symptoms and Killip class at admittance), Symptom Onset to Door time, in-hospital times (first ECG time, DB time), hospital organization (for example, admission during on/off hours) and clinical outcome (in-hospital survival) have been collected.

The Killip classification is a system used in individuals with an acute myocardial infarction, in order to risk stratify them into four severity classes. Individuals with a low Killip class are less likely to die within the first 30 days after their myocardial infarction than individuals with a high Killip class.

Previous frequentist and Bayesian analyses (for further details see Ieva and Paganoni, 2011; Guglielmi *et al.*, 2012) pointed out that age, total ischemic time (Symptom Onset to Balloon time, denoted by OB) in the logarithmic scale and Killip of the patient, categorized as a binary variable, corresponding to 0 for less severe (Killip class equal to 1 or 2) and 1 for more severe (Killip class equal to 3 or 4) infarction, are the most significant factors in order to explain survival probability from a statistical and clinical point of view.

There are $n = n_1 + \dots + n_J = 240$ patients, in $J = 17$ hospitals in the dataset; the number of patients per hospital ranges from 1 to 32, with a mean of 14.12. Each observation $y_i = 1$ if the i th patient survived; $y_i = 0$ otherwise.

In this study we fitted model (10.1)–(10.2) with $p = 3$, under (10.3), with the help of an R package called *DPpackage* (Jara, 2007).

In particular, we ran the function *DPglm*, which adopts a slightly different parameterization from (10.3); however it is only the prior of β_0 which changes (Jara *et al.*, 2011).

After some preliminary robustness analysis, the prior was fixed so that $\beta_1, \beta_2, \beta_3$ are i.i.d. according to $N(0,100)$, $\mu_b = 0$, $S_b = 100$, $\nu_0 = 5$, $t_0^{-1} = 30$, $a_0 = b_0 = 1$. We assumed such values since, in this case, the prior expected number of distinct values among the b_j s is 3 (which seems a sensible choice), however letting α be not too informative, while the prior expectation and variance of the conditional variance parameter Σ are 10 and 200, respectively, so that the marginal prior variance of each b_j is 110. Moreover, the robustness analysis showed that the inferences are not sensitive to different choices of μ_b and Σ_0 .

Summary inferences about regression parameters can be found in Table 10.1.

A look at the posterior distributions of β_i , for $i = 1, 2, 3$ (not included here for brevity) shows that Killip and age have a negative effect on the survival probability, while $\log(\text{OB})$ has a lighter influence on it.

Summary inferences about random-effects parameters can be found in Table 10.2. Their posterior means range from 3.058 to 4.783.

The marginal posterior densities of all the random-effects parameters, clustered in $k = 3$ groups, are depicted in Figure 10.1; while in Figure 10.2 the corresponding dissimilarity matrix D is showed.

Table 10.1 Posterior means and standard deviations of the fixed-effects regression parameters.

	Mean	sd	
Age	β_1	-0.0804	0.0339
$\log(\text{OB})$	β_2	-0.1758	0.3733
Killip	β_3	-1.6979	0.8747

Table 10.2 Posterior means and standard deviations of the random-effect regression parameters.

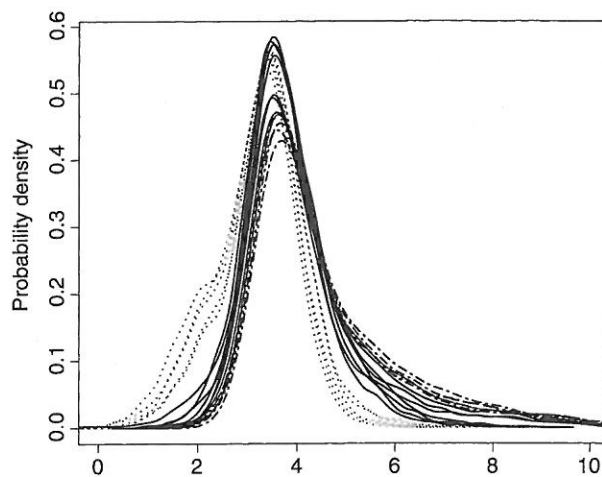
$\beta_0 + b_j$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$	$j = 9$
Mean	4.5472	4.1969	4.4122	4.6339	3.7119	3.9503	3.8226	4.0220	4.7827
sd	1.7269	1.5591	1.6909	1.7698	0.8452	0.9141	0.8585	1.5808	1.8754
$\beta_0 + b_j$	$j = 10$	$j = 11$	$j = 12$	$j = 13$	$j = 14$	$j = 15$	$j = 16$	$j = 17$	
Mean	3.1948	3.0584	4.4685	3.3995	3.9017	3.1969	3.3625	4.5776	
sd	0.8174	0.8824	1.6423	0.8809	0.8816	0.8735	0.8854	1.7906	

The PAM algorithm assigns hospitals 1, 4, 9, 12, 17 to the first group, hospitals 2, 3, 5, 6, 7, 8, 14 to the second and hospitals 10, 11, 13, 15, 16 to the third group.

According to the values of the posterior means of the b_j s, the related medoids (hospitals 4, 14 and 10) represent 'good', 'medium' and 'poor' performances, respectively. This classification is also in agreement with results in Guglielmi *et al.* (2012).

The clustering structure in three groups has been selected inspecting the boxplots of the dissimilarity between hospitals and the medoid of the cluster they belong to, obtained for different values of the number k of clusters, and evaluating average silhouette widths. From this inspection, the presence of either $k = 2$ or $k = 3$ clusters can be supported; however we decided to propose a three-group clustering structure which distinguishes strongly good, or strongly poor hospitals from the medium ones.

In Table 10.3 the estimated in-hospital survival probabilities for different case-mixes, in 'poor', 'medium' and 'good' medoids are shown.

**Figure 10.1** Posterior distributions of random-effects parameters $\beta_0 + b_j$, clustered in three groups: 'poor' (dotted), 'medium' (solid) and 'good' (dashed) hospitals.

egression

$j = 8$	$j = 9$
1.0220	4.7827
1.5808	1.8754
$i = 17$	
1.5776	
1.7906	

ospitals 2, 3, 5, 6,

medoids (hospi-
spectively. This

boxplots of the
to, obtained for
the widths. From
ted; however we
strongly good, or

it case-mixes, in

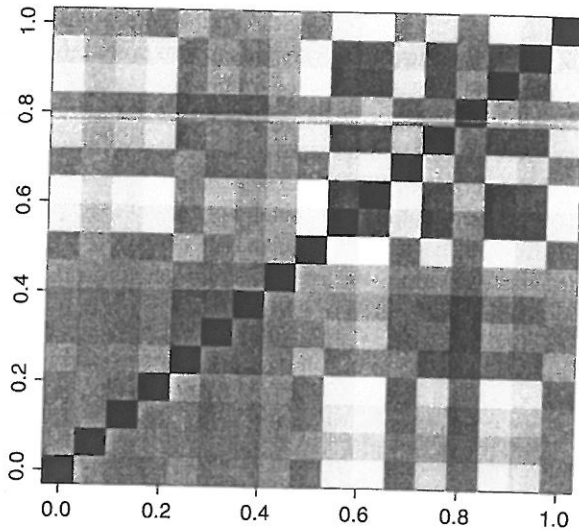


Figure 10.2 Dissimilarity matrix between hospitals, computed from (10.4).

In particular, when moving from a 'poor' hospital to a 'good' one, the in-hospital survival probability gain ranges from 1.5%, for a young patient with less severe infarction, to 32% in the case of an old patient with more severe infarction.

10.5 Summary

One of the major aims of this work is to measure the magnitude of the variations of healthcare providers and to assess the role of contributing factors, including patients' and providers' characteristics, on survival outcome.

Concerning patient features, we found out that Killip and age have a sharp negative effect on the survival probability, while the Symptom Onset to Balloon time has a lighter influence on it.

Moreover, the PAM algorithm applied to the posterior distributions of the hospitals' random effect enabled us to identify three clusters of providers: a group (hospitals 1, 4, 9, 12, 17) performing better than a medium group (hospitals 2, 3, 5, 6, 7, 8, 14) and a third group

Table 10.3 Estimated in-hospital survival probabilities for different case-mix, with average OB time 553 minutes in 'poor', 'medium' and 'good' medoids.

Patient		Estimated survival probability		
Age	Killip	Poor	Medium	Good
55	0	0.9805	0.9903	0.9953
85	0	0.8185	0.9014	0.9501
55	1	0.9021	0.9492	0.9749
85	1	0.4522	0.6260	0.7768

clustered in three

(hospitals 10, 11, 13, 15, 16) performing worse than the central one. Finally we estimated the effect of medoids for each group on in-hospital survival probability, to quantify loss/gain on survival due to provider's behaviour. This could be considered by healthcare governance as an instrument supporting healthcare decisions on optimizing network resources. Since the joint use of clinical registries and administrative databases proposed in this and previous analyses (see Ieva and Paganoni, 2010 and Barbieri, Grieco and Ieva, 2010) produced such useful results, a wider and more complete clinical register on STEMI, extended to the whole territory of Lombardia Region, has been planned and activated in 2010, called STEMI Archive.

As a future work we will apply the methodology and the models illustrated here to this larger dataset, to enable healthcare governance to establish benchmarks and to evaluate hospital network performances, then to offer better services to healthcare users.

References

- Antman, E.M., Hand, M., Armstrong, P.W. *et al.* (2008) Update of the ACC/AHA 2004 guidelines for the management of patients with ST elevation myocardial infarction. *Circulation*, **117**, 269–329.
- Barbieri, P., Grieco, N., Ieva, F. *et al.* (2010) Exploitation, integration and statistical analysis of Public Health Database and STEMI Archive in the Lombardia region, in *Complex Data Modeling and Computationally Intensive Statistical Methods* (eds P. Mantovan and P. Secchi), Contributions to Statistics, Springer, pp. 41–56.
- Cannon, C.P., Gibson, C.M., Lambrew, C.T. *et al.* (2000) Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. *Journal of American Medical Association*, **283**(22), 2941–2947.
- Ferguson, T.S. (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.
- Guglielmi, A., Ieva, F., Paganoni, A.M. and Ruggeri, F. (2012) A Bayesian random-effects model for survival probabilities after acute myocardial infarction. *Chilean Journal of Statistics*, **3**, 15–29.
- Hasday, D., Behar, S., Wallentin, L. *et al.* (2002) A prospective survey of the characteristics, treatments and outcomes of patients with acute coronary syndromes in Europe and the Mediterranean basin. The Euro heart survey of acute coronary syndromes. *European Heart Journal*, **23**, 1190–1210.
- Ieva, F. and Paganoni, A.M. (2010) Multilevel models for clinical registers concerning STEMI patients in a complex urban reality: a statistical analysis of MOMI² survey. *Communications in Applied and Industrial Mathematics*, **1**(1), 128–147.
- Ieva, F. and Paganoni, A.M. (2011) Process indicators for assessing quality of hospitals care: a case study on STEMI patients. *JP Journal of Biostatistics*, **6**(1), 53–75.
- Jara, A. (2007) Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews*, **7**, 17–26.
- Jara, A., Hanson, T., Quintana, F. *et al.* (2011) DPpackage: Bayesian non- and semi-parametric modelling in R. *Journal of Statistical Software*, **40**(5), 1–30.
- Jneid, H., Fonarow, G., Cannon, C. *et al.* (2008) Impact of time of presentation on the care and outcomes of acute myocardial infarction. *Circulation*, **117**, 2502–2509.
- Kaufman, L. and Rousseeuw, P.J. (1987) *Finding Groups in Data*, John Wiley & Sons, Inc., New York.
- Kleinman, K.P. and Ibrahim, J.G. (1998) A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine*, **17**, 2579–2596.
- MacNamara, R.L., Wang, Y., Herrin, J. *et al.* (2006) Effect of door to balloon time on mortality in patients with ST-segment elevation myocardial infarction. *Journal of American College of Cardiology*, **47**, 2180–2186.

- Müller, P. and Quintana, F.A. (2004) Nonparametric Bayesian data analysis. *Statistical Science*, **19**, 95–110.
- Normand, S.T., Glickman, M.E. and Gatsonis, C.A. (1997) Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association*, **92**, 803–814.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Racz, J. and Sedransk, J. (2010) Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. *Journal of the American Statistical Association*, **105**(489), 48–58.
- Saia, F., Marzocchi, A., Manari, G. *et al.* (2009) Patient selection to enhance the long-term benefit of first generation drug-eluting stents for coronary revascularization procedures: insights from a large multicenter registry. *Eurointervention*, **5**(1), 57–66.
- Struyf, A., Hubert, M. and Rousseeuw, P.J. (1997) Clustering in an object-oriented environment. *Journal of Statistical Software*, **1**(4), 1–30.
- Zeger, S.L. and Karim, M.R. (1991) Generalized linear models with random effects: a Gibbs Sampling approach. *Journal of the American Statistical Association*, **86**, 79–86.

ally we estimated
quantify loss/gain
healthcare governance
sources. Since the
this and previous
0) produced such
ended to the whole
d STEMI Archive.
illustrated here to
ks and to evaluate
users.

2004 guidelines for
1, **117**, 269–329.

l analysis of Public
*Data Modeling and
i*), Contributions to

m-onset-to-balloon
or acute myocardial

Annals of Statistics,

n-effects model for
ics, **3**, 15–29.

eristics, treatments
editerranean basin.
i, 1190–1210.

ng STEMI patients
ons in Applied and

pitals care: a case

ackage. *Rnews*, **7**,

ametric modelling

care and outcomes

s, Inc., New York.

generalized linear

ortality in patients
f Cardiology, **47**,