Ranking and Prioritization of Emergency Departments Based on Multi-indicator Systems

Enrico di Bella[1] [ID] · **Luca Gandullia[2]** · **Lucia Leporatti[1]** ·
Marcello Montefiori[1] · **Patrizia Orcamo[3]**

Abstract In this work we discuss how Emergency Departments (EDs) can be ranked on the basis of multiple indicators. This problem is of absolute relevance due to the increasing importance of EDs in regional healthcare systems and it is also complex as the number of indicators that have been proposed in the literature to measure ED performance is very high. Current literature faces this problem using synthetic (or numerically aggregated) indicators of a set of performance measures but, although simple, this solution has a number of drawbacks that make this choice inefficient: a compensation effect among the indicators; a high degree of subjectivism in the indicators weighting; opacity in the decision making; all the EDs are considered to be comparable. Indeed, the situations in which EDs are comparable (i.e. when all the performance of one ED are not lower than the performance indicators of the other) are a minority and incomparability is by itself a source of information that should be used to identify situations for which different policy actions should be designed. In this work we propose to use non compensatory composite indicators and partial ordering theory to rank and compare EDs giving value to the reasons of such an

✉ Enrico di Bella
  edibella@economia.unige.it

  Luca Gandullia
  luca.gandullia@unige.it

  Lucia Leporatti
  lucia.leporatti@unige.it

  Marcello Montefiori
  montefiori@unige.it

  Patrizia Orcamo
  patrizia.orcamo@regione.liguria.it

[1] Department of Economics and Business Studies, University of Genoa, Via Vivaldi, 5, 16126 Genoa, Italy

[2] Department of Political Sciences, University of Genoa, Largo Zecca 8/16, 16124 Genoa, Italy

[3] Department of Health, Liguria Region, Health Regional Agency, Piazza della Vittora 15, 16121 Genoa, Italy

🍂 Springer

incomparability. These methods are applied on a case study of 19 EDs in an administrative region in Italy.

Keywords Emergency departments · Multi-indicators systems · Partially ordered sets · Partial order scalogram analysis by coordinates · Non-compensatory composite indicators

1 Introduction

During the last decades, the use of Emergency Departments (EDs) in most of western countries significantly increased as a consequence of changes in the demographic structure of the population, in the epidemiology and in the patients' way of using emergency services (in particular, we refer to the increase of inappropriate use of EDs). It is a matter of fact that the increasing use of ED services risks to seriously compromise the quality of care and the efficiency of their management. For this reason, the evaluation of EDs, both in terms of cost-efficiency and quality, is a primary goal for policy makers committed in putting forth new policies intended to cost containment and quality improvement (Sørup et al. 2013; Graff et al. 2002). However, there is still not consensus on which indicators should be considered "more" accurate and extensive (Schull et al. 2011). The selection of specific measures strongly depends on the interest of different stakeholders (i.e. patients, policy makers, decision makers, administrators, physicians) (Tregunno et al. 2004) and the increasing need for reporting and monitoring ED performance resulted in a widespread effort put in the identification of a suitable subset of ED activity indicators able to capture the different aspects of the phenomenon. However, as stated by Schull et al. (2011), this effort brought to the identification of a large number of (not always fitting) indicators rather than to the identification of a smaller subset of appropriate and high-priority measures. The presence of multiple indicators of the same phenomenon, not perfectly equivalent but still in relation, may lead to results difficult to interpret and confusing for concrete policy making decisions. For this reason, most of the recent literature on ED performance focuses the analysis on prioritizing and selecting among the large set of available indicators (e.g. Lindsay et al. 2002; Schull et al. 2011; Sørup et al. 2013;). Since different measures of ED performance can bring to conflicting or not completely concordant conclusions, the information provided by the different indicators is often combined under some synthetic dimensions in order to come to non-contradictory policy actions for each particular ED.

There are many methods to reduce a large set of indicators to a univariate synthetic indicator (e.g.: OECD 2008; Huang et al. 2011), but the result tends often to depend on the functional form chosen for the synthetic indicator and, in particular, if a linear combination is selected, on the weights, which often come under scrutiny and controversy. Moreover, international literature has put in evidence that most of the commonly used techniques employed to build synthetic indicators suffer from a compensation effect[1] among the original indicators that may significantly affect policy making. Such an issue has resulted in a widespread effort to design non-compensatory techniques and two main research fields have been followed. Some techniques have been proposed to build non-compensatory composite indicators through the use of different aggregation schemes (e.g. the new version of the Human Development Index replaced the arithmetic mean with the geometric

---

[1] When a synthetic indicator is, for instance, a weighted sum of the elementary indicators, compensation means that a good value of such an indicator may be the results of a very good value for some indicators which masks potentially critical values for other indicators.

one, UNDP 2010) or by the introduction of penalties for those observations recording unbalanced values of the indicators (De Muro et al. 2011; Vidoli and Mazziotta 2013). These techniques for building non compensatory composite indicators bring to undeniable improvements over traditional methods and they represent a fruitful research branch since composite indicators are a particularly appreciated instrument for policy makers aiming at obtaining condensed and synthetic information of complex phenomena. An alternative, but partially complementary, approach to address the compensability issue and to avoid any subjective choice on the weighting has been proposed already during the 80s through the use of Theory of Partially Ordered Sets (POSET) of multi-indicator systems which have the benefit of keeping the valuable information derived from each single indicator distinct (see, for a review: Brüggemann and Patil 2011). POSET techniques are non-compensatory methods that do not require the reduction of a multifaceted and complex phenomenon to a unidimensional metric and their use is of particular interest when a phenomenon is described by a set of not concordant indicators complicating the establishment of a single ranking of observations. The main output of the POSET analysis is the Hasse diagram, a visual representation of comparabilities and incomparabilities across the elements of the set. The Hasse diagram is a very informative tool but it may result to be somehow too complex to be read and interpreted by policy makers. A more intuitive two dimensional representation of a partially ordered set can be obtained using the Partial Order Scalogram Analysis with base Coordinates (POSAC), which will be fully described in the following.

In this work we propose to approach the problem of EDs ranking and prioritization through the use of non-compensatory techniques, comparing the results obtained applying non-compensatory composite indicators with those obtainable using POSAC. To appreciate the effectiveness and usefulness of these techniques, we analyse and compare the performance of 19 EDs of the Liguria administrative region in Italy in 2013 measured through two sets of four indicators chosen according to the recent literature. Quality indicators measure the ability of EDs to guarantee a fast recognition of medical conditions and an adequate treatment. Previous researches (e.g.: Alessandrini and Knapp 2011; Hung and Chalut 2008; Schull et al. 2011) focus on waiting time spans, highlighting the relevance of time intervals between arrival and the first clinical assessment. In addition, most of previous papers on the topic (e.g.: Guttmann et al. 2006; Welch et al. 2011) use as a measure of ED performance the Length of Stay (LOS), which represents the time elapsing between the triage code assignment and the time of exit from the ED: ensuring low LOS is essential to prevent overcrowding and connected problems. The cost-efficiency of EDs may be defined as the ability to provide effective and high quality treatments minimizing the waste in equipment and expenditure (Solberg et al. 2003). These measures are mainly of interest for policy makers and ED managers who are continuously struggling with budget constraint issues. Measuring ED cost-efficiency may be pursed through an analysis of the average cost and average number of treatments prescribed in different EDs after correcting for differences in the patients' case-mix. Detecting EDs characterized by an excessive use of laboratory and non-laboratory exams may help in the identification of wasted resources to be restored.

2 Data and Methods

We use data provided by the registry of accesses to the EDs located in the administrative region of Liguria in Italy (1565,127 residents in 2013). Liguria has 19 EDs located across the region, seven of which are sited in Genoa, the regional capital (41% of total accesses).

The total number of accesses in 2013 was 633,982. The registry of ED accesses used contains information on patients (i.e. demographic characteristics such as age, gender and nationality) and on each access characteristics (e.g.: time of arrival, time of first examination, triage code, treatments received by each patient, diagnosis, outcome of the access, etc.). This database in our analysis represents the starting point from which indicators to measure ED activity are computed.

The two dominant categories of indicators used to measure ED quality are those connected to operational performance (or quality) and cost-efficiency (see, among the others, Sørup et al. 2013).

2.1 EDs Quality and Cost-Efficiency Indicators

Basing our choices on existing literature on the topic, we build up two sets of indicators to evaluate the activity of the 19 EDs of Liguria under a multi-indicator approach that will be fully detailed in the next section. To this extent four indicators for Quality and four indicators for Cost-efficiency were identified and computed for each ED.

Quality of services provided in EDs is strongly dependent on the timely provision of services. Three of the four indicators that we consider to capture the quality of services offered by each ED are based on yellow triage code[2] accesses which represent 21.8% of the 633,982 accesses. In fact, these patients represent an ideal category to be studied as they require ED care but without a concrete danger of death which, instead, characterizes red triage codes.[3] In particular, we compute two measures of patients' waiting time: the average waiting time before receiving the first visit for a yellow triage ($WAIT_Y$, in minutes) and the average length of staying in the ED for a yellow triage patient (LOS_Y, in minutes). In addition, we use the percentage of patients waiting more than 20 min as a measure of ED quality of services ($WAIT\text{-}20'_Y$): indeed, 20 min is considered the maximum waiting time for a patient classified as yellow, without compromising his/her safety. As a further measure of the quality of an ED we consider the waiting time spans for those patients characterized by a lower level of urgency of intervention (white and green patients), who represent the largest share of accesses (484.217 accesses, 76.4% of total accesses). We consider the percentage of white and green triage patients who wait more than 120 min before receiving the first treatment ($WAIT\text{-}120'_{WG}$); according to international standards (Ministero della Salute 2007) 2 h is indeed considered the maximum acceptable waiting time for non-urgent accesses. The four measures of ED Cost-efficiency are based on Yellow triage accesses as well (Cremonesi et al. 2010). Using the information on the number and type of each examination we are able to estimate the treatment cost per access, matching the examination code with the year 2013 regional official tariffs[4] for laboratory tests, non-laboratory tests and examinations. Therefore, we first obtain an estimate of the treatment cost for each access and, subsequently, we derive the average treatment cost of each Yellow code patient ($COST_Y$, in euros) in each ED. We also use the average number of treatments for Yellow triage codes measured in its partition into the

---

[2] In Italy the severity level of patients is classified using a four colours triage system (Levaggi and Montefiori 2013; Cremonesi et al. 2015): white = non-urgent/inappropriate access; green = non-urgent access; yellow = urgent access; and red = emergency.

[3] The concrete danger of death that characterizes red triage accesses may compromise the data quality on their waiting spans as in same situations medical staff gives priority to the patient's assistance rather than to a timely update of the patients' tracking system.

[4] In Italy regional tariffs are suggested by the Ministry of Health, but each regional district has the opportunity to arbitrarily revise them.

average number of visits (VISITS$_Y$), laboratory (LAB$_Y$) and non-laboratory exams (NOLAB$_Y$) for each access. These three last measures can be useful to identify those EDs that employ a disproportionate amount of treatments (split in the three main categories) given the same level of urgency of intervention (i.e. yellow triage code).

Due to the presence of missing information on treatments in several records, the final values of Cost-efficiency indicators are computed on 503.265 patients (80.625 of which are yellow codes).

As partial ordering techniques make use of ranking across observations, in order to reduce random incomparability due to non-significant differences among the indicator values (e.g. COST$_{Y(ED17)}$ = 83.01 €; COST$_{Y(ED15)}$ = 83.09 €) we run a cluster analysis for each single indicator with the aim of grouping EDs in an appropriate number of homogeneous groups. In particular, we selected the most adequate number of groups applying the Duda-Hart cluster stopping rule that requires to find the largest Je(2)/Je(1) values that correspond to a low pseudo-T-squared value that has much larger T-squared values next to it (Duda et al. 2001). In addition, to verify the results obtained using the Duda-Hart stopping rule we used a visual inspection of dendrograms that allow us to identify if the number of groups previously identified is confirmed under a hierarchical grouping approach. Some authors (e.g.: Raveh and Landau 1993; Brüggemann 2011; Brüggemann and Carlsen 2014a, b, 2015) argue that it may be convenient to use such interval rank-transformed data in order to render the analysis robust, in particular this occurs under the following circumstances: when the normal distribution of data is doubtful, when the number of observations in the dataset is small or when the dataset contains some outliers.

The values of the eight (Quality and Cost-efficiency) indicators computed for each ED are given in Table 1. For anonymity reasons, EDs are numbered sequentially from 1 to 19 and split into three categories according to the number of accesses in the study year: BIG (more than 40,000), MIDDLE (from 20,000 up to 40,000), SMALL (less than 20,000). The afore-mentioned groups are reported in the "capacity" column. Conversely the apices in Table 1 identify EDs group ranks, where 1 is the group with the lowest values (i.e. the group of the "best" EDs) whereas "worse" EDs are assigned to groups denoted with higher values. To measure the internal consistency of the indicators we applied the Cronbach's alpha on group ranked values for the two domains of Quality and Cost-efficiency; the coefficients turn out to be respectively equal to 0.865 and 0.763. These values denote an acceptable degree of concordance among the indicators, each of them bringing a different share of information.

2.2 Non-compensatory Composite Indicators

The construction of a composite indicator is generally developed through a series of subsequent steps (OECD 2008): (1) definition of the phenomenon to be measured; (2) selection of indicators; (3) normalization of individual indicators; (4) weighting and aggregation of single indicators. Although all the above-mentioned phases require a series of subjective decisions by the researcher (refer to Martinetti and Jacobi 2012) the aggregation step is the one responsible for the compensability criticism. Different approaches have been proposed to mitigate the compensability drawback: some (Mazziotta and Pareto 2015; Vidoli and Mazziotta 2013) propose the inclusion of a penalisation for unbalances among indicators while others (Munda and Nardo 2005) exploit Multi-Criteria Decision Analysis (MCDA) and build a ranking of observations in a complete pre-order through the use of pair-wise comparison of units. This second approach turns out to be computationally

Table 1 Values of the quality and cost-efficiency indicators

ED	Capacity	Quality indicators				Cost-efficiency indicators			
		$WAIT_Y$	$WAIT\text{-}20'_Y$	LOS_Y	$WAIT\text{-}120'_{WG}$	$COST_Y$	$VISIT_Y$	$NOLAB_Y$	LAB_Y
1	Middle	11.8[2]	15.2%[2]	207.5[2]	10.3%[3]	88.3[3]	1.6[4]	2.1[3]	7.0[3]
2	Big	19.9[4]	27.1%[3]	369.0[5]	28.0%[5]	114.3[5]	1.6[5]	2.4[4]	9.9[5]
3	Middle	25.9[5]	42.6%[4]	326.8[4]	11.6%[3]	106.6[4]	1.9[6]	1.9[3]	9.9[5]
4	Big	46.1[6]	55.3%[4]	333.0[4]	14.8%[4]	111.4[5]	1.4[3]	3.4[5]	16.5[7]
5	Middle	14.7[2]	19.2%[2]	121.7[1]	4.6%[2]	40.6[1]	1.1[2]	0.6[1]	4.2[2]
6	Middle	18.0[3]	26.7%[3]	195.5[2]	13.3%[3]	86.7[3]	1.4[3]	1.5[2]	10.0[5]
7	Small	13.1[2]	20.3%[2]	122.1[1]	0.5%[1]	44.5[1]	1.5[3]	1.2[2]	1.0[1]
8	Big	17.0[3]	25.7%[3]	265.3[3]	14.7%[4]	103.7[4]	1.4[3]	2.1[3]	10.1[5]
9	Big	22.7[4]	35.6%[3]	343.9[4]	21.7%[5]	101.0[4]	1.3[2]	2.1[3]	11.8[6]
10	Small	14.3[2]	19.6%[2]	650.8[6]	0.8%[1]	102.1[4]	1.5[4]	2.5[4]	8.9[4]
11	Small	7.7[1]	6.1%[1]	192.9[2]	0.7%[1]	111.3[5]	1.6[5]	2.1[3]	14.0[6]
12	Big	48.9[6]	67.3%[4]	632.1[6]	23.7%[5]	95.7[4]	1.2[2]	1.4[2]	12.7[6]
13	Small	16.3[3]	23.8%[3]	191.9[2]	2.3%[1]	120.2[5]	1.2[2]	2.6[4]	20.2[8]
14	Big	25.8[5]	44.1%[4]	325.7[4]	14.8%[4]	127.9[5]	2.2[7]	3.1[5]	7.9[4]
15	Big	19.0[4]	29.9%[3]	232.5[2]	15.0%[4]	83.1[3]	1.1[1]	2.0[3]	6.1[3]
16	Small	6.2[1]	3.7%[1]	148.2[1]	0.6%[1]	63.0[2]	1.0[1]	1.4[2]	6.7[3]
17	Middle	17.7[3]	28.0%[3]	405.2[5]	17.6%[4]	83.0[3]	1.5[3]	1.9[3]	6.2[3]
18	Middle	20.8[4]	32.8%[3]	622.4[6]	17.4%[4]	94.6[4]	1.5[4]	1.6[2]	8.6[4]
19	Middle	17.5[3]	27.8%[3]	209.6[2]	12.2%[3]	85.4[3]	1.3[2]	1.4[2]	10.3[5]
Number of groups		6	4	6	5	5	7	5	8

Apices indicate the groups assignments after the EDs clustering process

demanding when the number of observations to be ranked increases; in addition, it only allows to obtain a ranking across observations and not a composite indicator. As a consequence, in this context we exploit the first approach, ranking the 19 EDs using three different composite indicators built through the use of the Mazziotta–Pareto Index, geometric mean aggregation procedure and the Mean-Min function approach.

Let $X_{n \times m}$ be a data matrix whose generic element x_{ij} is the value of the j-th indicator (j in $J = \{1, \ldots, m\}$) of the i-th statistical unit (i in $I = \{1, \ldots, n\}$). All the procedures cited above require a preliminary standardisation of the indicators values from x_{ij} to z_{ij}. We will follow the Mazziotta and Pareto (2011) approach that brings to transformed indicators each having mean equal to 100 and standard deviation equal to 10:

$$z_{ij} = 100 \pm \frac{x_{ij} - \bar{x}_j}{s_j} 10 \tag{1}$$

with $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$ and $s_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}$; the sign $\pm$ is positive if the j-th indicator is directly correlated to the phenomenon under study and negative if inversely correlated.

2.2.1 Geometric Aggregation

The aggregation procedure implemented using the arithmetic mean has been widely used in previous years; however, additive aggregations imply full compensability among indicators, such that poor performance in some indicators can be compensated by high values in others. The increasing attention devoted to the mitigation of the compensability issue have led international institutions to consider geometric aggregation procedures for building composite indicators (OECD 2008): United Nations (UN) in 2010, started to use the geometric mean to compute the Human Development Index (UNDP 2010). The use of geometric mean instead of the arithmetic one implies that a low achievement in one dimension is no longer linearly compensated by a high achievement in another dimension. We will therefore provide estimations of composite indicators computed using a geometric aggregation of quality and cost efficiency standardized indicators.

2.2.2 Mazziotta–Pareto Index

The Mazziotta–Pareto Index (MPI) (De Muro et al. 2011; Mazziotta and Pareto 2011, 2015) is designed to take into account compensation among indicators penalising the discordant information among them. The generalized MPI is given by:

$$MPI_i^{-/+} = M_{z_i} \mp S_{z_i} cv_i \tag{2}$$

where M_{z_i} and S_{z_i} denote, respectively, the mean and standard deviation of the vector of the m standardised indicators of the i-th unit, $cv_i = S_{z_i}/M_{z_i}$ is the coefficient of variation of the i-th unit and the $\mp$ sign depends on the phenomenon measured: an "increasing/positive" composite indicator (e.g. quality and cost efficiency of EDs) is obtained using MPI^-, whereas for "decreasing/negative" composite indicators (e.g. inefficiency of EDs) MPI^+ is used. Formula (2) may also be written in this form:

$$MPI_i^{-/+} = M_{z_i} \left(1 \mp \frac{S_{z_i}}{M_{z_i}} cv_i \right) = M_{z_i} \left(1 \mp cv_i^2 \right) \tag{3}$$

where the coefficient $\left(1 \mp cv_i^2 \right)$ represents the aforesaid penalisation element for unbalances among indicators. Given a set of increasing/positive indicators and two units h and k such that $M_{z_h} = M_{z_k}$ and $S_{z_h} < S_{z_k}$, MPI_h^- will be higher than MPI_k^- because the information that the indicators provide on unit h is more concordant (and possible more reliable) than the information that they provide on unit k. Therefore, higher values of MPI for a unit, may be due to higher values of the original indicators or to a lower variability (called "horizontal variability").

2.2.3 Mean-Min Function Approach

The Mean-Min Function (MMF) is a two parameter function that allows the inclusion of a penalisation for compensation among indicators (Mazziotta and Pareto 2015) defined as:

$$MMF_i = M_{z_i} - \alpha \left(\sqrt{\left(M_{z_i} - \min_{j \in J} z_{ij} \right)^2 + \beta^2} - \beta \right) (0 \leq \alpha \leq 1; \beta \geq 0) \tag{4}$$

where M_{z_i} is the previously defined mean of the standardized values of indicators for each unit, $\min_{j \in J} z_{ij}$ is the minimum value among the m indicators of the i-th observation and α and β are the parameters used to define the intensity of penalisation for unbalance among indicators. If $\alpha = 0$ the function is reduced to the arithmetic mean (minimum penalisation), while if $\alpha = 1$ and $\beta = 0$ the function equals to the minimum standardized value of x_{ij}(maximum penalisation). Intermediate values of the parameters define specific penalisation levels. Although there are not criteria to select the values of α and β, in this paper we present the results obtained using $\alpha = 0.5$ and $\beta = 1$ because they provided stable results of the ranked transformed values of MMF.

2.3 The POSET Approach to EDs Ranking

Our starting point for a partially ordered set analysis is the construction of a matrix of ordinal data $ED_{n \times m}$ from the data matrix **X**, called scalogram, where the n rows (called *profiles*) represent the statistical units (in our case the EDs) and the m columns the ranks of the performance indicators. Denote with ed_i the profile of the i-th ED (i in $I = \{1, ..., n\}$) and with ed_{ij} the value of the j-th performance indicator (j in $J = \{1, ..., m\}$) for the i-th ED. Consider the profiles of two generic EDs ed_v and ed_w: if the m indicators that define the performances of the two units are consistently oriented (i.e. small values indicate higher levels of performance whereas larger values indicate lower levels), ed_v is claimed to be intrinsically worse than ed_w(and it is denoted with $ed_v > ed_w$) if all the indicators for ed_v are not smaller than the corresponding values for ed_w and at least one of them is bigger. It is possible to consistently define the cases of $ed_v < ed_w$ and $ed_v = ed_w$. Evidently, unanimity among indicators is not always met. When there is at least one element of ed_v bigger than the corresponding value for ed_w and at least one indicator for which the contrary is true, ed_v and ed_w are not comparable and this is denoted by $ed_v ed_w$. For instance, consider EDs 6, 7 and 19 and their $m = 4$ Cost-efficiency ranked indicators (Table 1)[5]: $_E ed_6 = (3, 3, 2, 5)$, $_E ed_7 = (1, 3, 2, 1)$ and $_E ed_{19} = (3, 2, 2, 5)$: $_E ed_6 > _E ed_7$, $_E ed_6 > _E ed_{19}$ but $_E ed_7 \|_E ed_{19}$. Although the Hasse diagrams are the most common form of posets representations (see Figs. 1, 2) they may result to be somehow confusing for policy makers that are generally used to have syntetic indicators and simplified rankings. For this reason, we think that a simpler, but still poset- founded, technique to rank EDs is the POSAC approach whose ease of interpretation for policy makers was already discussed in the literature (e.g.: di Bella et al. 2015; di Bella and Corsi 2016a, b).

POSAC is a non-metric procedure useful to reduce the dimensionality of a multivariate dataset trying to preserve as much as possible the order relations among the observations (Shye 1985; Shye and Amar 1985; Voigt et al. 2004; Bhat and Patil 2007; Brüggemann 2011). Compared to other methods that reduce the dimensionality of data (such as Principal Components Analysis, Multidimensional Scaling, etc.), the use of POSAC is particularly suitable when the objective is to rank the observations (e.g.: Patil and Taillie 2004; di Bella et al. 2015). It can thus be useful in our case, where the objective is to rank EDs from the least to the most performing in terms of Cost-efficiency and Quality of care provided. The POSAC technique is generally considered as a useful tool in multi-indicator systems (e.g.: Patil and Taillie 2004; Bhat and Patil 2007) but, to the best of our knowledge, this is the very first time this methodology is used to classify EDs.

---

[5] In the following $_E ed_i$ will indicate the profile of the ith ED for the Cost-efficiency set of indicators and $_E ed_i$ the profile of the i-th ED for the Quality set.
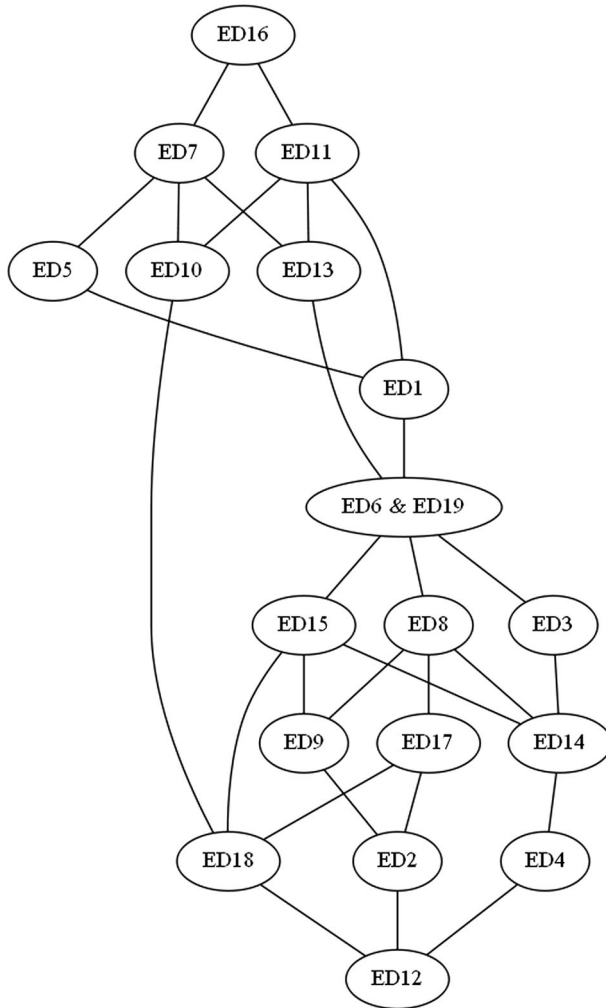
Fig. 1 Hasse diagram for the quality set of indicators

The POSAC technique uses an iterative procedure to assign two scores to each profile (represented by EDs in our dataset), X and Y (called base coordinates), so that the location of the points in the space reflects their partial ordering respect to the indicators mapping $ed_i \rightarrow (x_i, y_i)$ (i.e. $ED \rightarrow XY \in R^2$) such that:

$$ed_v > ed_w \leftrightarrow x_v \geq x_w \text{ and } y_v \geq y_w \tag{5}$$

$$ed_v \| ed_w \leftrightarrow \begin{cases} x_v \geq x_w \text{ and } y_v \leq y_w \\ \quad\quad or \\ x_v \leq x_w \text{ and } y_v \geq y_w \end{cases} \tag{6}$$

As described in Shye (1985, p. 181–185), the POSAC algorithm starts by computing the matrix of weak monotonicity coefficients among all the indicators and it identifies the two
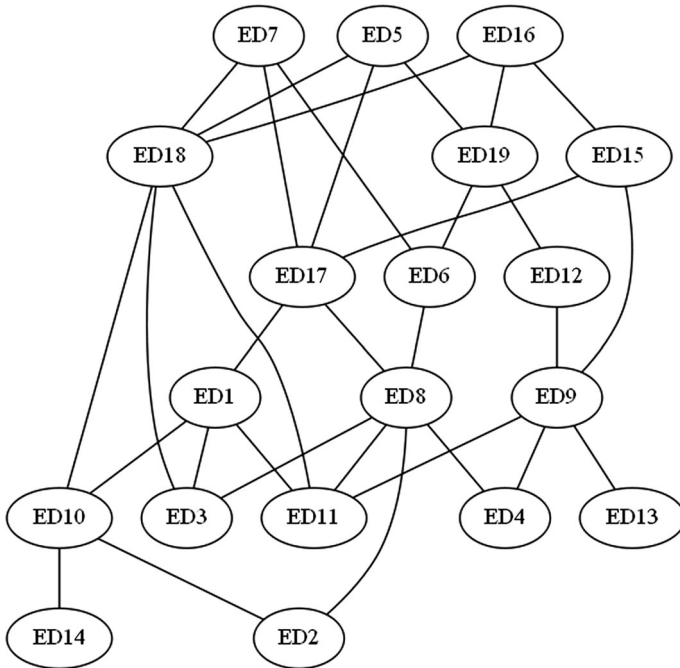
Fig. 2 Hasse diagram for the Cost-efficiency set of indicators

indicators that are the least positively correlated. Then, the initial (x_i, y_i) coordinates of i-th profile ed_i result from the following conditions:

$$x_i + y_i = \sum_{j=1}^{m} ed_{ij} \tag{7}$$

$$x_i - y_i = ed_{ia} - ed_{ib} \tag{8}$$

being ed_{ia} and ed_{ib} the scores of the two aforesaid least positively correlated indicators for the i-th ED. All the values computed using conditions (7) and (8) are transformed to place all profiles within the unit square. This initial approximation is improved minimizing a loss function defined on conditions (5) and (6). A steepest-descent process is carried out in the XY coordinates until it is not possible to improve the solution (see Shye and Amar (1985) for more details on the mathematical description of the algorithm). The final result is a Cartesian space the top right corner of which represents the best theoretical ED (top rank in all indicators, e.g. 1111) and the bottom left corner represents the worst theoretical ED (bottom rank in all variables, e.g. 6465 for Quality indicators). The line joining the best and the worst theoretical profiles (called Joint axis, or J-axis: $J = X + Y$) is the main dimension of the resulting two-dimensional space and its interpretation is straightforward: as we move along it, growing values of the coordinates indicate strict improvement in all rankings at the same time. Coherently with the Paretian approach, an ED is superior than another if every variable is at least equal to every corresponding variable of the other and at least one variable is lower with respect to the corresponding variable of the other ED. Therefore, efficient (or high quality) EDs will have high values of the J-axis whereas inefficient (or low quality) EDs will have low J-axis values. The line joining the two

remaining corners is called Lateral axis or L-axis ($L = X - Y$) and it represents the incomparability element of the EDs.[6] Moving along it can be interpreted as having some rankings that improve, while others deteriorate and a direct comparison is not possible: two profiles are considered incomparable if some variable is higher and some other is lower (for instance: 3334 with 4333). This is due to the fact that each variable has the same relevance, in terms of score, of another and no indicators weighting is done at all. Therefore, the L-axis has a role which may be considered to be similar to the "horizontal variability" of the MPI but, in this case, it has not only a numerical value but also a geometric interpretation. Consequently, determining which indicators are directly or inversely correlated to the J-axis explains why certain units of the dataset are represented closer to the bottom right or to the top left corner. Similarly, determining which indicators are directly or inversely correlated to the L-axis explains which are the indicators that may be identified as principal elements of incomparability among EDs.

In attempting to fit a set of empirically observed profiles into an order-preserving coordinate space of low dimensionality we do not always require a perfect fit but we need a measure to have an idea of how technically satisfactory a given solution is. The most natural measure of "goodness of fit" of the POSAC procedure (which can be interpreted analogously to the "stress measure" of Multidimensional Scaling) is the proportion of order relations (5) and (6), out of all profile pairs, correctly represented by their twofold coordinates. Considering a scalogram made of n profiles it is possible to have $n(n-1)/2$ pairwise profiles comparisons, so we have that such a goodness of fit measure is given by quantity:

$$POSAC_{gof} = \frac{\sum_{i=1}^{n(n-1)/2} I(i)}{\frac{n(n-1)}{2}} \times 100 \tag{9}$$

being $I(i)$ an indicator variable which assumes value 1 if the order relation relating the i-th comparison is correctly represented in the POSAC plane and 0 if not.

The EDs clustering shown in Table 1 was implemented in Stata (by StataCorp); the Hasse diagrams in Figs. 1 and 2 have been created using the PyHasse online software by Prof. R. Brüggemann[7] and the open source graph visualization software Graphviz[8]; POSAC analyses have been done using the Systat 11.0 package (by Systat Software Inc.); Figure 3 and the other analyses were run in R statistical environment (by R Core Team) and graphically enhanced (arrows) using the Paint.net open source images editing software.[9]

3 Results

3.1 Non-compensatory Composite Indicators

The use of non-compensatory composite indicators allows us to obtain a preliminary ranking of EDs according to their Quality and Cost-efficiency levels (Table 2). For what

---

[6] It is possible to find a sort of analogy between the L-axis and the penalisation coefficient of MPI in formula (2): both the two techniques try to include in the analysis a measure of the degree of concordance among the indicators.

[7] http://www.pyhasse.org.

[8] http://www.graphviz.org.
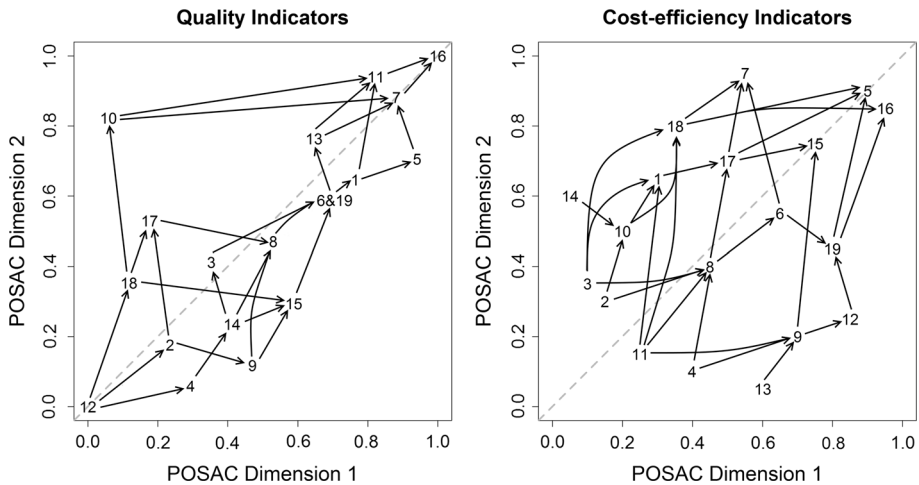
[9] http://www.getpaint.net.

Fig. 3 POSAC diagrams for the Quality and the cost-efficiency sets of indicators. Arrows correspond to the Hasse diagrams given in Figs. 1 and 2

concerns the Quality dimension, ED16 is unanimously identified as the best ED by all the non-compensatory composite indicators whilst ED12, ED4 and ED18 are the worst. Considering Cost-efficiency indicators, ED5 can be considered the most efficient, followed by ED7 and ED16. The difference in the composite indicators turns out to be particularly small for some EDs; let's consider for example the Cost-efficiency level of ED7 and ED16, which are alternatively classified as 2nd and 3rd depending on the use of MPI or MMF. The less cost-efficient EDs are clearly identified in ED14, ED13 and ED4. In accordance with previous results by Mazziotta and Pareto (2011), results turn out to be quite consistent over the different three methods employed, especially for what concerns MPI and Geometric aggregation. For Quality dimension, no ED changes its ranking when comparing MPI and Geometric aggregation and only 2 EDs (ED6 and ED12) change their ranking in Cost-efficiency performance. When also MMF ranking is considered, the number of EDs that modify their ranking is higher and respectively equal to 5 (26%) for Quality indicators and to 10 (53%) for cost-efficiency indicators. Generally, these differences are limited to one single ranking position although it is possible to observe that, especially for cost-efficiency indicators, the difference could be higher and one ED (ED10 for Quality indicators) changes its ranking for more than three positions (from 11 to 15).

3.2 Partial Order Analysis and POSAC Results

The Hasse diagrams in Figs. 1 and 2, respectively for the Quality and the Cost-efficiency sets of indicators, represent EDs ordering along with their incomparability component. From the bottom to the top we can read the position of each ED from the "worst" to the "best". Comparabilities are specified by the hedges that connect nodes. Considering Fig. 2 and referring to EDs 6, 7 and 19 it is immediate to see that EDs 6 and 7 and EDs 19 and 6 are connected by a line ($_Eed_7 < {}_Eed_6$, $_Eed_{19} < {}_Eed_6$) but EDs 7 and 19 ($_Eed_7 ||_E ed_{19}$) are not connected because they are incomparable. Weak forms of comparability are detectable considering the "levels" on which the two Hasse diagrams are organized as they represent weak forms of ordering among the EDs (Brüggemann 2011, Chapter 5). In other

Table 2 Composite indicators and ED ranking computed using the Geometric aggregation, the Mazziotta Pareto Index (MPI) and the Mean-Min function approach

ED	Quality indicators						Cost-efficiency indicators					
	Geometric mean	MPI	Mean Min function	Ranking Geometric mean	Ranking MPI	Ranking Mean Min function	Geometric mean	MPI	Mean Min Function	Ranking Geometric mean	Ranking MPI	Ranking Mean Min function
1	106.5	106.2	104.7	6	6	6	100.5	100.4	98.9	10	10	9
2	94.2	93.7	88.2	16	16	16	94.0	94.0	92.9	14	14	14
3	96.0	96.0	94.0	13	13	12	94.0	93.7	89.5	15	15	16
4	87.7	87.3	82.9	18	18	18	88.4	88.1	84.5	18	18	17
5	108.0	107.7	106.8	4	4	4	116.7	116.0	113.8	1	1	1
6	102.3	102.1	100.9	8	8	8	102.2	102.1	101.0	8	7	7
7	109.7	109.3	108.1	3	3	3	112.7	111.8	106.2	2	2	3
8	101.0	100.9	99.1	9	9	9	98.0	98.0	97.1	12	12	11
9	94.5	94.5	91.8	15	15	14	98.4	98.3	97.1	11	11	12
10	100.0	98.8	90.7	11	11	15	96.5	96.5	94.8	13	13	13
11	111.9	111.5	109.8	2	2	2	93.1	93.2	92.1	16	16	15
12	78.1	78.4	76.7	19	19	19	102.3	101.8	98.1	7	8	10
13	106.6	106.3	105.3	5	5	5	89.6	88.7	83.6	17	17	18
14	94.9	95.0	93.2	14	14	13	85.1	84.2	80.1	19	19	19
15	100.3	100.2	98.8	10	10	10	106.2	105.8	103.3	4	4	4
16	113.3	112.9	111.8	1	1	1	110.7	110.3	109.0	3	3	2
17	97.3	97.3	95.5	12	12	11	103.1	102.9	101.3	6	6	6
18	92.5	92.3	87.6	17	17	17	101.0	100.9	99.8	9	9	8
19	102.3	102.2	101.5	7	7	7	104.1	103.9	101.7	5	5	5

words, EDs that in a Hasse diagram are not connected are not comparable but it is possible to state that EDs in lower levels are generally better than EDs in upper levels. Moreover, identical profiles (i.e. $ed_v = ed_w$) are represented in the Hasse diagram by the same node (EDs 19 and 6 in Fig. 1) as there is neither quantitative, nor a qualitative basis for discriminating between them and only further analysis may suggest adequate policy actions.

Figure 3 gives the result of the POSAC procedure for the two sets of indicators while Table 3 reports the detailed values of the joint and lateral axes. The arrows represent the edges of the two Hasse diagrams in Figs. 1 and 2 and can be economically interpreted as a form of Pareto dominance (i.e.: EDs on the left of the arrow are dominated by the ones on the right of the arrow) in the partially ordered set based on the group ranks of Table 1.

The POSAC goodness of fit measure accounts to 88.8% for Quality indicators and to 92.4% for Cost-efficiency indicators. Similarly to what is done in Principal Components Analysis and Factorial Analysis, to interpret the meaning of the two axes (joint and lateral) generated by POSAC, the Spearman's rank correlation between them and each original indicators provides an interpretation of their meaning (Table 4). As expected, all the indicators are strongly and negatively correlated with the J-axis which assumes the role of POSAC synthetic indicator of the original indicators. The correlation values between the L-axis and the original indicators are less straightforward to interpret because the L-axis is not directly involved in EDs ranking but it plays a key role to represent the strength of incomparability among EDs: the farther two EDs are plotted along the L-axis, the more

Table 3 Joint and Lateral axes resulting from POSAC procedure for the two sets of indicators and ranking for J-axis values

ED	Quality indicators			Cost efficiency indicators		
	J-axis	L-axis	Ranking J-axis	J-axis	L-axis	Ranking J-axis
1	0.839	0.535	5	0.677	0.371	10
2	0.453	0.532	17	0.468	0.420	17
3	0.618	0.476	11	0.454	0.362	18
4	0.392	0.650	18	0.474	0.658	16
5	0.905	0.565	4	0.949	0.500	1
6	0.804	0.537	7	0.774	0.532	6
7	0.939	0.500	2	0.858	0.383	4
8	0.707	0.521	9	0.652	0.519	11
9	0.514	0.671	15	0.642	0.695	12
10	0.575	0.168	12	0.577	0.370	13
11	0.939	0.469	2	0.444	0.556	19
12	0.000	0.500	19	0.711	0.711	9
13	0.839	0.465	5	0.499	0.775	14
14	0.563	0.578	14	0.499	0.225	14
15	0.655	0.612	10	0.866	0.500	3
16	1.000	0.500	1	0.948	0.526	2
17	0.574	0.346	13	0.772	0.435	7
18	0.469	0.374	16	0.743	0.349	8
19	0.804	0.537	7	0.783	0.612	5

Table 4 Spearman's rank correlation between Joint and Lateral axes and original indicators

	Quality indicators				Cost efficiency indicators			
	$WAIT_Y$	$WAIT\text{-}20'_Y$	LOS_Y	$WAIT\text{-}120'_{WG}$	$COST_Y$	$VISIT_Y$	$NOLAB_Y$	LAB_Y
Joint	−0.866	−0.840	−0.887	−0.853	−0.915	−0.669	−0.778	−0.630
Lateral	0.329	0.301	−0.209	0.293	0.039	−0.649	−0.025	0.661

incomparable they are (and vice versa). Therefore, correlations between L-axis and elementary indicators explain which is the role that each elementary indicator has in the definition of dissimilarity among EDs (i.e. how relevant is its influence in the "horizontal variability"). Table 4 shows that some indicators have a weak correlation with the lateral (incomparability) axis stating that all of them bring coherent and useful information for EDs ranking. On the other side, the indicators that are more relevant in defining incomparabilities are LOS_Y for Quality indicators and $VISIT_Y$ and LAB_Y for Cost-efficiency indicators. These indicators are the ones responsible for incomparability among EDs and they may become the focus of targeted policy actions because they are the indicators that generate "out from the J-axis" situations.

For what it concerns Quality indicators, the best (ED16) and the worst EDs (ED12) can be easily identified both using composite indicators or the POSAC diagram in the coordinates (1; 1) and (0; 0) (Fig. 3; Table 3) as they are identified coherently in all indicators as the best (1,1,1,1) and worst (6,4,6,5) profiles. The high number of levels of the Hasse diagram of this first set of indicators (Fig. 1) corresponds in the POSAC representation to a structured group of EDs that can be quite clearly ranked following the arrows along the J-Axis from the "worst" to the "best". Whereas it is quite easy to rank all the EDs simply using composite indicators, such an ordering may not provide to policy makers all the useful information because "horizontal variability" is embedded inside the composite indicator itself, whereas POSAC makes it explicit through the L-axis that makes "horizontal variability" explicit: incomparable but opposite situations (e.g.: 22,344 and 44,322) may have the same value of a composite indicator but they are plot on opposite sides of the L-axis, pointing out that there are some internal differences that should be taken into account and suggesting that they may be object of different policy actions. Consider ED10 which has good values for three Quality indicators but a bad value for indicator LOS_Y: it lies far from the J-axis due to the effect of the correlation that the indicator has with the L-axis (Table 4). Without the POSAC representation and using only the non-compensatory composite indicators given in Table 2, it would be difficult to note this particular behavior.

Cost-efficiency indicators determine a situation that is less clear and this results in a Hasse diagram with a smaller number of levels (6 against 9) and a POSAC plot which has more EDs plotted along the L-axis. The combination of the Hasse diagram and the POSAC given in Fig. 3 for Cost-efficiency indicators outlines that it is possible to identify two main structured groups of EDs that can be analyzed separately: EDs 11, 4, 13, 8, 9, 6, 19, 12, and 16 (below the J-axis) and EDs 2, 3, 10, 14, 1, 17, 18 and 7 (above the J-axis). The EDs that are below the J-axis suffer LAB_Y issues (the indicator has a positive correlation with the L-axis) and, vice versa, the EDs that are above the J-axis have $VISIT_Y$ issues. EDs 5 and 16 are the closest to the top right corner of the plot and can be considered as two benchmarks or best practices to whom the other hospitals might refer for improving their profile.

Table 5 Spearman's rank correlation between *J*-axis rankings and non-compensatory composite indicators

	Quality indicators			Cost efficiency indicators		
	Geometric mean	MPI	Mean Min function	Geometric mean	MPI	Mean Min function
MPI	1.000			0.998		
Mean Min function	0.983	0.983		0.983	0.988	
J-axis	0.992	0.992	0.983	0.934	0.939	0.940

Looking at Fig. 3 and Table 3 it is evident that some EDs can be easily ranked for Quality and Cost-efficiency level using the POSAC graph and the ranked values of the *J*-axis; ranking is generally consistent with that established using composite indicators. The Spearman correlations between the ranked values of the *J*-axis coordinates and of the three non-compensatory indicators (Table 5) turn out to be higher than 0.93, suggesting the stability of results across the applied methods; in this sense, the ranked values of POSAC *J*-axis seem to be able to capture well the information derivable from non-compensative composite indicators. However, it is important to notice that, despite the high correlation, the ranking of *J*-axis coordinates is similar but not equal to that established using composite indicators with 8 and 15 EDs that slightly change rankings when comparing POSAC results with MPI indicators respectively for Quality and Cost-Efficiency indicators. In addition to the establishment of a single ranking, Fig. 3 allows us to identify also incomparable situations in which a single ranking cannot be established (e.g.: ED9 and ED19 for quality indicators). All this is not understandable looking solely at the values of the composite indicators, but it becomes immediately evident if we look at Fig. 3. The worst EDs are in the lower left corner of the two plots, and as long as we move along the main diagonal we find the EDs whose indicators are concordantly better. The EDs that are far from this diagonal (e.g. ED 10 for Quality indicators and ED 12 or ED 18 for Cost-efficiency indicators) are characterized by a sub-set of indicators that are substantially better (or worse) than the others. In short: it is possible to read the ranking of the EDs along the diagonal, easily identifying the EDs that the indicators concordantly identify as the best and the worst and this is generally consistent with the ranking given by the non-compensatory composite indicators in Table 2. Moreover, similar EDs are plotted close to each other and different EDs are plotted on opposite sides of the *J*-axis and this is an additional information that is not provided by such non compensatory indicators.

4 Conclusions

The ranking of EDs according to their costs and quality is propaedeutic to detect best practices or, on the opposite, critical issues that require attention and, eventually, targeted interventions. In order to do this, we believe that the use of multi-indicators systems and the partial ordering techniques herein discussed could be a useful tool at decision maker disposal.

The use of synthetic indicators can bring only partial (if not even misleading) information to policy makers, maybe not in the best or worst cases (which are generally well known and clearly identifiable even without any advanced statistical measurement

technique) but very likely in intermediate situations where a compensation effect flattens different situations. The non-compensatory approaches applied in this work (i.e., non compensatory composite indicators and POSAC) seem to mitigate the issue and they can become two complementary instruments for policy makers aiming at synthesizing complex phenomena in a ranking without losing relevant information. Indeed, incomparability among EDs (in particular shifts from the *J*-axis in POSAC) can become a source of additional information which may suggest, with due caution, feasible policy actions to cope with the problems at stake. At the same time, incomparability may become a nuisance that makes policy making difficult and should be limited (Brüggemann and Carlsen 2014a, b, 2015). The arrangement of too large multi-indicator systems may result in too many incomparabilities and the *L*-axis may be dominant over the *J*-axis with a blurred and uninformative POSAC output. For this reason, it may be appropriate to split all the indicators into separate systems (e.g. Quality and Cost-efficiency) each one representing one different domain of the problem.

Our suggestion to policy makers is to use systems of multi-indicators and partial ordering techniques like POSAC because they are suitable for helping the decision maker in choosing the best actions. The solutions suggested by these techniques are multi-dimensional and might provide the decision maker with a number of different lines of actions. In our case study, POSAC extends a mere ranking of EDs with a full map. Such a graphical tool may help policy makers to compare the relative positioning of each ED in relation with the others, quickly identifying the best and the worst ones, but also the one that are affected by specific issues. Certainly, this fact leaves an open question on how such a method can provide more information for policy makers without, at the same time, puzzling them, used to have simpler statistical tools. However, we believe that POSAC, compared to the Hasse diagram, still represents a viable solution for policy making and it can become the right balance between the use of synthetic indicators and POSET theory in policy making.

Anyhow, POSAC has some drawbacks that, if overcome, can incentivize its broader use in several research fields (di Bella and Corsi 2016a). At the moment, the POSAC routine can indeed be developed only using SYSTAT environment or the original FORTRAN code, which are not practically editable or customizable according to the researcher's needs. In addition, as POSAC is a two dimensional graph, the presence of a large number of elementary indicators, risks to result in a graph reporting a very high number of incomparabilities and often obliges researchers to divide variables in subsets of dimensions defining the phenomenon (in our case quality and cost efficiency). Another relevant limitation of POSAC concerns profiles weighting: when POSAC is applied on several units, the current procedure does not allow to weight profiles and if one profile occurs multiple times in a dataset all the duplicates are dropped and POSAC is plotted with respect of conditions (5) and (6) assigning the same importance to frequent and rare profiles. The definition of a more flexible POSAC procedure can mitigate these issues and significantly enlarge the application possibilities. As a consequence, we are currently evaluating the possibility of conducting POSAC technique to a specific case of more commonly used approaches in particular Multidimensional Scaling that has various similitudes with POSAC but a wider diffusion and a stronger mathematical formalization. All that said, POSAC and non-compensatory indicators or their possible developments should become important and synergic tools for policy evaluation.

References

Alessandrini, E. A., & Knapp, J. (2011). Measuring quality in pediatric emergency care. *Clinical Pediatric Emergency Medicine, 12*(2), 102–112.

Bhat, K., & Patil, G. (2007). Posac, data based weights, and mutual probability methods for multicriterion prioritization: A study in the theory and application of ranking methods, Center for Statistical Ecology and Environmental Statistics, Technical Reports and Reprint Series, p. 0703

Brüggemann, R., & Carlsen, L. (2014a). Incomparable—what now? *Match, 71*(3), 699–716.

Brüggemann, R., & Carlsen, L. (2014b). Incomparable: what now II? Absorption of incomparabilities by a cluster method. *Quality and Quantity, 49*(4), 1633–1645.

Brüggemann, R., & Carlsen, L. (2015). Incomparable - what Now, III. Incomparabilities, elucidated by a simple version of ELECTRE III and a fuzzy partial order approach. *Match, 73*(2), 277–302.

Brüggemann, R., & Patil, G. P. (2011). *Ranking and prioritization for multi-indicator systems—introduction to partial order applications.* Berlin: Springer.

Cremonesi, P., di Bella, E., & Montefiori, M. (2010). Cost analysis of emergency department. *Journal of Preventive Medicine and Hygiene, 51*(4), 157–163.

Cremonesi, P., di Bella, E., Montefiori, M., & Persico, L. (2015). The robustness and effectiveness of the triage system at times of overcrowding and the extra costs due to inappropriate use of emergency departments. *Applied Health Economics and Health Policy, 13*(5), 507–514.

De Muro, P., Mazziotta, M., & Pareto, A. (2011). Composite Indices of development and poverty: An application to MDGs. *Social Indicators Research, 104*(1), 1–18.

di Bella, E., Corsi, M & Leporatti L. (2016a). POSET analysis of panel data with POSAC. In M. Fattore and R. Brüggeman (eds), Partial order concepts in applied sciences, Springer, in press.

di Bella, E., Corsi, M., & Leporatti, L. (2015). A multi-indicator approach for smart security policy making. *Social Indicators Research, 122*(3), 653–675.

di Bella, E., Corsi, M., Leporatti, L., & Cavalletti, B. (2016). Wellbeing and sustainable development: A multi-indicator approach. *Agriculture and Agricultural Science Procedia, 8,* 784–791.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley.

Graff, L., Stevens, C., Spaite, D., & Foody, J. (2002). Measuring and improving quality in emergency medicine. *Academic Emergency Medicine, 9*(11), 1091–1107.

Guttmann, A., Razzaq, A., Lindsay, P., Zagorski, B., & Anderson, G. M. (2006). Development of measures of the quality of emergency department care for children using a structured panel process. *Pediatrics, 118*(1), 114–123.

Huang, I. B., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. *Science of the Total Environment, 409*(19), 3578–3594.

Hung, G. R., & Chalut, D. (2008). A consensus-established set of important indicators of pediatric emergency department performance. *Pediatric Emergency Care, 24*(1), 9–15.

Levaggi, R., & Montefiori, M. (2013). Definition of a prospective payment system to reimburse emergency departments. *BMC Health Services Research, 13,* 409.

Lindsay, P., Schull, M., Bronskill, S., & Anderson, G. (2002). The development of indicators to measure the quality of clinical care in emergency departments following a modified-delphi approach. *Academic Emergency Medicine, 9*(11), 1131–1139.

Martinetti, E. C., & von Jacobi, N. (2012). Light and shade of multidimensional indexes: How methodological choices impact on empirical results. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in Italy: Research and Reflections, Social Indicators Research Series 48* (pp. 69–103). Dordrecht: Springer.

Mazziotta, M., & Pareto, A. (2011). Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario. *Rivista di Statistica Ufficiale, 1,* 63–79.

Mazziotta, M., & Pareto, A. (2015). Comparing two non-compensatory composite indices to measure changes over time: a case study. *Statistika, 95*(2), 44–53.

Ministero Della Salute (2007) Progetto Mattoni SSN. Milestone 1.3—Definizione del sistema di valutazione dei pazienti (triage PS e 118). Available online: http://www.mattoni.salute.gov.it/ Accessed 10 July 2016.

Munda, G., & Nardo, M. (2005). *Non-compensatory composite indicators for ranking countries: A defensible setting* (p. 21833). EUR: EUR Report.

OECD. (2008). *Handbook on constructing composite indicators. Methodology and user guide.* Paris: OECD Publications.

Patil, G., & Taillie, C. (2004). Multiple indicators, partially ordered sets, and linear extensions: Multicriterion ranking and prioritization. *Environmental and Ecological Statistics, 11,* 199–228.

Raveh, A., & Landau, S. (1993). Partial order scalogram analysis with base coordinates (POSAC): Its application to crime patterns in all the states in the United States. *Journal of Quantitative Criminology, 9*(1), 83–99.

Schull, M. J., Guttmann, A., Leaver, C. A., Vermeulen, M., Hatcher, C. M., Rowe, B. H., et al. (2011). Prioritizing performance measurement for emergency department care: consensus on evidence based quality of care indicators. *Canadian Journal of Emergency Medicine, 13*(05), 300–309.

Shye, S. (1985). *Multiple scaling. The theory and application of partial order scalogram analysis*. Amsterdam: North-Holland.

Shye, S., & Amar, R. (1985). Partial order scalogram analysis by base coordinates and lattice mapping of items by their scalogram roles. In D. Canter (Ed.), *Facet theory: Approaches to social research* (pp. 277–298). New York: Springer.

Solberg, L. I., Asplin, B. R., Weinick, R. M., & Magid, D. J. (2003). Emergency department crowding: consensus development of potential measures. *Annals of Emergency Medicine, 42*(6), 824–834.

Sørup, C. M., Jacobsen, P., & Forberg, J. L. (2013). Evaluation of emergency department performance-a systematic review on recommended performance and quality-in-care measures. *Scandinavian Journal of Trauma Resuscitation & Emergency Medicine, 21*(1), 62–76.

Tregunno, D., Ross Baker, G., Barnsley, J., & Murray, M. (2004). Competing values of emergency department performance: Balancing multiple stakeholder perspectives. *Health Services Research, 39*(4 Pt 1), 771–792.

UNDP. (2010). *Human Development Report 2010*. New York: Palgrave MacMillan.

Vidoli, F., & Mazziotta, C. (2013). Robust weighted composite indicators by means of frontier methods with an application to European infrastructure endowment. *Statistica Applicata, Italian Journal of Applied Statistics, 23*(2), 259–282.

Voigt, K., Welzl, G., & Brüggemann, R. (2004). Data analysis of environmental air pollutant monitoring systems in Europe. *Environmetrics, 15*(6), 577–596.

Welch, S. J., Asplin, B. R., Stone-Griffith, S., Davidson, S. J., Augustine, J., Schuur, J., et al. (2011). Emergency department operational metrics, measures and definitions: results of the second performance measures and benchmarking summit. *Annals of Emergency Medicine, 58*(1), 33–40.