SIS 2017
Statistics and Data Science: new challenges, new generations

28–30 June 2017
Florence (Italy)

Proceedings of the Conference of the Italian Statistical Society

edited by
Alessandra Petrucci
Rosanna Verde

FIRENZE UNIVERSITY PRESS
2017

SIS 2017. Statistics and Data Science: new challenges, new generations : 28-30 June 2017 Florence (Italy) : proceedings of the Conference of the Italian Statistical Society / edited by Alessandra Petrucci, Rosanna Verde. – Firenze : Firenze University Press, 2017.
(Proceedings e report ; 114)

http://digital.casalini.it/9788864535210

ISBN 978-88-6453-521-0 (online)

Peer Review Process
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

Firenze University Press Editorial Board
A. Dolfi (Editor-in-Chief), M. Boddi, A. Bucelli, R. Casalbuoni, M. Garzaniti, M.C. Grisolia, P. Guarnieri, R. Lanfredini, A. Lenzi, P. Lo Nostro, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, G. Nigro, A. Perulli, M.C. Torricelli.

This work is licensed under a Creative Commons Attribution 4.0 International License
(CC BY 4.0: https://creativecommons.org/licenses/by/4.0/legalcode)

CC 2017 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

SIS2017 Committees

Scientific Program Committee:
Rosanna Verde (chair), Università della Campania "Luigi Vanvitelli"
Maria Felice Arezzo, Sapienza Università di Roma
Antonino Mazzeo, Università di Napoli Federico II
Emanuele Baldacci, Eurostat
Pierpaolo Brutti, Sapienza Università di Roma
Marcello Chiodi, Università di Palermo
Corrado Crocetta, Università di Foggia
Giovanni De Luca, Università di Napoli Parthenope
Viviana Egidi, Sapienza Università di Roma
Giulio Ghellini, Università degli Studi di Siena
Ippoliti Luigi, Università di Chieti-Pescara "G. D'Annunzio"
Matteo Mazziotta, ISTAT
Lucia Paci, Università Cattolica del Sacro Cuore
Alessandra Petrucci, Università degli Studi di Firenze
Filomena Racioppi, Sapienza Università di Roma
Laura M. Sangalli, Politecnico di Milano
Bruno Scarpa, Università degli Studi di Padova
Cinzia Viroli, Università di Bologna

Local Organizing Committee:
Alessandra Petrucci (chair), Università degli Studi di Firenze
Gianni Betti, Università degli Studi di Siena
Fabrizio Cipollini, Università degli Studi di Firenze
Emanuela Dreassi, Università degli Studi di Firenze
Caterina Giusti, Università di Pisa
Leonardo Grilli, Università degli Studi di Firenze
Alessandra Mattei, Università degli Studi di Firenze
Elena Pirani, Università degli Studi di Firenze
Emilia Rocco, Università degli Studi di Firenze
Maria Cecilia Verri, Università degli Studi di Firenze

Supported by:
Università degli Studi di Firenze
Università di Pisa
Università degli Studi di Siena
ISTAT
Regione Toscana
Comune di Firenze
BITBANG srl

Index

Preface
XXV

Alexander Agapitov, Irina Lackman, Zoya Maksimenko
Determination of basis risk multiplier of a borrower default using survival analysis
1

Tommaso Agasisti, Alex J. Bowers, Mara Soncin
School principals' leadership styles and students achievement: empirical results from a three-step Latent Class Analysis
7

Tommaso Agasisti, Sergio Longobardi, Felice Russo
Poverty measures to analyse the educational inequality in the OECD Countries
17

Mohamed-Salem Ahmed, Laurence Broze, Sophie Dabo-Niang, Zied Gharbi
Quasi–Maximum Likelihood Estimators For Functional Spatial Autoregressive Models
23

Giacomo Aletti, Alessandra Micheletti
A clustering algorithm for multivariate big data with correlated components
31

Emanuele Aliverti
A Bayesian semiparametric model for terrorist networks 37

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

Giorgio Alleva
Emerging challenges in official statistics: new sources, methods and skills
43

Rémi André, Xavier Luciani and Eric Moreau
A fast algorithm for the canonical polyadic decomposition of large tensors
45

Maria Simona Andreano, Roberto Benedetti, Paolo Postiglione, Giovanni Savio
On the use of Google Trend data as covariates in nowcasting: Sampling and modeling issues
53

Francesco Andreoli, Mauro Mussini
A spatial decomposition of the change in urban poverty concentration
59

Margaret Antonicelli, Vito Flavio Covella
How green advertising can impact on gender different approach towards sustainability
65

Rosa Arboretti, Eleonora Carrozzo, Luigi Salmaso
Stratified data: a permutation approach for hypotheses testing
71

Marika Arena, Anna Calissano, Simone Vantini
Crowd and Minorities: Is it possible to listen to both? Monitoring Rare Sentiment and Opinion Categories about Expo Milano 2015
79

Maria Felice Arezzo, Giuseppina Guagnano
Using administrative data for statistical modeling: an application to tax evasion
83

Monica Bailot, Rina Camporese, Silvia Da Valle, Sara Letardi, Susi Osti
Are Numbers too Large for Kids? Possible Answers in Probable Stories
89

Simona Balbi, Michelangelo Misuraca, Germana Scepi
A polarity–based strategy for ranking social media reviews
95

A. Balzanella, S.A. Gattone, T. Di Battista, E. Romano, R. Verde
Monitoring the spatial correlation among functional data streams through Moran's Index
103

Oumayma Banouar, Said Raghay
User query enrichment for personalized access to data through ontologies using matrix completion method
109

Giulia Barbati, Francesca Ieva, Francesca Gasperoni, Annamaria Iorio, Gianfranco Sinagra, Andrea Di Lenarda
The Trieste Observatory of cardiovascular disease: an experience of administrative and clinical data integration at a regional level
115

Francesco Bartolucci, Stefano Peluso, Antonietta Mira
Marginal modeling of multilateral relational events
123

Francesca Bassi, Leonardo Grilli, Omar Paccagnella, Carla Rampichini, Roberta Varriale
New Insights on Students Evaluation of Teaching in Italy
129

Mauro Bernardi, Marco Bottone, Lea Petrella
Bayesian Quantile Regression using the Skew Exponential Power Distribution
135

Mauro Bernardi
Bayesian Factor–Augmented Dynamic Quantile Vector Autoregression
141

Bruno Bertaccini, Giulia Biagi, Antonio Giusti, Laura Grassini
Does data structure reflect monuments structure? Symbolic data analysis on Florence Brunelleschi Dome
149

Gaia Bertarelli and Franca Crippa, Fulvia Mecatti
A latent markov model approach for measuring national gender inequality
157

Agne Bikauskaite, Dario Buono
Eurostat's methodological network: Skills mapping for a collaborative statistical office
161

Francesco C. Billari, Emilio Zagheni
Big Data and Population Processes: A Revolution?
167

Monica Billio, Roberto Casarin, Matteo Iacopini
Bayesian Tensor Regression models
179

Monica Billio, Roberto Casarin, Luca Rossini
Bayesian nonparametric sparse Vector Autoregressive models
187

Chiara Bocci, Daniele Fadda, Lorenzo Gabrielli, Mirco Nanni, Leonardo Piccini
Using GPS Data to Understand Urban Mobility Patterns: An Application to the Florence Metropolitan Area
193

Michele Boreale, Fabio Corradi
Relative privacy risks and learning from anonymized data
199

Giacomo Bormetti, Roberto Casarin, Fulvio Corsi, Giulia Livieri
A stochastic volatility framework with analytical filtering
205

Alessandro Brunetti, Stefania Fatello, Federico Polidoro
Estimating Italian inflation using scanner data: results and perspectives
211

Guénael Cabanes, Younès Bennani, Rosanna Verde, Antonio Irpino
Clustering of histogram data : a topological learning approach
219

Renza Campagni, Lorenzo Gabrielli, Fosca Giannotti, Riccardo Guidotti, Filomena Maggino, Dino Pedreschi
Measuring Wellbeing by extracting Social Indicators from Big Data
227

Maria Gabriella Campolo, Antonino Di Pino
Assessing Selectivity in the Estimation of the Causal Effects of Retirement on the Labour Division in the Italian Couples
235

Stefania Capecchi, Rosaria Simone
Composite indicators for ordinal data: the impact of uncertainty
241

Stefania Capecchi, Domenico Piccolo
The distribution of Net Promoter Score in socio–economic surveys
247

Massimiliano Caporin, Francesco Poli
News, Volatility and Price Jumps
253

Carmela Cappelli, Rosaria Simone, Francesca di Iorio
Growing happiness: a model-based tree
261

Paolo Emilio Cardone
Inequalities in access to job-related learning among workers in Italy: evidence from Adult Education Survey (AES)
267

Alessandro Casa, Giovanna Menardi
Signal detection in high energy physics via a semisupervised nonparametric approach
273

Claudio Ceccarelli, Silvia Montagna, Francesca Petrarca
Employment study methodologies of Italian graduates through the data linkage of administrative archives and sample surveys
279

Ikram Chairi, Amina El Gonnouni, Sarah Zouinina, Abdelouahid Lyhyaoui
Prediction of Firm's Creditworthiness Risk using Feature Selection and Support Vector Machine
285

Sana Chakri, Said Raghay, Salah El Hadaj
Contribution of extracting meaningful patterns from semantic trajectories
293

Chieppa A., Ferrara R., Gallo G., Tomeo V.
Towards The Register-Based Statistical System: A New Valuable Source for Population Studies
301

Shirley Coleman
Consulting, knowledge transfer and impact case studies of statistics in practice
305

Michele Costa
The evaluation of the inequality between population subgroups
313

Michele Costola
Bayesian Non–Negative l_1–Regularised Regression
319

Lisa Crosato, Caterina Liberati, Paolo Mariani, Biancamaria Zavanella
Industrial Production Index and the Web: an explorative cointegration analysis
327

Francesca Romana Crucinio, Roberto Fontana
Comparison of conditional tests on Poisson data
333

Riccardo D'Alberto, Meri Raggi
Non-parametric micro Statistical Matching techniques: some developments
339

Stefano De Cantis, Mauro Ferrante, Anna Maria Parroco
Measuring tourism from demand side
345

Lucio De Capitani, Daniele De Martini
Optimal Ethical Balance for Phase III Trials Planning
351

Claudia De Vitiis, Alessio Guandalini, Francesca Inglese, Marco D. Terribili
Sampling schemes using scanner data for the consumer price index
357

Ermelinda Della Valle, Elena Scardovi, Andrea Iacobucci, Edoardo Tignone
Interactive machine learning prediction for budget allocation in digital marketing scenarios
365

Marco Di Marzio, Stefania Fensore, Agnese Panzera, Charles C. Taylor
Nonparametric classification for directional data
371

Edwin Diday
Introduction to Symbolic Data Analysis and application to post clustering for comparing and improving clustering methods by the Symbolic Data Table that they induce
379

Carlo Drago
Identifying Meta Communities on Large Networks
387

Neska El Haouij, Jean-Michel Poggi, Raja Ghozi, Sylvie Sevestre Ghalila, Mériem Jaidane
Random Forest–Based Approach for Physiological Functional Variable Selection for Drivers Stress Level Classification
393

Silvia Facchinetti, Silvia A. Osmetti
A risk index to evaluate the criticality of a product defectiveness
399

Federico Ferraccioli, Livio Finos
Exponential family graphical models and penalizations
405

Mauro Ferrante, Giovanna Fantaci, Anna Maria Parroco, Anna Maria Milito, Salvatore Scondotto
Key–indicators for maternity hospitals and newborn readmission in Sicily
411

Ferretti Camilla, Ganugi Piero, Zammori Francesco
Change of Variables theorem to fit Bimodal Distributions
417

Francesco Finazzi, Lucia Paci
Space–time clustering for identifying population patterns from smartphone data
423

Annunziata Fiore, Antonella Simone, Antonino Virgillito
IT Solutions for Analyzing Large-Scale Statistical Datasets: Scanner Data for CPI
429

Michael Fop, Thomas Brendan Murphy, Luca Scrucca
Model–based Clustering with Sparse Covariance Matrices
437

Maria Franco-Villoria, Marian Scott
Quantile Regression for Functional Data
441

Gallo M., Simonacci V., Di Palma M.A.
Three–way compositional data: a multi–stage trilinear decomposition algorithm
445

Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson, Linda Sharples
Nonparametric shared frailty model for classification of survival data
451

Stefano A. Gattone, Angela De Sanctis
Clustering landmark–based shapes using Information Geometry tools
457

Alan E. Gelfand, Shinichiro Shirota
Space and circular time log Gaussian Cox processes with application to crime event data
461

Abdelghani Ghazdali
Blind source separation
469

Massimiliano Giacalone, Antonio Ruoto, Davide Liga, Maria Pilato, Vito Santarcangelo
An innovative approach for Opinion Mining : the Plutchick analysis
479

Massimiliano Giacalone, Demetrio Panarello
A G.E.D. method for market risk evaluation using a modified Gaussian Copula
485

Chiara Gigliarano, Francesco Maria Chelli
Labour market dynamics and recent economic changes: the case of Italy
491

Giuseppe Giordano, Giancarlo Ragozini, Maria Prosperina Vitale
On the use of DISTATIS to handle multiplex networks
499

Michela Gnaldi, Silvia Bacci, Samuel Greiff, Thiemo Kunze
Profiles of students on account of complex problem solving (CPS) strategies exploited via log–data
505

Michela Gnaldi, Simone Del Sarto
Characterising Italian municipalities according to the annual report of the prevention–of–corruption supervisor: a Latent Class approach
513

Silvia Golia
A proposal of a discretization method applicable to Rasch measures
519

Anna Gottard
Tree–based Non–linear Graphical Models
525

Sara Hbali, Youssef Hbali, Mohamed Sadgal, Abdelaziz El Fazziki
Sentiment Analysis for micro–blogging using LSTM Recurrent Neural Networks
531

Stefano Maria Iacus, Giuseppe Porro, Silvia Salini, Elena Siletti
How to Exploit Big Data from Social Networks: a Subjective Well–being Indicator via Twitter
537

Francesca Ieva
Network Analysis of Comorbidity Patterns in Heart Failure Patients using Administrative Data
543

Antonio Irpino, Francisco de A.T. De Carvalho, Rosanna Verde
Automatic variable and components weighting systems for Fuzzy cmeans of distributional data
549

Michael Jauch, Paolo Giordani, David Dunson
A Bayesian oblique factor model with extension to tensor data
553

Johan Koskinen, Chiara Broccatelli, Peng Wang, Garry Robins
Statistical analysis for partially observed multilayered networks
561

Francesco Lagona
Copula–based segmentation of environmental time series with linear and circular components
569

Alessandro Lanteri, Mauro Maggioni
A Multiscale Approach to Manifold Estimation
575

Tiziana Laureti, Carlo Ferrante, Barbara Dramis
Using scanner and CPI data to estimate Italian sub–national PPPs
581

Antonio Lepore
Graphical approximation of Best Linear Unbiased Estimators for Extreme Value Distribution Parameters
589

Antonio Lepore, Biagio Palumbo, Christian Capezza
Monitoring ship performance via multi–way partial least–squares analysis of functional data
595

Caterina Liberati, Lisa Crosato, Paolo Mariani, Biancamaria Zavanella
Dynamic profiling of banking customers: a pseudo–panel study
601

Giovanni L. Lo Magno, Mauro Ferrante, Stefano De Cantis
A comparison between seasonality indices deployed in evaluating unimodal and bimodal patterns
607

Rosaria Lombardo, Eric J Beh
Three–way Correspondence Analysis for Ordinal–Nominal Variables
613

Monia Lupparelli, Alessandra Mattei
Log–mean linear models for causal inference
621

Badiaa Lyoussi, Zineb Selihi, Mohamed Berraho, Karima El Rhazi, Youness El Achhab, Adiba El Marrakchi , Chakib Nejjari
Research on the Risk Factors accountable for the occurrence of degenerative complications of type 2 diabetes in Morocco: a prospective study
627

Valentina Mameli, Debora Slanzi, Irene Poli
Bootstrap group penalty for high-dimensional regression models
633

Stefano Marchetti, Monica Pratesi, Caterina Giusti
Improving small area estimates of households' share of food consumption expenditure in Italy by means of Twitter data
639

Paolo Mariani, Andrea Marletta, Mariangela Zenga
Gross Annual Salary of a new graduate: is it a question of profile?
647

Maria Francesca Marino, Marco Alfò
Dynamic random coefficient based drop-out models for longitudinal responses
653

Antonello Maruotti, Jan Bulla
Hidden Markov models: dimensionality reduction, atypical observations and algorithms
659

Chiara Masci, Geraint Johnes, Tommaso Agasisti
A flexible analysis of PISA 2015 data across countries, by means of multilevel trees and boosting
667

Lucio Masserini, Matilde Bini
Impact of the 2008 and 2012 financial crises on the unemployment rate in Italy: an interrupted time series approach
673

Angelo Mazza, Antonio Punzo, Salvatore Ingrassia
An R Package for Cluster–Weighted Models
681

Antonino Mazzeo, Flora Amato
Methods and applications for the treatment of Big Data in strategic fields
687

Letizia Mencarini,Viviana Patti, Mirko Lai, Emilio Sulis
Happy parents' tweets
693

Rodolfo Metulini, Marica Manisera, Paola Zuccolotto
Space–Time Analysis of Movements in Basketball using Sensor Data
701

Giorgio E. Montanari, Marco Doretti, Francesco Bartolucci
An ordinal Latent Markov model for the evaluation of health care services
707

Isabella Morlini, Maristella Scorza
New fuzzy composite indicators for dyslexia
713

Fionn Murtagh
Big Textual Data: Lessons and Challenges for Statistics
719

Gaetano Musella, Gennaro Punzo
Workers' skills and wage inequality: A time-space comparison across European Mediterranean countries
731

Marta Nai Ruscone
Exploratory factor analysis of ordinal variables: a copula approach
737

Fausta Ongaro, Silvana Salvini
IPUMS Data for describing family and household structures in the world
743

Tullia Padellini, Pierpaolo Brutti
Topological Summaries for Time–Varying Data
747

Sally Paganin
Modeling of Complex Network Data for Targeted Marketing
753

Francesco Palumbo,Giancarlo Ragozini
Statistical categorization through archetypal analysis
759

Michela Eugenia Pasetto, Umberto Noè, Alessandra Luati, Dirk Husmeier
Inference with the Unscented Kalman Filter and optimization of sigma points
767

Xanthi Pedeli, Cristiano Varin
Pairwise Likelihood Inference for Parameter–Driven Models
773

Felicia Pelagalli, Francesca Greco, Enrico De Santis
Social emotional data analysis. The map of Europe
779

Alessia Pini, Lorenzo Spreafico, Simone Vantini, Alessandro Vietti
Differential Interval–Wise Testing for the Inferential Analysis of Tongue Profiles
785

Alessia Pini, Aymeric Stamm, Simone Vantini
Hotelling meets Hilbert: inference on the mean in functional Hilbert spaces
791

Silvia Polettini, Serena Arima
Accounting for measurement error in small area models: a study on generosity
795

Gennaro Punzo, Mariateresa Ciommi
Structural changes in the employment composition and wage inequality: A comparison across European countries
801

Walter J. Radermacher
Official Statistics 4.0 – learning from history for the challenges of the future
809

Fabio Rapallo
Comparison of contingency tables under quasi–symmetry
821

Valentina Raponi, Cesare Robotti, Paolo Zaffaroni
Testing Beta–Pricing Models Using Large Cross–Sections
827

Marco Seabra dos Reis, Biagio Palumbo, Antonio Lepore, Ricardo Rendall, Christian Capezza
On the use of predictive methods for ship fuel consumption analysis from massive on–board operational data
833

Alessandra Righi, Mauro Mario Gentile
Twitter as a Statistical Data Source: an Attempt of Profiling Italian Users Background Characteristics
841

Paolo Righi, Giulio Barcaroli, Natalia Golini
Quality issues when using Big Data in Official Statistics
847

Emilia Rocco
Indicators for the representativeness of survey response as well as convenience samples
855

Emilia Rocco, Bruno Bertaccini, Giulia Biagi, Andrea Giommi
A sampling design for the evaluation of earthquakes vulnerability of the residential buildings in Florence
861

Elvira Romano, Jorge Mateu
A local regression technique for spatially dependent functional data: an heteroskedastic GWR model
867

Eduardo Rossi, Paolo Santucci de Magistris
Models for jumps in trading volume
873

Renata Rotondi, Elisa Varini
On a failure process driven by a self–correcting model in seismic hazard assessment
879

M. Ruggieri, F. Di Salvo and A. Plaia
Functional principal component analysis of quantile curves
887

Massimiliano Russo
Detecting group differences in multivariate categorical data
893

Michele Scagliarini
A Sequential Test for the C_{pk} Index
899

Steven L. Scott
Industrial Applications of Bayesian Structural Time Series
905

Catia Scricciolo
Asymptotically Efficient Estimation in Measurement Error Models
913

Angela Serra, Pietro Coretto, Roberto Tagliaferri
On the noisy high–dimensional gene expression data analysis
919

Mirko Signorelli
Variable selection for (realistic) stochastic blockmodels
927

Marianna Siino, Francisco J. Rodriguez-Cortés, Jorge Mateu, Giada Adelfio
Detection of spatio–temporal local structure on seismic data
935

A. Sottosanti, D. Bastieri, A. R. Brazzale
Bayesian Mixture Models for the Detection of High-Energy Astronomical Sources
943

Federico Mattia Stefanini
Causal analysis of Cell Transformation Assays
949

Paola Stolfi, Mauro Bernardi, Lea Petrella
Estimation and Inference of SkewStable distributions using the Multivariate Method of Simulated Quantiles
955

Paola Stolfi, Mauro Bernardi, Lea Petrella
Sparse Indirect Inference
961

Peter Struijs, Anke Consten, Piet Daas, Marc Debusschere, Maiki Ilves, Boro Nikic, Anna Nowicka, David Salgado, Monica Scannapieco, Nigel Swier
The ESSnet Big Data: Experimental Results
969

Jérémie Sublime
Smart view selection in multi-view clustering
977

Emilio Sulis
Social Sensing and Official Statistics: call data records and social media sentiment analysis
985

Matilde Trevisani, Arjuna Tuzzi
Knowledge mapping by a functional data analysis of scientific articles databases
993

Amalia Vanacore, Maria Sole Pellegrino
Characterizing the extent of rater agreement via a non–parametric benchmarking procedure
999

Maarten Vanhoof, Stephanie Combes, Marie-Pierre de Bellefon
Mining Mobile Phone Data to Detect Urban Areas
1005

Viktoriya Voytsekhovska, Olivier Butzbach
Statistical methods in assessing the equality of income distribution, case study of Poland
1013

Ernst C. Wit
Network inference in Genomics
1019

Dilek Yildiz, Jo Munson, Agnese Vitali, Ramine Tinati, Jennifer Holland
Using Twitter data for Population Estimates
1025

Marco Seabra dos Rei
Structured Approaches for High-Dimensional Predictive Modeling
1033

Key-indicators for maternity hospitals and newborn readmission in Sicily

Un sistema di indicatori per la classificazione dei punti nascita e re-ricoveri neonatali in Sicilia

Mauro Ferrante, Giovanna Fantaci, Anna Maria Parroco, Anna Maria Milito, Salvatore Scondotto

Abstract This paper proposes a composite indicator for the classification of maternity units, which takes into account for the different dimensions of service delivery, as potential predictors of health outcomes. As a measure of outcome, infant readmissions is considered, being a proxy of morbidity. The results highlight that after controlling for risk factors of the newborn, and for the presence of neonatal intensive unit, infants born in lower level hospitals show readmission rates higher than infants born in higher level hospitals.

Riassunto *Il presente articolo propone un indicatore composito per la classificazione dei punti nascita che tenga conto delle diverse dimensioni coinvolte nella qualità deli servizio erogato, quali potenziali predittori di esiti di salute. Come misura di esito vengono presi in esame i re-ricoveri neonatali, quale spia di possibili complicanze. I risultati, controllando per la presenza di unità di terapia intensive neonatale ed altri fattori di rischio, mostrano tassi di riammissione più elevate in strutture di basso livello, sulla base della classificazione proposta.*

Key words: *Composite indicator, birth-at-risk, healthcare evaluation, regionalization.*

Mauro Ferrante, Dipartimento Culture e Società, Università degli Studi di Palermo; email: mauro.ferrante@unipa.it

Giovanna Fantaci, Dipartimento Attività Sanitarie e Osservatorio Epidemiologico della Regione Sicilia; email: giovanna.fantaci@regione.sicilia.it

Anna Maria Parroco, Dipartimento di Scienze Psicologiche, Pedagogiche e della Formazione, Università degli Studi di Palermo; email: annamaria.parroco@unipa.it

Anna Maria Milito, Dipartimento Culture e Società, Università degli Studi di Palermo; email: annamaria.milito@unipa.it

Salvatore Scondotto, Dipartimento Attività Sanitarie e Osservatorio Epidemiologico della Regione Sicilia; email: salvatore.scondotto@regione.sicilia.it

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

1 Introduction

The relevance and quality of hospitals depend on various factors, among which the delivery volume, the geographic location, and the public or private ownership represent only a small component [2]. In regionalization programs the most used index to evaluate the quality of service delivery is given by the volumes of activity [4,7]. These programs generally aim at creating centers of excellence at which patients decide to receive care. Examples of this kind may be found in cancer care or complex survey procedures [6]. Nonetheless, there are situations, for example those related to acute situations for acute myocardial infarction [3] or in perinatal care [10], in which delivery volume alone is not able to provide an adequate picture of the complexity of factors which need to be evaluated in orienting regional healthcare programs.

Within the frame of maternity hospitals regionalization programs, the National Health and Medical Research Council recommend that pregnancies less than 33 week gestation be delivered at hospitals with neonatal intensive unit, in order to reduce the risk of mortality and morbidity. However, despite several studies have demonstrated better outcomes in high-level hospitals, with neonatal intensive unit and with high levels of delivery volume, for birth-at-risk (e.g. pre-term births) [1,9], less marked seem to be the differences in the case of low- or no risk childbirths [5,11]. Moreover, regionalization programs in perinatal care in Europe and North America, have determined a decrease in maternity hospitals with direct consequences also in terms of travel times required to reach the hospital [8]. This phenomenon determined a greater attention on the geographic location of maternity hospitals and, more in general, on issues related to maternity hospital accessibility.

Starting from these considerations, this Paper proposes a composite indicator for monitoring maternity hospitals in order to assist regionalization programs and to perform a classification of maternity hospitals in Sicily based on the proposed indicator. In order to evaluate the relationship between the proposed indicator and newborn readmission within 30 days from the childbirth event and hospital's category is performed.

2 Materials and methods

For the purposes of the present study several information sources were considered. The main information source is represented by Birth Certificates Records (CeDAP), which represent a unique information source for obtaining information on childbirth-related characteristics of both the mother and the infant. As a second information sources, Hospital Discharge Records have been used which, through an integration with CeDAP, allowed for the reconstruction of some characteristics of the childbirth event, e.g. transfers, complications, and readmissions. Finally, distance matrix among Sicilian municipalities has been considered in order to determine travel time

among mo M of and M In order to construct a composite indicator for maternity hospitals classification, several dimensions have been considered and for each dimension a set of variables have been selected, as reported in Table1.

For the delivery *volumes* dimension, both the total number of childbirths and the caesarian section rates have been considered, given the relevance that both these aspects represent within the actual national evaluation system of quality of care. However, in order to take into account also for the degree of *complexity* of the childbirth event, also caesarian section rates in Robson's categories 01 and 03 have has been included in the indicator, since the presence of a caesarian section childbirth may be a signal of inappropriateness. The third dimension considered relates to the territorial *basin* of the hospital, where the set of municipalities with at least 5% of its childbirths in the selected hospital defines the *basin* of the considered maternity hospital. For this dimension, the composite indicator includes both the total number of childbirths of women residing in the hospital's basin municipalities and the share of basin's childbirths in the hospital over the total number of basin's childbirths. As for the *travel time* dimension, the average travel time of total childbirths in the hospital has been included. Finally, for the *transfer* dimension, the share of childbirths with transfers within one day of both the mother and the newborn have been considered.

Table 1: Dimension, variables direction and weights of the composite indicator for maternity unit classification.

Dimension	Variables	Direction	Weight
Volumes	Total number of newborns **(V1)**	(+)	0.10
	% of caesarean section births **(V2)**	(-)	0.10
Complexity	% of caesarean section births in Robson = 01 or 03 **(C1)**	(-)	0.10
Basin	Total newborns of women living in hospital's basin municipalities **(B1)**	(+)	0.10
	% of basin's birth in the hospital **(B2)**	(+)	0.15
Travel time	Average travel time from municipality of residence and hospital's municipality **(T1)**	(-)	0.15
Transfer	% of transfers of the mother within one day from the birth **(TR1)**	(-)	0.15
	% of transfers of the newborn within one day from the birth **(TR2)**	(-)	0.15

Once derived these information, a standardization procedure has been used for variable transformation, and the elementary indices have been aggregated through a weighted average, with the weights system indicated in Table 1. Weights have been derived after a sensitivity analysis with different weights systems and through technical meeting with experts. The produced indicator allowed for a classification of maternity hospitals into three categories. Finally, in order to evaluate the association between maternity unit's category and birth outcome, multiple logistic regression analyses were performed in which readmission of newborns has been

considered as an outcome measure, as a function of the proposed classification, and of other factors, such as the presence of neonatal intensive care unit and by controlling for other childbirth-related risk factors. For the definition of childbirth-at-risk the following criteria have been applied: mother aged less than 20 or above 40 years old; very low birth weight (<1500 gr); gestational age lower than 37 weeks; multiple childbirth; small for gestational age (SGA).

3 Results

In 2014 a total of 44,436 newborns have been delivered in the 56 maternity hospitals of Sicily region. Nonetheless, after the application of the exclusion criteria, and record-linkage between Birth Certificate Records and Hospital Discharge Records, the valid cases resulted equal to 34,861.The average volume of activity resulted equal to 794, with a high degree of variability, with a minimum of 13 newborns in the case of Lipari Island hospital, and a maximum of 2538. Three hospitals have not been included in the analysis having less than 10 newborns in the year considered. A rather high degree of variability appeared also for all the other dimensions considered in the composite indicator.

In Figure 1 the relationship among delivery volumes, the score for the composite indicator proposed and the current regional classification are analyzed. Moreover, dashed lines indicates the cut points chosen for the categorization of the proposed indicator into three categories. It can be seen that, despite there is a direct relationship between delivery volumes and the proposed indicator, there are differences in terms of maternity units classification between the proposed classification and the current regional classification.

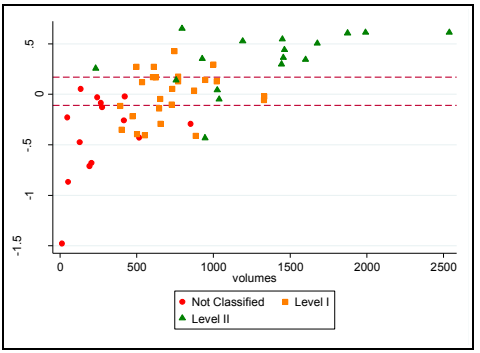


Figure 1: Delivery volumes, regional classification of maternity units, and composite indicator score, in Sicily, year 2014.

In Table 2 the results of logistic regression model are reported, in which the outcome variable assumes value 1 if the newborn has been readmitted within 30 days after the childbirth, and 0 elsewhere. In order to control for some risk-factors of the newborn and for the degree of severity, the presence of neonatal intensive care unit and other childbirth-related risk factors have been considered.

Table 2: Results of logistic regression model for infant readmission within 30 days after the childbirth by composite indicator categories, birth-at-risk and presence of neonatal intensive unit in the hospital

Variable	Category	OR	Inf CI 95%	Sup CI95%
Composite Indicator	High (Reference)			
	Medium	0.903	0.764	1.068
	Low	1.320**	1.065	1.636
Risk	Newborn-at-risk= No (Reference)			
	Newborn-at-risk = Yes	1.483***	1.267	1.736
UTIN	Neonatal intensive unit = No (Reference)			
	Neonatal intensive unit = Yes	4.431***	3.677	5.339
Constant		0.011***	0.009	0.013

*, **, and *** indicate significance level at 0.1, 0.05 and 0.01 respectively.

From the analysis of the results in Table 2, a strong direct association between both the presence of risk factor (OR=1.48) and the presence of the Neonatal Intensive Unit (OR=4.43) with infant readmission within 30 days after the birth event appears. Nonetheless, after adjusting for these factors, low-category maternity units show a higher risk of readmission compared to births happened in high- and medium-level hospitals (OR=1.320). By considering that more complicated situations should be assisted by higher-level hospitals, this result calls for monitoring actions aimed at exploring the reasons of these readmissions in low-level maternity units. Moreover, the low-level category of the proposed classification comprises not only low-volumes maternity units, but also level I and one level II maternity units, according with the current regional classification criterion.

4 Discussion and conclusion

The classification of maternity units represents an important issue, not only from the scientific perspective, related to the analysis of the determinants of health outcomes, but especially from the health policy perspective [2]. It has been showed that several elements such as geographic location, transfers, post-partum length of stay, and childbirth risk factors, should be considered in evaluating the quality of service delivery [1,5,8,9]. The proposed approach tries to overcome some of the limits of the current classification system of maternity units, which is based mainly on delivery volumes. The proposed indicator takes into account for a number of

these dimensions and it may constitute the basis for classifying maternity units into different levels based on a set of dimensions which take into account for both some volume-related aspects of service delivery, and for some quality and territorial aspects related to the childbirth event.

The results showed that the proposed classification may predict comorbidity conditions, as measured by 30-days infant readmissions after the birth event. Further research is required to better highlight the determinants of these readmission, jointly with the analysis of the relationship between the proposed classification, and other potential proxies of quality of care (e.g. mother readmissions, mother and infant mortality).

References

1. Bartels DB, Wypij D, Wenzlaff P, Dammann O, Poets CF. Hospital volume and neonatal mortality among very low birth weight infants. Pediatrics 2006; 117: 2206-14.
2. Falster MO, Roberts CL, Ford J, Morris J, Kinnear A, Nicholl M. Development of a maternity hospital classification for use in perinatal research. *NSW Public Health Bull*, 2012; 23(1-2): 12-16.
3. Ferrante M, Scondotto S, De Luca G, Fantaci G, Pollina-Addario S. Distance from the nearest hospital and mortality for acute myocardial infarction (AMI) in Sicily Region (Southern Italy). Epidemiologia e Prevenzione, 2014; 38(6): 373-378.
4. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*, 2002; 137(6): 511-520.
5. Hemminki E, Heino A, Gissler M. Should births be centralised in higher level hospitals? Experiences from regionalised health care in Finland. BJOG: An International Journal of Obstetrics & Gynaecology, 2011; 118(10): 1186-1195.
6. Hollenbeak CS, Rogers AM, Barrus B, Wadiwala I, Cooney RN. Surgical volume impacts bariatric surgery mortality: a case for centers of excellence. *Surgery*, 2008; 144(5): 736-743.
7. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA, 1996; 276(13): 1054-1059.
8. Pilkington H, Blondel B, Papiernik E, Cuttini M, Charreire H, Maier RF, Petrou S, Combier E, Kunzel W, Bréart G, Zeitlin, J. Distribution of maternity units and spatial access to specialised care for women delivering before 32 weeks of gestation in Europe. Health & place, 2010; 16(3): 531-538.
9. Rautava L. The Effect of the Birth Hospital and the Time of Birth on the Outcome of Finnish Very Preterm Infants. Annales Universitatis Turkuensis, ser D 902. Turku: University of Turku, 2010.
10. Ravelli ACJ, Jager KJ, de Groot MH, Erwich JJHM, Rijninks-van Driel GC, Tromp M, Eskes M, Abu-Hanna A, Mol BWJ. Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. BJOG: An International Journal of Obstetrics & Gynaecology, 2011; 118(4): 457-465.
11. Tracy SK, Sullivan E, Dahlen H, Black D, Wang YA, Tracy MB. Does size matter? A population-based study of birth in lower volume maternity hospitals for low risk women. BJOG 2006; 113: 86-96.