Journal of Productivity Analysis (2020) 53:355–375
https://doi.org/10.1007/s11123-020-00578-6

Operational efficiency vs clinical safety, care appropriateness, timeliness, and access to health care

The case of Portuguese public hospitals

Diogo Cunha Ferreira[1] · Alexandre Morais Nunes[1,2] · Rui Cunha Marques[1]

Published online: 25 May 2020

Abstract

Health care systems face resource scarcity that may jeopardise their financial sustainability as well as the quality of delivered health care. In view of that, the association between technical efficiency, access, and quality of services should be investigated, despite some past attempts that led to mixed, unclear, and perhaps biased results. We use a dataset composed of financial resources, hospital services, appropriateness and timeliness of care, patients' clinical safety, access to health care services, demographics, and epidemiology variables to study the aforementioned link regarding the Portuguese public hospitals (operating between 2013 and 2016). Quality and access data are aggregated into three main composite indicators, through Grey Relational Analysis (GRA). Bias- and environmentally corrected efficiency scores are estimated via bootstrap-based directional Data Envelopment Analysis. A double bootstrap algorithm is employed, using GRA-based quality indicators as predictors of technical efficiency. Evidence suggests that (1) Portuguese public hospitals exhibit low performance in terms of quality, while the different indicators present considerable correlation among them and with hospital size and patients' complexity characteristics; (2) patients' clinical safety, appropriateness and timeliness, as well as access to health care services are consistent and significant predictors of technical efficiency; and (3) the association between efficiency, quality, and access depends on the interaction between appropriateness, timeliness, and access. Therefore, quality and access can be improved with no efficiency sacrifice and vice versa.

Highlights

- We investigate the relationship between efficiency, quality and access in hospitals
- We use the Gray Relational Analysis to merge and simplify social performance data
- We use Data Envelopment Analysis to assess efficiency of public hospitals
- We use the double bootstrap for multiple regression (explanatory) analysis
- Efficiency, quality, and access to healthcare services evolve in the same sense.

Keywords Hospitals · Efficiency · Quality · Access · Data Envelopment Analysis · Grey Relational Analysis

JEL Classification C12 · C14 · C18 · C61 · I14 · I15 · H51

1 Introduction

Evaluating hospitals performance has been an exercise undertaken by a considerable number of researchers (Hollingsworth 2008; Hollingsworth and Peacock 2008). Most of them rely on technical efficiency assessment (Dervaux et al. 2009; Ferrier and Valdmanis 2004), still disregarding the presence and the impact of important care quality features (Shimshak et al. 2009). Indeed, health care services should improve patients' quality of life, efficiently manage

✉ Diogo Cunha Ferreira
diogo.cunha.ferreira@tecnico.ulisboa.pt

[1] CERis, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

[2] CAPP, Instituto Superior de Ciências Sociais e Políticas, University of Lisbon, R. Almerindo Lessa, 1300-666 Lisboa, Portugal

resources, and mitigate or erase barriers to health care (Ferreira et al. 2017b; Khushalani and Ozcan 2017).

Let quality be defined as the hospital's ability to provide safe, appropriate, and timely care to their patients. Patients' clinical safety is the ability of hospitals (and their staff) to safeguard patients from health care complications, including septicaemia, other infections, trauma, and tumbles. Appropriateness and timeliness of health care services regard the capacity of delivering timely patient-centred care supported by evidence-based guidelines. Timeliness is the capacity of providing timely care to the patient so as to prevent complications resulting from long waits.

Hospitals' performance encompasses not only the efficiency but also the social performance level that regards, among others, quality and access dimensions. While they should be maximised, resources waste must be erased for the sake of sustainability. Meanwhile, if a hospital is effective, then it can improve patients' health status (the care service's main goal), with the best quality levels possible. Performance can be seen as the merging of both efficiency and effectiveness concepts, such that if a health care provider has a good performance level, then it is simultaneously efficient and effective.

The interaction between quality, access, and efficiency in health care services is not well known (Gok and Sezen 2013) as the results found in the literature are rather controversial; see Ozcan (2014). Kissick (1994), for instance, coined the term "*iron triangle of health care*" to refer to three competing dimensions of health care: access, quality, and cost containment. This author argued that improvements in one dimension come at the expense of declines in one or both of the other two dimensions. One may argue that (not so unusual) cost constraints may exert an adverse effect upon the quality of provided services (Dismuke and Sena 2001). Others have concluded that quality improvements may lead to efficiency deterioration (Singaroyan et al. 2006) and to hospital costs increasing (Morey et al. 1992; Valdmanis et al. 2008). Poor quality and poor technical efficiency may also *walk hand in hand* (Clement et al. 2008; Mobley and Magnussen 2002). Still, one can improve efficiency by decreasing resource waste and/or increasing outputs (care services) production, with no quality sacrifice, see Chang et al. (2011), Ferrando et al. (2005), and Nayar and Ozcan (2008), or even with the latter's improvement, see Arocena and García-Prado (2007), and Helling et al. (2006). These mixed results lead Hvenegaard et al. (2011) to propose a *U*-shaped relationship between quality and efficiency. In other cases, a weak association (or even no association at all) between efficiency and quality was observed (Laine et al. 2005; Navarro-Espigares and Torres 2011). The relationship between those two concepts may also depend on some hospital features, such as size, investments on information technologies, and teaching status, see Gholami et al. (2015), Gok and Sezen (2013), Nayar et al. (2013), and Yang and Zeng (2014), or even on the variables adopted to characterise quality of provided services, as in Khushalani and Ozcan (2017), Martini et al. (2014), and Varabyova et al. (2016a). In short, the link efficiency-quality is still far from being fully understood.

Ferreira and Marques (2018) argue that the association between efficiency and access to health care services has been frequently disregarded from most of the analyses and it remains not understood. Despite their efforts to overcome such a shortcoming, they ignore the possible interactions that access can arguably have with quality, namely the clinical safety, the appropriateness, and the timeliness of the hospital services. Another relevant study that accounted for both outcomes and process measures of quality, in line with the Donabedian's model (Donabedian 1988, 2005), is the one of Ferrier and Trivitt (2013). They used a double Data Envelopment Analysis (DEA) approach to integrate thirteen quality dimensions in several ways, such as considering them either as extra outputs or adjustment factors. Ferrier and Trivitt concluded that quality, especially the outcomes, does indeed impact on efficiency. Nonetheless, the method used to include quality into the performance assessment model does matter, although no ultimate model was defined.

Although the relationship between efficiency and quality has been widely discussed in the literature, separating the term *quality* into two distinct (although related) concepts and analysing their interactions with access to health care and their conjoint impact on bias- and environment corrected efficiency distribution has never been done before. Hence, some questions deserve to be answered. Therefore, our research is concerned with testing four assumptions, as detailed and discussed below.

Hypothesis 1. Technical efficiency and clinical safety are positively associated.

Complications during care may be more or less dangerous depending on features such as the patient's age, comorbidities, and gender. Low levels of complications reduce needs in terms of length of hospital stay or new clinical procedures, contributing to patients' safety as it decreases their exposure to risk factors. Because good clinical safety levels lead to fewer interventions and, consequently, smaller expenditures per patient, clinical safety is expected to be positively related to efficiency.

Hypothesis 2. Technical efficiency and appropriateness and timeliness are positively associated.

Each patient typically requires differentiated and appropriate care procedures. If these are neither patient-centred nor supported by evidence-based guidelines, then health care is likely inappropriate/unsuitable, putting the patient under unnecessary risk. The same applies for untimely care, increasing the impact of risk factors. Therefore, appropriateness and timeliness of care are both positively associated with clinical safety, and for this reason we expect that the former are also positively associated with technical efficiency.

Hypothesis 3. Technical efficiency and access to health care services are positively associated.

A patient has access to a health care service if and only if she/he can use it whenever necessary and at her/his own will (Ferreira and Marques 2018). Thus, access can be measured by the service availability and the absence of barriers, either personal, financial, or organizational, including waiting time (Gulliford et al. 2002; Peters et al. 2008). The lower access levels, the higher the patients' severity of illness and the more resources consumed by the hospital to treat them, making it less efficient. Hence, we expect that technical efficiency and access to health care services are positively associated.

Hypothesis 4. At least two of the social performance dimensions—clinical safety, care appropriateness and timeliness, and access—interact, and this interaction is positively associated with efficiency of hospitals.

By the previous discussion, the dimensions associated with the social performance may interact and this interaction may be related to the efficiency of hospital services. For instance, we have concluded that clinical safety and appropriateness may be associated, thus they may interact. The safer and the more appropriate the care provided, the more efficient the hospital. Other interactions are also possible.

To analyse these four hypotheses and evaluate the association level between technical efficiency and quality/access, we combine nonradial directional distance-based DEA (Charnes et al. 1979) and Grey Relational Analysis (GRA) (Deng 1982; Girginer et al. 2015; Kuo et al. 2008), using the double bootstrap technique of Simar and Wilson (2007). Double bootstrap allows for a Multivariate Regression Analysis (MRA) in the presence of interaction terms, if required. Furthermore, it accounts for the data generating process and permits the bias-correction of efficiency scores. However, it does not allow the correction of efficiency estimates by the external (operational) nondiscretionary environment, which consists of market conditions, epidemiology, and demographics in the case of hospitals. Hence, we make a small change in the double bootstrap to consider those two issues. Although the MRA is useful to account for more than one potential predictor for efficiency, a multicollinearity and a discriminating power loss problem may be present. They imply that we should be parsimonious in the choice of these predictors. GRA is a technique that condenses the number of predictor variables into some composite indicators correlated with the raw data, without substantial loss of information. Due to the simplicity of GRA, we use it to mitigate that dimensionality problem. Therefore, we seek to contribute to both health care management and operational research fields. The former not only tries to answer to some questions that remain unclearly answered in the literature but also to give a step forward on the utilisation of double bootstrap that has shown to be a robust technique to evaluate potential associations between efficiency and some predictors.

The present paper is structured as follows. Section 2 describes the adopted methods (DEA, GRA, and double bootstrap). Section 3 deals with the case study, describing data and variables. Section 4 provides and discusses the main findings. Section 6 concludes this document.

2 Methods

2.1 Overview

Given the mixed results in extant literature and the shortcomings in previously adopted methods, we propose the utilisation of some alternative methods to analyse the association between efficiency, appropriateness and timeliness of care, patients' safety, and access to care services, and their potential interactions (if any). These measures of (secondary) health care quality can be understood as the process quality and outcomes from Donabedian's triad (Donabedian 2005). To analyse the aforementioned interaction in hospitals, we combine the bias- and environment adjusted nonradial directional-based DEA, the GRA, and the double bootstrap method.

2.2 Data envelopment analysis and the nonradial directional distance functions

DEA (Banker et al.1984) assumes that a set of n hospitals,

$$J = (1, \ldots, j, \ldots, n),$$

can produce the same kind of outputs (goods or services), from similar inputs (resources).

Let $\mathbf{x}^k = (x_1^k, \ldots, x_i^k, \ldots, x_m^k)$ denote the vector whose m components are the input levels used by hospital $k \in J$. Likewise, let $\mathbf{y}^k = (y_1^k, \ldots, y_r^k, \ldots, y_s^k)$ be the vector of s different outputs delivered by the hospital $k \in J$.

The standard formulation of DEA assumes radial contraction (resp. expansion) of inputs (resp. outputs), which is difficult to achieve in practice. For that reason, nonradial DEA-based models are usually preferred for empirical applications; see Zhang and Choi (2014). To obtain optimal levels of inputs and outputs and, consequently, the efficiency score of hospital $k \in J$ we use a nonradial Directional Distance Function (DDF)-based DEA; see Chambers et al. (1996), and Chambers et al. (1998). This model is more flexible than the standard DEA because it allows for both input contraction and output expansion at different rates (Ferreira and Marques 2016c; Zhang and Choi 2014; Zhou et al. 2012).

Let $\lambda_t^k \in [0, 1]$, $k \in J$, $t = 1, \dots, n$, be the tth coefficient associated with the kth hospital. The n components of $\boldsymbol{\lambda}^k$ should be optimised, for instance, through linear programming. Denote optimal values of variables with $\star$ (superscript). Let $f_n : \mathbb{R}_+^{2n} \mapsto \mathbb{R}_+ \cup \{0\}$ be an aggregating function with two vectorial entries of the same size, n. Using a vector of weights, $\boldsymbol{\lambda}^k = (\lambda_1^k, \dots, \lambda_j^k, \dots, \lambda_n^k)$, and linear combination, f_n transforms the vector $\mathbf{B} = (B_1, \dots, B_j, \dots, B_n)$ into a scalar B, *i.e.* $f_n(\mathbf{B}, \boldsymbol{\lambda}^k) = B \geqslant 0$. Should those weights range between 0 and 1 and obey to $\sum_{j=1}^n \lambda_j^k = 1$, then f_n represents the weighted arithmetic mean of $\mathbf{B}$:

$$f_n(\mathbf{B}, \boldsymbol{\lambda}^k) = \langle \mathbf{B}, \boldsymbol{\lambda}^k \rangle = \mathbf{B}^\top \boldsymbol{\lambda}^k = \sum_{j=1}^{n} \lambda_j^k \cdot B_j = B.$$

DEA estimates targets for both inputs and outputs using the linear combination of efficient hospitals located in the boundary, as follows

$$(x_i^k)^\star = \langle \boldsymbol{\lambda}^k, \mathbf{x}_i \rangle = \sum_{j=1}^{n} \lambda_j^k \cdot x_i^j, i = 1, \dots, m,$$

and

$$(y_r^k)^\star = \langle \boldsymbol{\lambda}^k, \mathbf{y}_r \rangle = \sum_{j=1}^{n} \lambda_j^k \cdot y_r^j, r = 1, \dots, s,$$

where $\mathbf{x}_i = (x_i^1, \dots, x_i^j, \dots, x_i^n)^\top$, $i = 1, \dots, m$, and $\mathbf{y}_r = (y_r^1, \dots, y_r^j, \dots, y_r^n)^\top$, $r = 1, \dots, s$. We verify that $(x_i^k)^\star x_i^k$ and $(y_r^k)^\star y_r^k$, for any i, r. Inequalities can be transformed into equations by using (nonnegative) slacks: $(x_i^k)^\star + (S_i^k)^- = x_i^k$, and $(y_r^k)^\star - (S_r^k)^+ = y_r^k$, for any i, r. Assume that these slacks result from unknown components, $(\boldsymbol{\delta}^k)^+$ and $(\boldsymbol{\delta}^k)^-$—to be optimised—and known directional components, $(\mathbf{d}^k)^+$ and $(\mathbf{d}^k)^-$, such that $(S_i^k)^- = (\delta_i^k)^- \cdot (d_i^k)^-$, $i = 1, \dots, m$, and $(S_r^k)^+ = (\delta_r^k)^+ \cdot (d_r^k)^+$, $r = 1, \dots, s$. For the sake of simplicity, we assume that

$$(d_i^k)^- = \frac{\langle \mathbf{1}, \mathbf{x}_i \rangle}{n} \text{ and } (d_r^k)^+ = \frac{\langle \mathbf{1}, \mathbf{y}_r \rangle}{n}, \forall i, r,$$

where $\mathbf{1}$ is a n-length vector of ones. In other words, the directional vector results from the arithmetic average of observable values in dataset. Given this framework, the nonradial DDF program becomes:

$$\mathcal{D} = \max_{\boldsymbol{\lambda}, \boldsymbol{\delta}^-, \boldsymbol{\delta}^+} \sum_{i=1}^{m} (\delta_i^k)^- + \sum_{r=1}^{s} (\delta_r^k)^+ \tag{1a}$$

subject to:

$$\langle \boldsymbol{\lambda}^k, \mathbf{x}_i \rangle + \frac{\langle \mathbf{1}, \mathbf{x}_i \rangle}{n} \cdot (\delta_i^k)^- = x_i^k, \ i = 1, \dots, m, \tag{1b}$$

$$\langle \boldsymbol{\lambda}^k, \mathbf{y}_r \rangle - \frac{\langle \mathbf{1}, \mathbf{y}_r \rangle}{n} \cdot (\delta_r^k)^+ = y_r^k, \ r = 1, \dots, s, \tag{1c}$$

$$\langle \mathbf{1}, \boldsymbol{\lambda}^k \rangle = 1, \tag{1d}$$

$$\boldsymbol{\lambda}^k, (\boldsymbol{\delta}^k)^-, (\boldsymbol{\delta}^k)^+ 0. \tag{1e}$$

Once $\boldsymbol{\lambda}^k$, $(\boldsymbol{\delta}^k)^+$, and $(\boldsymbol{\delta}^k)^-$ have been optimised by model (1a)–(1e), one can estimate the efficiency score of hospital $k \in J$, denoted by θ^k:

$$\begin{aligned}\theta^k &= \tfrac{1}{m} \cdot \sum_{i=1}^{m} \tfrac{(x_i^k)^\star}{x_i^k} \Big/ \tfrac{1}{s} \cdot \sum_{r=1}^{s} \tfrac{(y_r^k)^\star}{y_r^k} \\ &= \frac{s}{m} \frac{m - \frac{1}{n} \sum_{i=1}^m \frac{(\delta_i^k)^-}{x_i^k} \sum_j x_i^j}{s + \frac{1}{n} \sum_{r=1}^s \frac{(\delta_r^k)^+}{y_r^k} \sum_j y_r^j}\end{aligned} \tag{2}$$

Quantities $(\boldsymbol{\delta}^k)^-$ and $(\boldsymbol{\delta}^k)^+$ are nonnegative because of $\sum_{j=1}^n \lambda_j^k \cdot x_i^j x_i^k$ and $\sum_{j=1}^n \lambda_j^k \cdot y_r^j y_r^k$, for any i, r. Thus, $0 < \theta^k \leqslant 1$. A finding that $\theta^k = 1$ for $k \in J$ indicates that hospital k is technically efficient. Moreover, $1 - \theta^{k'}$ identifies the inefficiency level of hospital $k' \in J$.

2.3 Correction of efficiency estimates for bias and operational conditions

Efficiency scores must be adjusted for nondiscretionary variables characterising the exogenous environment under which hospitals operate. It is worth noting that these exogenous environmental variables do not include the quality ones (safety, appropriateness, timeliness, and access). Rather, by exogenous variables or operational conditions we mean the demographic and epidemiological dimensions that are prone to drive technical efficiency and are non-discretionary to hospital managers (quality and access are not exogenous to the production process). The adjustment for exogenous factors is compulsory because hospitals usually face quite heterogeneous environments, meaning that some act on unfair conditions when compared with the others, see Bădin et al. (2012), Daraio and Simar (2005), Daraio and Simar (2007b), and Kontodimopoulos et al. (2010).

Efficiency scores θ^k are likely biased due to the presence of noise. To mitigate such a bias, we employ the conditional subsampling method. The conditionality introduced here allows the correction of efficiency for operational conditions and bias, simultaneously. Let $\mathbf{z}^k = (z_1^k, \dots, z_v^k, \dots, z_w^k)$ be the vector of w exogenous dimensions featuring the hospital $k \in J$. Following Ferreira and Marques (2016a), we construct a comparability subset for the hospital $k \in J$ (under analysis) from the original sample. The subset, denoted by Ω^k, replaces J in (1a)–(1e), and is specified as follows:

$$\Omega^k = \left\{ \bigcup_{j \in J} j \text{ s.t. } \begin{matrix} z_v^j > z_v^k - h_v, \\ z_v^j < z_v^k + h_v, \end{matrix} \; v = 1, \ldots, w \right\}, \tag{3}$$

where $h_v > 0$ is a variable-specific bandwidth that can be estimated, for instance, through the Silverman's rule of thumb (Silverman1986). For the second-order Epanechnikov kernel, $h_v \approx 2.34 \cdot \hat{\sigma}_v \cdot n^{-1/5}$, where $\hat{\sigma}_v$ is the empirical standard deviation of z_v, $v = 1, \ldots, w$ (Ferreira et al. 2017a).[1] We obtain conditional efficiency measures when h_v is finite. Should $h_v \mapsto +\infty$, then $\Omega^k \mapsto J$, and we estimate unconditional efficiency scores.

Once defining the comparability subset Ω^k for $k \in J$, one can construct the conditional subsampling procedure to correct efficiency estimates for bias and the environment. Let $0 \leqslant \gamma \leqslant 1$ be a quantity and $N = [n^\gamma]$ the size of a subsample.[2] Daraio and Simar (2007a) suggest using $\gamma > 0.50$. In the present case, $\gamma = 0.90$, although results have shown stability for $0.85 \leqslant \gamma \leqslant 0.95$. The procedure is defined in Algorithm A (see Appendix A),[3] whose final outcome is the efficiency estimate corrected for bias and the environment, $\mathbb{B}(\theta^k)$, for the hospital under analysis, $k \in J$. More details regarding subsampling can be found in Daraio and Simar (2007a).

2.4 Grey relational analysis

The dimensionality related to the number of independent variables is a problem underlying multiple regressions. Should the sample be small, then the statistical power of the method may be worsened, which can be problematic. At this point, one should be aware that a considerable number of quality-related variables are available and can be used to conduct the second stage of the present study (vide infra). Using them all as explanatory variables may bring undesirable outcomes from the multiple regression due to multicollinearity problems. Alternatively, one may merge those independent quality and access variables into a smaller set, in an intuitive way. Thereby, we use the GRA method of Deng (1982) to reduce the set of explanatory variables into three composite indicators: appropriateness and timeliness, patients' clinical safety, and access to health care services. These indicators will be used as explanatory variables of hospital technical efficiency.

GRA is an ordinal multicriteria decision aiding tool that ranks several different alternatives (hospitals) given their performances in a set of criteria or attributes (Kuo et al. 2008). Hereinafter, we use the term *attribute* to denote each quality/access variable. GRA has been applied in a number of fields, including health care (Girginer et al. 2015), watermarking schemes (Lin et al. 2011), hiring (Olson and Wu 2006), and biomass boilers optimisation (Morán et al. 2006).

Let a set of attributes be denoted by $\mathcal{Q}_\ell^{j(g)}$, $j \in J$, $\ell = 1, \ldots, Q$, $g = 1, 2, 3$. They can be grouped into "*the larger, the better*" attributes, $\mathcal{Q}_\ell^{j(g)}$, $j \in J$, $\ell = 1, \ldots, q$, $g = 1, 2, 3$, and "*the smaller, the better*" attributes, $\tilde{\tilde{\mathcal{Q}}}_\ell^{j(g)}$, $j \in J$, $\ell = q + 1, \ldots, Q$, $g = 1, 2, 3$. The range of the ℓth attribute in the observed dataset is:

$$\mathcal{V}_\ell^{(g)} = \max_{j \in J} \mathcal{Q}_\ell^{j(g)} - \min_{j \in J} \mathcal{Q}_\ell^{j(g)}, \quad \begin{matrix} \ell = 1, \ldots, Q, \\ g = 1, 2, 3, \end{matrix}$$

Consider the following normalisation (Girginer et al. 2015; Kuo et al. 2008):

$$\mathcal{Q}_\ell^{j(g)\star} = \begin{cases} \dfrac{\tilde{\mathcal{Q}}_\ell^{j(g)} - \min_{j \in J} \tilde{\mathcal{Q}}_\ell^{j(g)}}{\mathcal{V}_\ell^{(g)}}, \ell = 1, \ldots, q, \\ \dfrac{\max_{j \in J} \tilde{\mathcal{Q}}_\ell^{j(g)} - \tilde{\mathcal{Q}}_\ell^{j(g)}}{\mathcal{V}_\ell^{(g)}}, \ell = q + 1, \ldots, Q, \end{cases}, \quad j \in J, g = 1, 2, 3. \tag{4}$$

There is a reference associated with each attribute:

$$\mathcal{R}(\mathcal{Q}_\ell^{(g)}) = 1, \; \forall \ell = 1, \ldots, Q, \; g = 1, 2, 3.$$

The reference denotes the best possible quality-related level that can be achieved by hospitals in the corresponding attribute. The closer $\mathcal{Q}_\ell^{j(g)\star}$ is to the reference, the higher the quality of the jth hospital in the ℓth attribute of the kth attributes' cluster. Consider the quantities:

$$\mathcal{Q}_\ell^{j(g)\star\star} = |\mathcal{R}(\mathcal{Q}_\ell^{(g)}) - \mathcal{Q}_\ell^{j(g)\star}| \geq 0, \begin{cases} j \in J, \\ \ell = 1, \ldots, Q, \\ g = 1, 2, 3, \end{cases} \tag{5}$$

[1] Bandwidths can be estimated through distinct ways, including the least squares cross-validation procedure of Bădin et al. (2010) or the nearest-neighbour method proposed by [Daraio and Simar (2007a), p.109–110]. These two approaches produce local hospital-specific bandwidths. It is interesting to note that, in the current case, these two approaches lead to very similar results as the ones achieved via the Silverman's (global) approach. To keep it simple, we present only the results using the latter.

[2] $[a]$ is the nearest integer of a towards $-\infty$; e.g. $[2.95] = 2$.

[3] To avoid too large a manuscript, a document with the Appendix was stored in https://drive.google.com/drive/folders/1WICqxHmpUNESRqgjjZ2Yoe1dtvQ3ZYzt?usp=sharing, which is accessible to anyone.

$$\underline{Q^{(g)\star\star}} = \min_{j,\ell} Q_{\ell}^{j(g)\star\star},\ g = 1, 2, 3, \tag{6}$$

and

$$\overline{Q^{(g)\star\star}} = \max_{j,\ell} Q_{\ell}^{j(g)\star\star},\ g = 1, 2, 3, \tag{7}$$

and, finally, the grey relational coefficients

$$\Gamma^{j(g)} = \frac{1}{Q}\sum_{\ell=1}^{Q} \frac{\underline{Q^{(g)\star\star}} + \frac{1}{2}\cdot\overline{Q^{(g)\star\star}}}{Q_{\ell}^{j(g)\star\star} + \frac{1}{2}\cdot\overline{Q^{(g)\star\star}}},\ \begin{cases} j \in J, \\ g = 1, 2, 3, \end{cases} \tag{8}$$

and

$$\Gamma_{\ell}^{(g)} = \frac{1}{Q}\sum_{j=1}^{n} \frac{\underline{Q^{(g)\star\star}} + \frac{1}{2}\cdot\overline{Q^{(g)\star\star}}}{Q_{\ell}^{j(g)\star\star} + \frac{1}{2}\cdot\overline{Q^{(g)\star\star}}},\ \begin{cases} \ell = 1, \ldots, Q, \\ g = 1, 2, 3. \end{cases} \tag{9}$$

$\Gamma^{j(g)}$ is the level of relationship between the (normalised) empirical observation and the reference. The higher $\Gamma^{j(g)}$, the better the quality of the jth hospital in the corresponding set (either appropriateness and timeliness, safety, or access to care).

2.5 Double bootstrap

Past studies handling with the association between efficiency and quality in health care encompass some shortcomings limiting the extrapolation of their findings and the policy and managerial implications' validity. To estimate efficiency, most of them rely on linear programming techniques, particularly DEA. Varabyova et al. (2017) identify, at least, five different ways of including quality in benchmarking analyses:

(a) One-stage approach, where quality variables are included in DEA as intangible outputs; see Chang et al. (2011), Ferreira and Marques (2017), Garavaglia et al. (2011), and Navarro-Espigares and Torres (2011);
(b) Two-stage approach, where efficiency scores are estimated in the first stage and regressed against several predictors in the second ones; see Gok and Sezen (2013);
(c) Congestion analysis, so that undesirable quality variables can be assumed as weakly disposable outputs, see Dismuke and Sena (2001), Matranga and Sapienza (2015), and Valdmanis et al. (2008);
(d) Multiple objective DEA that assumes the maximisation of two objective functions (efficiency and quality), see Shimshak et al. (2009); and
(e) Conditional approach that imposes quality variables as exogenous to the production process, see Ferreira and Marques (2018), and Varabyova et al. (2016a).

Because of the *curse of dimensionality* exhibited by DEA and models alike, the inclusion of quality/access as additional outputs in one-stage approaches (including congestion analysis) may not be the best solution. Thus, using the one-stage approach to analyse the relationship between efficiency and quality through DEA estimates does not seem appealing.

Two-stage approaches are, indeed, better alternatives. Existing research adopting this strategy usually employs MRA in the second stage. Two problems can be found in the literature. First, efficiency scores are serially correlated (Hirschberg and Lloyd 2002), biasing the regression coefficients' estimates. Therefore, a bootstrap mechanism might be compulsory for multiple regressions (Simar and Wilson 2007). Second, those regressions rely on the separability condition (Daraio and Simar 2007a), meaning that predictors are not likely to affect the efficient frontier shape. Rather, they are expected to impact on efficiency distributions, only. Varabyova et al. (2016a) and Ferreira and Marques (2018) use the conditional approach as a potential alternative to overcome such a problem. The latter concluded that the separability condition holds for Portuguese public hospitals. It means that neither quality nor access seem to affect the efficient boundary. Moreover, it is difficult to use the conditional approach for MRA, especially whenever interactions should be accounted for. Therefore, we choose the bootstrapped based multivariate analysis, also known as double bootstrap, proposed by Simar and Wilson (2007), as one of the suitable alternatives to study the impact of quality and access on hospitals' efficiency distribution.

The standard double bootstrap model (model 2 of Simar and Wilson (2007)) was formulated to use radial DEA alongside the traditional bootstrap of Simar and Wilson (1998). This means that the first step of their model cannot correct technical efficiency for environmental effects. Moreover, the procedure does not allow the utilisation of nonradial directional models for efficiency assessment. To be consistent with the first stage of this study and the subsampling procedure, an adaptation of the original double bootstrap model (model 2) should be made. It can be found in Algorithm B (Appendix A, online). The algorithm regresses (conditional) slacks against the quality-related variables using the maximum likelihood method, and repeatedly draws a pseudo-residual with left truncation to estimate new slacks. These newly estimated slacks are used to *perturb* the original data (just like the original bootstrap algorithm does) and then to estimate bootstrapped efficiency scores. Once the efficiency scores space has been reduced to a Euclidean-space $\mathbb{R}_{+}^{n\times 1}$, one can employ (once more) the maximum likelihood method within a second stage to achieve ten sets of L_2 regression coefficients, useful to estimate confidence intervals (Simar and Wilson 2007). We used $L_1 = 200$ and $L_2 = 2{,}000$.

Once the bias- and environmentally-corrected efficiency estimate has been assessed in a first stage, they are regressed against a set of three quality/access composites in a second stage through a single truncated regression with bootstrap. In short, the double bootstrap achieves bias-corrected and confidence intervals for the coefficients β_0, β_g and $\beta_{gg'}$, g, $g' = 1, 2, 3$, of the truncated regression

$$\begin{aligned}(\theta^j)^{-1} &= \beta_0 + \sum_{g=1}^{3} \beta_g \cdot \Gamma^{j(g)} \\ &+ \xi \cdot \sum_{g=1}^{3} \sum_{g'=1}^{3} \beta_{gg'} \cdot \Gamma^{j(g)} \cdot \Gamma^{j(g')} \\ &+ \varepsilon^j, j \in J,\end{aligned} \tag{10}$$

where $\varepsilon \sim \mathcal{N}(0, 1)$ is an error term, and θ^j is the bias-corrected conditional efficiency estimate associated with the jth hospital. $\xi \in \{0, 1\}$ is a binary variable, which is equal to 0 if we do not consider interaction terms ($\Gamma^{(g)} \cdot \Gamma^{(g')}$, g, $g' = 1, 2, 3$); 1, otherwise. Note that $\beta_{gg'} = \beta_{g'g}$ for any g, $g' = 1, 2, 3$. Since the double bootstrap relates quality and the reciprocal of θ, negative coefficients β_g, $\beta_{gg'}$, g, $g' = 1, 2, 3$, indicate that both technical efficiency and quality (aggregated) attributes vary in the same sense. Double bootstrap obtains confidence intervals for coefficients β_0, β_g, and $\beta_{gg'}$, g, $g' = 1, 2, 3$ We assume the 5% significance level. Thus, if the 95% confidence intervals do contain the null value, *i.e.* $\{0\} \in CI(\beta_\star)$, then the associated coefficient is not statistically significant (different from 0).

3 Case study: efficiency and quality in Portuguese public hospitals

3.1 An overview of the Portuguese public hospitals

The Portuguese National Health Service (NHS) provides universal care to the population. There are three health care levels: primary (health centres), secondary (hospitals), and tertiary. The 2018 report of the World Health Organization[4] mentions that Portugal has 113 public hospitals (which belong to the NHS) and 96 private hospitals, including small clinics. These figures represent the number of infrastructures or buildings, not the Decision Making Unit (DMU) *per se* as it can be composed of more than one physical entity, as detailed below. From the sample of more than two hundred infrastructures, only 27 Portuguese public general hospitals operating between 2013 and 2016 can be considered. Private health care providers do not belong to the NHS and then do not deliver their data, publicly. Additionally, in 2013 there were 10,500 beds in private hospitals and 25 thousand in the public ones, which results roughly into one hundred beds per private provider, a quantity that doubles for public providers. This is because private entities are, in general, smaller than their public counterparts. In fact, only a few of them are comparable (in size) with a small/medium-size public hospital. Therefore, even if private providers would be included into the analysis, there would expectedly a problem of scale between both groups. The analysis reduces, thus, to the case of public hospitals.

The 113 public hospital facilities in Portugal comprise general (acute) hospitals, as well as specialized providers (maternities, oncology centres, rehabilitation centres,...). The latter were removed from the sample because their production function is arguably different from the one of the first group of hospitals. Since 2002/2003 the remaining infrastructures suffered some reforms carried out in the light of the New Public Management paradigm, aiming at exploiting both economies of scale and economies of scope. Reforms included the merging of hospitals: horizontal merging of two or more near facilities to create a hospital centre, and vertical merging of a hospital facility and several close primary care centres, creating a Local Health Unit (LHU). Although the number of buildings did not significantly change, the quantity of DMUs heavily reduced since 2002. The introduction of Public-Private Partnerships (PPPs) into the Portuguese NHS was also a strategy inspired on the aforementioned paradigm. Four PPPs were introduced to either replace outdated infrastructures (substitution hospitals) or to reinforce supply in regions where it was low.

In 2018, hospitals operating in the NHS included seven singular hospitals (or singletons), 20 hospital centres, eight LHUs, and the four PPPs. Data related to LHUs contain information of both primary and secondary levels of care; hence, they were removed from the sample to avoid incorrect benchmarking comparisons. Furthermore, PPPs are not forced to publicly deliver their data of inputs (namely, costs and number of staff). Because of missing data for these cases, PPPs were also erased from the dataset. Finally, based on such problems, the dataset concerns 27 Portuguese public general (acute) hospitals—20 hospital centres and seven singletons—operating between 2013 and 2016. In other words, the sample is composed of one hundred and eight DMUs-year ($=27 \times 4$).

Data pooling concerning the four years could introduce some bias sources if technology considerably changes over time (Ferreira and Marques 2015). In fact, between 2013 and 2016 a frontier shift (resulting from technological

[4] http://www.euro.who.int/__data/assets/pdf_file/0008/366983/portugalreview-report-eng.pdf.

progress or regress) may have occurred, which is not analysed here (nor it is the focus of the current research). To overcome that problem we slightly change Eq. (3) by introducing a new constraint:

$$t^k - \epsilon < t^j < t^k + \epsilon,$$

where $t^j \in \{2013, 2014, 2015, 2016\}$ is the year associated with the jth hospital and ϵ is a nonnegative scalar smaller than 1 (Ferreira et al. 2016), say $\epsilon = 0.50$. This change in Ω^k imposes that hospitals are compared only to their peers operating in the very same year.[5]

3.2 Data and variables

3.2.1 Data source(s)

Data concerning health care providers are made (freely) available by the official sources (Central Administration of the Health System).[6] Exogenous variables (demography and epidemiology) were collected from the Statistics Portugal.[7] Selected variables are defined below.

3.2.2 Inputs

Hospitals need to consume resources to deliver health care services to the population. The less waste they produce, the better their technical efficiency. Table 1 presents the adopted input variables. Following the literature, inputs (or resources) can be either physical or monetary. We solely considered time spent in inpatient services and costs associated with the whole hospital. Note that time (hospital days) can be considered a resource that need to be consumed for the healing process Ferreira et al. (2017a). Costs were adjusted by the Gross Domestic Product price deflator (basis: 2016). Inputs consumption is considerably heterogeneous among the Portuguese hospitals. On average, hospitals spent roughly €310 million annually, from which a meaningful share regarded cost inefficiency (Ferreira and Marques 2015, 2016a, 2016b).

The number of beds is used as input in several studies. We have initially considered this variable as an extra input, but it would reduce even more the sample (which is rather small) because of missing data for a considerable number of hospitals. However, considering only those observations with data on costs, hospital days, and beds, we computed the Pearson's correlation coefficient and observed that it

Table 1 Inputs

Input	Description	$\bar{x}$	σ_x	Range	References
Staff costs, x_1	Expenses with doctors, nurses, and other staff (either clinical or not), excluding outsourcing	€70.89 M	€55.52 M	€212.74 M	Ferreira and Marques (2015), Ferreira and Marques (2016a), Ferreira and Marques (2016b), Marques and Carvalho (2013), Simões and Marques (2011)
Operating costs, x_2	Operational costs excluding staff, outsourcing, and costs of goods sold and consumed	€147.50 M	€118.18 M	€430.27 M	Ferreira and Marques (2015), Ferreira and Marques (2016a), Ferreira and Marques (2016b), Khushalani and Ozcan (2017), Marques and Carvalho (2013), Simões and Marques (2011)
Costs of goods sold and consumed, x_3	Expenses with drugs and clinical material	€75.77 M	€78.86 M	€294.54 M	Ferreira and Marques (2015), Ferreira and Marques (2016a), Ferreira and Marques (2016b), Marques and Carvalho (2013), Simões and Marques (2011)
Costs with outsourcing, x_4	Costs with external external labour outsourcing	€16.52 M	€14.99 M	€60.39 M	Ferreira and Marques (2015)
Hospital days[a], x_5	Total number of days spent by all inpatients in nursery	177,243.95	114,367.99	484,488.00	Ferreira and Marques (2016a), Ferreira and Marques (2016b), Perez (1992)

M millions

[a]Time is a resource that should be consumed to improve inpatients' health status

[5] For instance, suppose that $t^k = 2015$, $k \in J = \{1, \dots, 108\}$. Then, $t^j \in T_{2015} =]2015 - 0.50, 2015 + 0.50[$. Years 2013, 2014, and 2016 do not belong to T_{2015}, thus samples operating in these years cannot act as potential peers for hospitals in t^k.

[6] http://benchmarking.acss.min-saude.pt/.

[7] https://ine.pt/xportal/xmain?xpgid=ine_main xpid=INE xlang=en.

was always larger than 0.95 ($p < 0.001$). That is to say that, either considering or not beds as input, it seems to be pointless in this case. As detailed before, we conduct an exploratory two-stage analysis based on regressions to relate efficiency with quality and access. Thus, small changes on efficiency estimates resulting from the inclusion of beds as input are not expected to promote variations on the coefficients of regressions.

3.2.3 Outputs

Hospitals production is another required measure for DEA. Table 2 defines eight outputs representing the production of the main hospital services: inpatient services, medical appointments, operating theatre, emergency room, and delivery room. Outputs were adjusted by the Case-Mix Index (CMI) (either internment, medical, or surgical CMI) so as to homogenise produced services. Note that hospital days can also be viewed as an output. This is a common assumption in the literature and the logic behind this is that it directly relates to the costs of care (Worthington 2004).

3.2.4 Demographics and epidemiology

It is well known that age structure impacts on costs and, eventually, on final outputs and quality. Indeed, older population is prone to more severe diseases that consume more resources and the in-hospital probability of death is much higher than the one for youngsters. Additionally, the education level and the purchasing power tend to have a positive effect on costs. There is evidence that education and purchasing power (resp. average age) are positively (resp. negatively) correlated. These (and other) issues were predicted by the Grossman model (Grossman 1972). Therefore, municipal and regional variables like population size, purchasing power, and population education level are considered to adjust both data and results for epidemiology and demographics, see Table 3. Note that such a large number of variables (13) may result on empty set Ω^k. In view of that, the Principal Component Analysis (PCA) technique was employed to aggregate them into three uncorrelated components, explaining more than 96% of the original data variance. These three components replaced the exogenous nondiscretionary variables in Eq. (3).

3.2.5 Appropriateness and timeliness of health care, patients' safety, and access to health care

We considered the outcomes presented in Table 4. Outcomes were clustered into three criteria: clinical safety, care appropriateness and timeliness, and access to health care. Safety, appropriateness and timeliness, and access are related concepts. On the one hand, the number of major surgeries on potentially minor operations may reflect lack of scientific knowledge and patient-centring. On the other hand, the higher the waiting time for a hip surgery, the higher the probability of complications, especially in the age group of elders. This means that a timely answer (surgery) reflects the quality of care. Finally, access to health care can be evaluated by the number of first visits and surgeries within the maximum ensured response time. It is worth mentioning that the maximum time is established by law and low rates result from poor access to secondary health care facilities. Note that both variables' selection and clustering have strictly followed the official database, which in turn reflects the State's own perception about appropriateness and timeliness, safety, and access.

3.2.6 Different models for robustness

It is well known that nonparametric benchmarking models, including DEA, are prone to problems related to dimensionality. Additionally, hospital days have been considered as inputs and as outputs (vide supra), requiring at least two different models. To account for this fact and to reduce the *curse of dimensionality effect*, we propose eight different models, M_l $l = 1, \ldots, 8$ see Table A in Appendix B. For instance, models M_1 and M_2 consider all variables as in Tables 1 and 2. The difference between them is the treatment given to hospital days). The remaining models use the total costs, Sum($x_1, \ldots, x_4$), to reduce the number of inputs. We also use PCA to narrow down some or all of the outputs into a single variable. Note that models M_7 and M_8 consider only a share of the presented outputs because some of them (such as births and surgeries) may represent subcategories of inpatient discharges. The difference between these two models is the aggregating level between output variables.

4 Results

This subsection provides the results achieved through GRA, DEA, and double bootstrap methods. We also include a bivariate regression analysis to strengthen the results. We developed all routines using the integration of MATLAB® programming software and the optimisation package IBM ILOG CPLEX®, v12.6.3.

4.1 GRA results

Clinical safety. No hospital has reached efficiency in safe events as $\max_{j \in J}\{\Gamma^{j(1)}\} = 0.89$. The small average, 0.66, may reveal a considerable lack of clinical safety in Portuguese public hospitals. No considerable heterogeneity was observed: the coefficient of variation of $\Gamma^{(1)}$ is nearly 16%. As expected, $\Gamma^{(1)}$ exhibits statistically significant and

Table 2 Outputs. Note: values are expressed in thousand

Output	Description	$\bar{y}$	σ_y	Range	References[a]
Inpatient discharges, y_1	Total number of treated inpatients in inpatient services	23.24	13.08	61.69	Chowdhury et al. (2014), Marques and Carvalho (2013), Simões and Marques (2011)
Emergency cases, y_2	Total number of emergencies	167.03	66.68	290.48	Khushalani and Ozcan (2017), Marques and Carvalho (2013), Simões and Marques (2011)
First medical appointments, y_3	Total number of first medical appointments	44.11	41.58	224.26	N.A./N.F.[a]
Follow-up scheduled medical appointments, y_4	Total number of follow-up scheduled medical appointments	235.34	171.27	656.25	
Outpatient surgeries, y_5	Total number of outpatient scheduled (minor) surgeries	8.67	5.35	19.39	N.A./N.F.[b]
Conventional surgeries, y_6	Total number of conventional scheduled (major) surgeries	6.45	4.54	16.40	
Urgent surgeries, y_7	Total number of urgent surgeries	2.93	1.98	7.99	
Births, y_8	Total number of births	1.88	1.08	4.95	Clement et al. (2008), Valdmanis et al. (2008)
Hospital days[c], y_9	Total number of days spent by all inpatients in nursery	177.24	114.37	484.49	Hollingsworth (2008), Rego et al. (2010)

N.A./N.F. not available or not found

[a]Several studies have used the number of medical appointments as outputs (see e.g. Khushalani and Ozcan (2017), Safdar et al. (2016)), but none has considered their classification into first and follow-up scheduled appointments

[b]As in the medical appointments case, most studies use the number of surgeries with no distinction between major and minor surgeries. Examples of studies using the overall number of surgeries as outputs include Caballer-Tarazona et al. (2010), Chang et al. (2004), Khushalani and Ozcan (2017), to name a few

[c]Treating hospital days as output is a conventional practice. The logic behind is that the variable is directly related to the costs of care (inputs)

Table 3 Exogenous nondiscretionary variables

Exogenous nondiscretionary variable[a]	Description	$\bar{z}$	σ_z	Range	References
Population size, z_1	Number of inhabitants targeted by the hospital	852,417.19	1,167,660.39	3,543,363.00	Guerrini et al. (2017)
Population density, z_2	Number of inhabitants per square kilometre	502.52	584.33	2,409.40	Allin et al. (2016), Ferreira and Marques (2015), Ferreira and Marques (2016a), Guerrini et al. (2017)
Elderly rate[b], z_3	Total number of elderly (>65 years old) per 100 inhabitants	20.65	3.87	14.17	Allin et al. (2016), Guerrini et al. (2017), Varabyova et al. (2016b)
Young rate, z_4	Total number of youths (<15 years old) per 100 inhabitants	14.07	1.56	6.37	N.A./N.F.; variables z_4 and z_5 complement z_3 in order to characterise the age structure of the target population.
Dependence index, z_5	Total number of elderly per 100 working age population	31.97	7.39	27.33	
Crude death rate[c], z_6	Total number of deaths per 1,000 inhabitants	10.23	2.51	9.40	Allin et al. (2016)
Child mortality rate, z_7	Total number of child (<1 year old) deaths per 1,000 childbirths	3.07	1.22	5.90	Hadad et al. (2013), Retzlaff-Roberts et al. (2004)
Crude birth rate, z_8	Total number of live births per 1,000 inhabitants	7.61	1.27	4.10	N.A./N.F.
Stillbirth rate, z_9	Total number of stillbirth cases per 100 child deaths	34.30	8.06	30.00	Hadad et al. (2013), Retzlaff-Roberts et al. (2004)
Illiteracy rate[d], z_{10}	Total number of illiterates per 100 inhabitants	5.46	2.24	8.50	Andrulis and Brach (2007), Fiscella et al. (2000), Sudore et al. (2006)
Inhabitants per doctor[e], z_{11}	Inhabitants per doctor	313.80	126.58	574.90	Bhat (2005), Hadad et al. (2013), Retzlaff-Roberts et al. (2004)
Inhabitants per pharmacist, z_{12}	Inhabitants per pharmacist	994.14	222.25	889.60	N.A./N.F.; z_{11} and z_{12} both characterise the access to the health care
Purchasing power *per capita*, z_{13}	Purchasing power *per capita* (national level: 100%)	94.67	13.27	53.60	Allin et al. (2016), Hadad et al. (2013), Varabyova et al. (2016b)

N.A./N.F. not available or not found

[a]Non-discretionary variables were collected from the official source (*vide* the Statistics Portugal website, https://www.ine.pt/). Most of the variables regard the years of 2014 and 2015, but in few cases data is slightly older. Still, it is assumed that the exogenous environment is stationary

[b]The age structure impacts on the health care resources consumption, as the higher the population average age, the more severe are the diseases

[c]Both death and birth rates can be associated with the access to health care facilities, appropriateness and timeliness and safety of provided services

[d]Education is highly related to the access to the health care system, as the more educated the population, the higher the prevention against diseases, the higher the purchasing power and the higher the access to the private health care sector

[e]The higher the number of doctors and pharmacists to attend for a population, the better the access to the health care system

Table 4 Appropriateness and timeliness, access, and safety variables. Values are presented in percentage (%)

	Outcomes	Direction[a]	Description	$\bar{x}$	σ_x	Range	References
Clinical safety	Decubitus ulcers, $\mathcal{Q}_1^{(1)}$	↙	Number of pressure ulcers in bedridden per 100 inpatients	1.50	1.09	4.50	Lyder et al. (2012), Sullivan and Schoelles (2013), Tsang et al. (2008)
	Pulmonary embolism/deep vein thrombosis, $\mathcal{Q}_2^{(1)}$	↙	Postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures	0.19	0.12	0.58	Tsang et al. (2008)
	Septis, $\mathcal{Q}_3^{(1)}$	↙	Postoperative septicaemia cases per 100 inpatients	0.69	0.60	2.81	Schang et al. (2016), Tsang et al. (2008)
	Trauma on vaginal delivery, $\mathcal{Q}_4^{(1)}$	↙	Cases of trauma on vaginal delivery (third and fourth degree lacerations), with/ without instrumentation, per 100 assisted deliveries	2.84	1.94	8.67	Tsang et al. (2008)
	In-hospital death rate, $\mathcal{Q}_5^{(1)}$	↙	In-hospital death cases for low severity levels (first two out of four severity levels) per 100 inpatients	2.84	1.94	8.67	Lindlbauer and Schreyögg (2014), Morey et al. (1992)
Care appropriateness and timeliness	Readmissions, $\mathcal{Q}_6^{(2)}$	↙	Number of readmissions within 30 days after discharge per 100 inpatients	8.69	1.65	8.57	Allin et al. (2016), Chowdhury and Zelenyuk (2016), Dahl and Kongstad (2017), Khushalani and Ozcan (2017)
	Outpatient surgeries on potential outpatient procedures, $\mathcal{Q}_7^{(2)}$	↗	Rate of outpatient surgeries on potential outpatient procedures	75.92	9.05	53.40	N.A./N.F.
	Inpatients staying more than 30 days, $\mathcal{Q}_8^{(2)}$	↙	Number of inpatients staying more than 30 days per 100 inpatients	3.19	1.05	4.46	N.A./N.F.
	Hip surgeries in the first 48 h, on elderly, $\mathcal{Q}_9^{(2)}$	↗	Rate of hip surgical procedures on elderly, in the first 48h	46.66	21.01	88.76	Bottle and Aylin (2006)
	Caesarean sections on total deliveries, $\mathcal{Q}_{10}^{(2)}$	↙	Rate of caesarean sections on total deliveries	27.03	8.57	40.00	N.A./N.F.; unless strictly necessary, C-sections should be avoided. The Portuguese government aims at reducing the caesarean section rates, by following the European Union guidelines. C-section rates should be below 15% of total deliveries. These outcomes are useful to check how far the hospitals are from the predefined national goal(s).
	Caesarean sections on TUCPs, $\mathcal{Q}_{11}^{(2)}$	↙	Rate of caesarean sections on TUCPs	17.25	10.62	35.00	
	First caesarean sections on TUCPs, $\mathcal{Q}_{12}^{(2)}$	↙	Rate of first caesarean sections on TUCPs	0.17	0.06	0.30	
	Vaginal deliveries after caesarean sections on TUCPs, $\mathcal{Q}_{13}^{(2)}$	↙	Rate of vaginal deliveries after caesarean sections on TUCPs	0.28	0.14	0.52	
Access	Rate of first medical appointments within time, $\mathcal{Q}_{14}^{(3)}$	↗	Rate of first medical appointments within the maximum ensured response time	73.54	12.67	49.10	N.A./N.F.; the Portuguese government uses both $\mathcal{Q}_{14}^{(3)}$ and $\mathcal{Q}_{15}^{(3)}$ to measure how hospitals answer to demand for care, within the appropriate and legislated time[b]
	Rate of surgeries within time, $\mathcal{Q}_{15}^{(3)}$	↗	Rate of surgeries within the maximum ensured response time	87.49	7.96	29.80	

N.A./N.F. not available or not found

[a] ↙ The lower, the better. ↗ The higher, the better. Directions strictly follow the official sources' indications, *vide* http://benchmarking.acss.min-saude.pt/

[b] See *Diário da República, series 1, no. 86, 4th May 2017* [in Portuguese]

negative correlation with clinical safety variables, being $R \in [-0.73, -0.25]$,*.[8] Negative correlations are expected because safety-based variables are defined as "*the lower, the better*", see Table 4 (third column). Thus, $\Gamma^{(1)}$ is a good proxy for clinical safety data. The analysis of (normalised) $\Gamma_\ell^{(1)}$ reveals that each safety variable contributes equally to the GRA coefficient, with scores ranging from 0.18 (in-hospital death rates for low severity levels) to 0.22 (septicaemia rates). Furthermore, if one takes the full-time equivalent doctors as a proxy for the hospital size, then the clinical safety proxy, $\Gamma^{(1)}$, and size are negatively correlated, $R = -0.34$,* which means that larger hospitals are more likely to observe complication episodes than the smaller ones. In fact, size (doctors) and complexity/severity of illness (inpatient services' CMI) are highly and positively related, $R = 0.87$,* since larger hospitals typically handle the most complex cases. By construction, safety and complexity of diseases are negatively correlated, $R = -0.36$,* that is an expected result as the former is not just the outcome resulting from medical care but also from case-by-case severity.

Appropriateness and timeliness of care. As in the case of clinical safety, there is not a single hospital providing the best possible quality of care. In this case, the average value of $\Gamma^{(2)}$ is 0.52 and its maximum is 0.77, values that are even lower than the ones achieved for clinical safety. Appropriateness and safety are positively correlated, $R = 0.46$,*. Some of the best performers in $\Gamma^{(1)}$ are also best performers in $\Gamma^{(2)}$. Interestingly, these units have small sizes, when measured by the full-time equivalent doctors. Indeed, size (resp. complexity) and quality of care are negatively correlated, $R = -0.38$ (resp. $R = -0.25$)*. In short, larger hospitals tend to provide care services with lower appropriateness, timeliness, and clinical safety.

Access to health care services. Access, $\Gamma^{(3)}$, is the quality dimension where hospitals exhibit the highest heterogeneity (coefficient of variation of 25.56% for a mean of 0.57). This is also the dimension where benchmarks present the best performance levels, ranging from 0.78 to 0.94. Access is highly correlated with both safety and appropriateness of care: $R = 0.52$ and $R = 0.48$,* respectively. That is, hospitals providing good quality of care and safe services are more likely to exhibit good access levels. However, access is neither related to the hospital size nor to the case-mix of patients. Both variables describing $\Gamma^{(3)}$ seem to have different impacts on this composite index, as stated by the two-sample t-test for equal means. In this case, the rate of surgeries within the maximum defined time is the variable that most impacts on $\Gamma^{(3)}$.

[8] R denotes the Pearson's correlation coefficient. The asterisk means that one rejects the null hypothesis of no correlation, at the common significance level of 5%.

4.2 DEA results

Evaluating the possible association between quality and efficiency requires a robust methodology to estimate the latter. Using an integrated approach of both nonradial DEA and subsampling (either conditional or not), we have achieved eight sets of $\mathcal{B} = 1500$ efficiency estimates per hospital. Each set corresponds to a single model, M_l, $l = 1, \ldots, 8$. Basic statistics of bias-corrected efficiency estimates are presented in Table B (Appendix B). It exhibits the results of both conditional and unconditional models. Looking at this table, no meaningful impact of the three PCA components of exogenous nondiscretionary dimensions on efficiency are found. In fact, Table C (Appendix B, online) compares those two models using three well known statistical tests (Kruskal–Wallis, two-sample Kolmogorov–Smirnov, and two-sample Student's t tests) applied to the possible $(1500)^2$ combinations of efficiency estimates.[9] According to this table, there is no evidence to reject the null hypotheses. In other words, the external nondiscretionary environment is not sufficient to explain hospitals' performance variance. Hence, the investigation of the trade-offs between efficiency, and process quality and outcomes can be done using any of these estimates as one may expect that they have the same behaviour.

Back to Table B of Appendix B and considering the conditional efficiency estimates, we observe a considerable heterogeneity among the average bias-corrected efficiency estimates returned by each model. As expected, more parsimonious models, such as M_5 and M_8, provide smaller efficiency levels than models using more variables, as M_1 and M_2. For instance, the upper bound of the 95% confidence interval associated with the average efficiency score of model M_5 is smaller than the lower bound of the interval for M_2. It means that one may expect statistically significant differences among these two models. To confirm it, we have compared the eight models using the Kruskal–Wallis and the two-sample Kolmogorov–Smirnov tests, and getting the probability of observing p values above 5%; see Table D (Appendix B, online).[10] According to this table (matrix), models M_1 and M_2, M_3 and M_4, and M_5 and M_8 are equivalent. The comparability between M_1 (or M_2) and M_3 (or M_4) is weak/unclear as both tests returned distinct

[9] If the probability of getting p values above the significance level of 5%, Prob ($p > 0.05$), is large (say, larger than 90%), then one may conclude that no substantial evidence exists to reject the null hypothesis associated with each one of the three statistical tests. In this case, not rejecting the null hypothesis implies not rejecting the hypothesis that conditional and unconditional efficiency estimates are statistically equal.

[10] The smaller the value of each matrix's entry, the larger the probability of rejecting the null hypothesis and of concluding that two models deliver different efficiency estimates. Naturally, the diagonal of the matrix is composed of 100% entries, only.

outcomes. Model M_6 seems to be equivalent to M_3 and M_4 according to the Kruskal–Wallis test, but not according to the two-sample Kolmogorov–Smirnov test; thus, their comparability is, at least, questionable. The comparability of M_7 and the remaining models is also both weak and non compatible. For these reasons, the following discussion is based on models M_2, M_3, M_6, M_7, and M_8.

Because M_1 and M_2 are equivalent (as it also happens to M_3 and M_4), we firstly conclude that considering hospital days either as an input or an output variable seems to be pointless, at least in the present case. Additionally, we observe that M_5 and M_8 are equivalent. The differences between these two models are the hospital days as input/output variable and the number of variables used to construct the PCA aggregate. Given the fact that considering the hospital days either as an extra input or an extra output is meaningless, then considering subcategories of inpatient discharges, such as births and surgeries, appears to be redundant as well. Indeed, let $y' = PCA\ (y_1,\ \ldots,\ y_8)$ and $y'' = PCA\ (y_1,\ \ldots,\ y_4)$; the Pearson's correlation coefficient is $\mathrm{corr}(y',\ y'') = 0.9999$*.

4.3 Double bootstrap results: a bivariate analysis

In this subsection, technical efficiency and the different GRA-based quality indicators are regressed using the double bootstrap technique to verify if there is a simple linear relationship (association) between those dimensions of hospital performance.

Figure 1 portrays the dispersion diagrams of bias-corrected efficiency estimates against GRA quality aggregates for two models: M_2 and M_8. Both efficiency and quality aggregates were rescaled by the standard deviation and recentred on average. Hence, we can identify four main regions of performance: (R1) poor quality and poor efficiency, (R2) good quality and poor efficiency, (R3) poor quality and high efficiency, and (R4) good quality and high efficiency.

Additionally, Table E (Appendix B, online) provides the estimates and some statistical tests applied to residuals resulting from the linear regression. Residuals must obey to three conditions: normality, absence of autocorrelation, and homoskedasticity. To evaluate if residuals follow the Gaussian distribution, $\mathcal{N}(0, \hat{\sigma}_\varepsilon)$, we used the one-sample Kolmogorov–Smirnov test after residuals' rescaling by $1/\hat{\sigma}_\varepsilon$. The Durbin–Watson test evaluates the null hypothesis of uncorrelated residuals from the linear regression, against the alternative hypothesis of autocorrelation among them. Finally, to test the residuals' homoskedasticity we used the Spearman's ρ associated with the pair $(\Gamma^{j(g)}, |e^j|)$. If ρ is statistically not different from 0, then residuals are likely homoskedastic. To test whether the simple linear model is a good explainer of the relationship between efficiency and quality, the R^2 should be high, and the probability of observing p values associated with the three previously described tests above 5% should be large (ideally, close to 100%).

Clinical safety vs technical efficiency. According to Fig. 1 and to the model M_2, there are technically efficient hospitals delivering unsafe care to their patients but there are also simultaneously efficient and safe units. However, M_2 may suffer from the so-called curse of dimensionality affecting DEA based models. In fact, a meaningful number of hospitals considered efficient by M_2 are inefficient under M_8. Following this parsimonious model, most of the technically efficient hospitals belong to the region (R4)—good quality and high efficiency. Also, the dispersion diagram exhibits a positive trend for both θ and $\Gamma^{(1)}$: when one performance dimension grows, the other one is expected to increase as well. According to Table E (Appendix B, online), coefficients β_1 associated with $\Gamma^{(1)}$ are consistent among the different models and are always negative and statistical significant at the 5% level. However, the linear relationship between those two dimensions is, in general, weak, especially for models with a considerable quantity of variables. For instance, the simple linear model is unable to explain the relationship between $1/\theta$ (as assessed via M_2) and $\Gamma^{(1)}$. The goodness-of-fit is better in model M_8, but the homoscedasticity assumption over residuals is often violated. In other words, we cannot use the simple linear model to explain the relationship between technical efficiency and clinical safety in hospitals. Perhaps, some interaction terms have to be considered. Still, we may argue that it is possible to provide safe care services with a minimal waste of resources.

Appropriateness and timeliness vs technical efficiency. The relationship between appropriateness of care and the technical efficiency is even harder to discern than before. There is a greater dispersion in plots of Fig. 1, although a positive trend can be observed from M_8. That is, one may expect that when the performance of hospitals improve in terms of appropriateness and timeliness of care, the technical efficiency is also likely to improve. Most of the efficient hospitals present good performance levels in terms of $\Gamma^{(2)}$. Nevertheless, only 23% (at the most) of the variance of $1/\theta$ can be explained by the simple linear model, although residuals obey the three conditions underlying the regression model. Again, some additional dimensions, including interactions, should be considered to explain the evolution of hospital efficiency through quality of care.

Access to health care services vs technical efficiency. Access to health care services can be measured by both the quantity of first medical appointments and surgeries within the maximum guaranteed response time regarding the quantity of patients requiring care delivery. These two dimensions are, per se, measures of hospital productivity, as

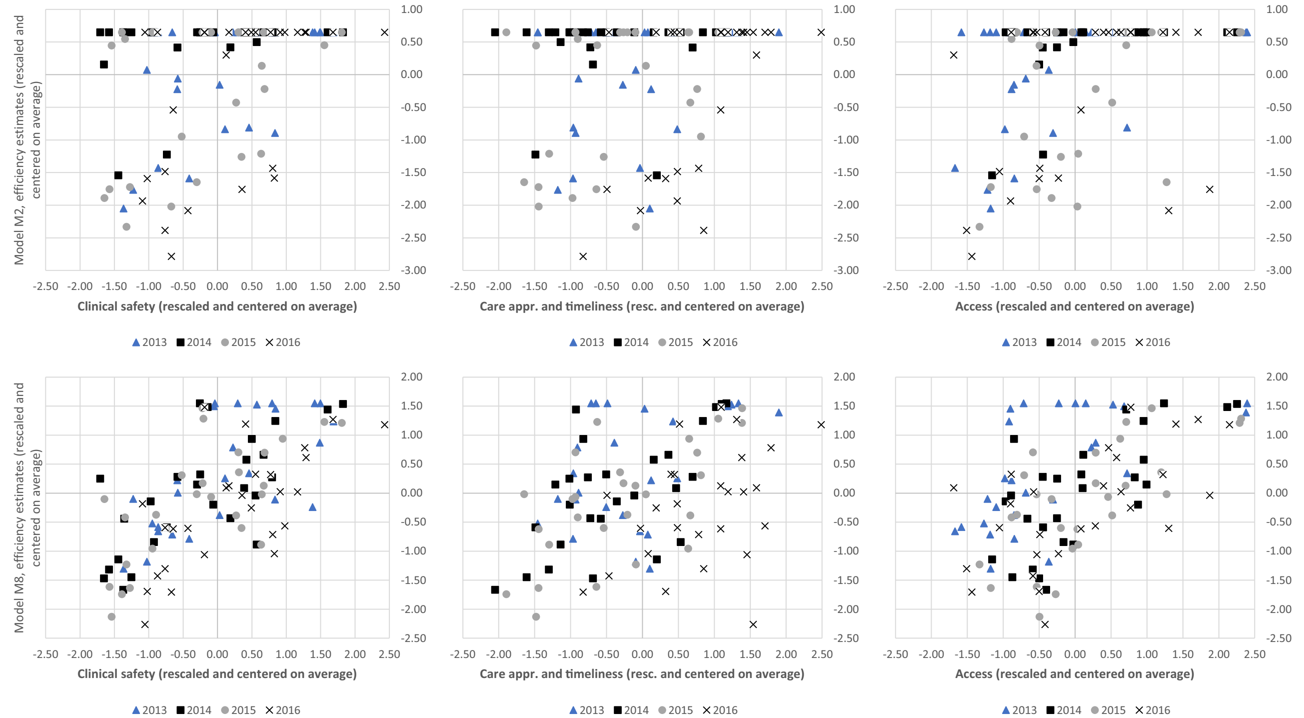

Fig. 1 Dispersion diagrams of bias-corrected efficiency estimates against GRA quality aggregates: models M_2 (top) and M_8 (bottom)

they measure how fast the entity answers to health care demand. The trend observed in diagrams of Fig. 1 for the model M_8 shows a positive relationship between efficiency and access, as it happens to the other two quality dimensions. The dispersion, in this case, lies within the other two cases, implying that the coefficient of determination associated with the simple linear model is smaller than 0.2436 but larger than 0.1189, *i.e.*, the expected values for the goodness of fit of the first and the second simple linear models. Yet, as we can suspect from the dispersion diagram, residuals resulting from this model are likely heteroskedastic, as it happens to the linear model associated with the clinical safety. Based on these results, we conclude that the bivariate analysis does not seem adequate to investigate the association between efficiency and the three dimensions of quality. A multivariate model, Eq. (10) with $\xi = 1$, is probably a more appropriate approach as it allows the inclusion of interaction terms.

4.4 Double bootstrap results: a multivariate approach with and without interactions

The (adapted) double bootstrap of Simar and Wilson (2007), *vide* Algorithm 2 in appendix, is used in the second-stage to investigate whether quality-related variables are related or not to the technical efficiency distribution. In the real world, quality dimensions do not act in isolation, thus we employ a multivariate linear model with and without interaction terms, to verify whether they play a meaningful role on efficiency.

Table 5 provides the estimates and statistical tests resulting from the multivariate linear regression models with $\xi = 0$ and $\xi = 1$. As before, we recentred and rescaled the explanatory variables through their averages and standard deviations.[11]

Admitting no interaction terms ($\xi = 0$), the three quality aggregated variables can explain at the most half of the variation of hospitals' technical efficiency. This happens with model M_8. It is not necessarily true that fewer variables will return better linear models.[12] Likewise, models with a considerable number of inputs/outputs, such as M_2 and M_3, also have poor goodness of fit. Even though M_8 has a substantial coefficient of determination, a meaningful share of efficiency estimates result in autocorrelated residuals when those estimates are regressed using the multivariate linear model (10) with $\xi = 0$, which means that the linear model without interaction terms could not be the most

[11] Because models are multivariate, the coefficient of determination should be adjusted using Adjusted $R^2 = 1 - (1 - R^2) \cdot (n-1)/(n-\eta-1)$, where $n = 108$ is the sample size and η is the number of predictors (explanatory variables). Thus, $\eta = 3$ if $\xi = 0$ and $\eta = 9$ if $\xi = 1$.

[12] After all, M_6 considers a single input and a single output and the goodness of fit of its associated multivariate linear model is poor (the fraction of variance unexplained is, at least, 86%).

Table 5 Multivariate linear regression: estimates and statistical tests

	M_2			M_3			M_6			M_7			M_8		
	Estimate	LB[a]	UB[b]	Estimate	LB	UB	Estimate	LB	UB	Estimate	LB	UB	Estimate	LB	UB
$(1/\theta^j) = \beta_0 + \sum_{g=1}^{3} \beta_g \cdot \Gamma^{j(g)} + e^j$, $(\xi = 0)$															
β_0	1.1048	1.0774	1.1331	1.1891	1.1643	1.2201	1.1365	1.1265	1.1468	1.2261	1.2110	1.2411	1.3101	1.2880	1.3326
β_1	−0.0456	−0.0700	−0.0217	−0.0859	−0.1165	−0.0568	−0.0134	−0.0217	−0.0063	−0.0569	−0.0708	−0.0440	−0.1147	−0.1404	−0.0884
β_2	0.0042	−0.0235	0.0334	0.0013	−0.0334	0.0338	−0.0090	−0.0194	0.0014	−0.0306	−0.0471	−0.0140	−0.0221	−0.0434	−0.0002
β_3	−0.0384	−0.0691	−0.0090	−0.0614	−0.0871	−0.0358	−0.0286	−0.0369	−0.0209	−0.0217	−0.0345	−0.0070	−0.0608	−0.0761	−0.0452
R^2	0.1319	0.0739	0.2040	0.3119	0.2291	0.4503	0.1168	0.0798	0.1651	0.2600	0.1945	0.3263	0.4686	0.4090	0.5280
Adjusted R^2 [c]	0.1069	0.0472	0.1810	0.2920	0.2069	0.4344	0.0913	0.0533	0.1411	0.2386	0.1713	0.3069	0.4533	0.3920	0.5144
F-statistic[d]	97	100	99	100	100										
KS test[e]	0	97	96	100	100										
DW test[f]	99	98	98	100	78										
$(1/\theta^j) = \beta_0 + \sum_{g=1}^{3} \beta_g \cdot \Gamma^{j(g)} + \sum_{g=1}^{3} \sum_{g'=1}^{3} \beta_{gg'} \cdot \Gamma^{j(g)} \cdot \Gamma^{j(g')} + e^j, (\xi = 1)$															
β_0	1.1118	1.0846	1.1400	1.1897	1.1593	1.2240	1.1369	1.1265	1.1469	1.2209	1.2052	1.2366	1.3154	1.2927	1.3406
β_1	−0.1623	−0.7773	0.4300	−1.0812	−1.8106	−0.2567	−0.1164	−0.3707	0.0780	0.1740	−0.2131	0.5336	−1.1493	−2.1083	−0.1670
β_2	0.6394	−0.0493	1.4212	0.3583	−0.8871	1.4771	0.5788	0.2426	0.8618	0.6247	0.1564	1.0622	−0.7437	−1.5869	0.1090
β_3	−0.1438	−0.7987	0.3428	0.4560	−0.2601	0.9883	0.0984	−0.1184	0.3013	0.1034	−0.2178	0.3741	0.4396	−0.1172	0.9998
β_{11}	−0.0559	−0.7374	0.5653	0.8212	−0.0559	1.5713	−0.3929	−0.6454	−0.1215	−1.0065	−1.4565	−0.5880	0.4747	−0.4259	1.3782
β_{12}	−0.3345	−0.9588	0.3108	0.0915	−0.9786	1.3367	0.4887	0.2691	0.7441	0.6712	0.2736	1.1082	0.5526	−0.2936	1.2906
β_{13}	0.9217	0.5401	1.2876	0.2420	−0.2462	0.8027	0.3846	0.2401	0.5210	0.7574	0.4388	1.0973	0.4103	−0.0700	0.8984
β_{22}	0.1044	−0.9286	0.9936	0.2121	−1.2326	1.5805	−0.6834	−1.0125	−0.3546	−0.6509	−1.1943	−0.1353	0.7363	−0.2958	1.8647
β_{23}	−1.6005	−2.5310	−0.4752	−1.7132	−2.2524	−1.1044	−0.4044	−0.5806	−0.1980	−0.9224	−1.1882	−0.6331	−0.7938	−1.2524	−0.4482
β_{33}	0.7470	0.2299	1.2345	0.7081	0.4284	0.9831	−0.0693	−0.1828	0.0484	0.0861	−0.0839	0.2499	−0.1539	−0.3656	0.1071
R^2	0.3221	0.2055	0.4413	0.4825	0.3652	0.6158	0.1846	0.1452	0.2322	0.3464	0.2712	0.4401	0.5807	0.4917	0.6636
Adjusted R^2	0.2598	0.1326	0.3900	0.4350	0.3069	0.5805	0.1098	0.0666	0.1617	0.2864	0.2043	0.3887	0.5421	0.4450	0.6327
F-statistic	100	100	83	100	100										
KS test	8	95	99	100	100										
DW test	99	97	100	99	98										

[a] Lower bound of the 95% confidence interval

[b] Upper bound of the 95% confidence interval

[c] Adjusted $R^2 = 1 - (1 - R^2) \cdot (n - 1)/(n - \eta - 1)$, where n is the sample size ($n = 108$) and η is the total number of explanatory variables in the model (excluding the constant term)

[d] This row represents the probability of the F-statistic's p value is smaller than the significance level of 5%. If this probability is large, then one may expect that a significant linear regression relationship exists between the response (the reciprocal of the efficiency estimates) and the predictor variables (GRA aggregates of quality)

[e] This row exhibits the probability of the residuals of the linear regression model following the standard Gaussian distribution. We use the one-sample Kolmogorov–Smirnov test and the null hypothesis is H_0: residuals come from a standard normal distribution. If the probability Prob($p > 0.05$) is large, then residuals from the estimated linear regression are likely normal, with null average and unitary standard deviation

[f] The Durbin–Watson test of the linear model evaluates the null hypothesis H_0: residuals are uncorrelated, against the alternative H_1: there is autocorrelation among residuals. This row shows the probability Prob($p > 0.05$) for the Durbin–Watson test. Hence, large values of this probability indicate that residuals' autocorrelation is unlikely

appropriate one to explain the association between efficiency and quality.

When interaction terms, $\Gamma^{(g)} \cdot \Gamma^{(g')}$, $g, g' = 1, 2, 3$, are accounted for ($\xi = 1$), the goodness of fit of models increase considerably for most of the considered models. For instance, the fraction of efficiency variance unexplained by quality decays to 45% in model M_8. Likewise, considering M_3, that fraction goes from 71 ($\xi = 0$) to 57% ($\xi = 1$). In the best scenario for these two models, the nine explanatory variables used in the regression could explain 63 and 58% of the technical efficiency variance. They are, then, the best two multivariate linear models from the considered eight. We remark that the conditions over residuals are almost always fulfilled in these two regressions. Those models, M_3 and M_8, are consistent in two aspects: (a) coefficients associated with $\Gamma^{(2)}$, $\Gamma^{(3)}$, $\Gamma^{(1)} \cdot \Gamma^{(1)}$, $\Gamma^{(1)} \cdot \Gamma^{(2)}$, $\Gamma^{(1)} \cdot \Gamma^{(3)}$, and $\Gamma^{(2)} \cdot \Gamma^{(2)}$ are not statistically different from 0; and (b) coefficients associated with $\Gamma^{(1)}$ and $\Gamma^{(2)} \cdot \Gamma^{(3)}$ are statistically different from 0 and are negative. This means that the technical efficiency of hospitals is positively associated with the clinical safety and with the interaction between the appropriateness, timeliness, and access to health care. Thus, the effect of care appropriateness and timeliness on efficiency depends on access levels and vice-versa. In the light of statistical evidence, Hypotheses 1 and 4 cannot be rejected.

Let us consider the results associated with M_8 and an efficient hospital k. Suppose that k decreases $\Gamma^{k(1)}$ from 0.87 to 0.835 (keeping the appropriateness, the timeliness, and the access levels unchanged). As $\hat{\sigma}_1 \approx 0.07$, the worsening of $\Gamma^{k(1)}$ corresponds to a decrease of $\partial\Gamma^{k(1)} = \hat{\sigma}_1/2$. Since $\partial(1/\theta^k) \approx \beta_1 \cdot \partial\Gamma^{k(1)} = -1.1493/2$, the resulting efficiency worsening is about 36%, *i.e.*, the expected final efficiency score of k is $\theta^k = 0.64$. Now, let us suppose that the hospital keeps the patients' clinical safety unchanged, but the appropriateness level decreases by 0.05 ($= \hat{\sigma}_2$). In such a case, the final efficiency estimate will be $\theta^k = 1/(1 + \beta_{23} \cdot \Gamma^{k(3)})$. That is, the final efficiency estimate will depend on the performance of hospital k in terms of access to health care. Let $\Gamma^{k(3)} = 0.75$; hence, $\theta^k = 0.63$. If the access level would be smaller, say $\Gamma^{k(3)} = 0.50$, then $\theta^k = 0.72$, which means that the worse the performance of the hospital in terms of access, the smaller the impact of appropriateness reduction on efficiency. Since $\Gamma^{(2)}$ and $\Gamma^{(3)}$ are both strictly positive and β_{23} is statically significant and negative, then efficiency is expected to improve following improvements on appropriateness, timeliness, and access to health care services. Therefore, in the light of statistical evidence, we cannot reject Hypotheses 2 and 3.

This conjoint effect of appropriateness, timeliness, and access on efficiency makes sense because some of the variables adopted to describe appropriateness could also be good proxies for access, *e.g.*, rate of hip surgical procedures on elderly in the first 48 h (*i.e.*, within the appropriate time). For instance, according to Ferreira and Marques (2018), timeliness of services can be defined as the "*capacity of delivering health care services at the right time, whenever required and appropriate*"; thus, it can be a measure of access. Providing adequate care at the optimal access, preferably through linkage between primary and secondary health care, allows efficiency improvements. This is possible because the patient is more closely followed and the severity of the disease reduces, avoiding complications (surgeries outside the recommended time) as well as the incidence of chronic diseases responsible for large expenditures in the public health service. Another situation that contributes to the improvement of efficiency is compliance with the Maximum Guaranteed Response Times (clinically defined according to priority and severity), which, when overcome, call into question efficiency because there are complications and aggravation of a situation that could be previously solved with fewer resources.

5 Discussion

Based on econometric and statistical methods applied to Portuguese public hospitals, we can drawn some important empirical results on the efficiency-quality-access link. Evidence seems to suggest that the improvement of technical efficiency should be followed or is expected to follow enhancements in terms of quality and/or access to health care. Specifically, poor efficiency is associated with poor quality of provided services, in line with Clement et al. (2008) and Mobley and Magnussen (2002), at the same time that best practices can deliver safe, appropriate and timely services to their patients, as concluded by Arocena and García-Prado (2007) and Helling et al. (2006). In opposition, the results seem to refute the findings of Morey et al. (1992), Singaroyan et al. (2006), and Valdmanis et al. (2008). Moreover, the U-shaped relationship between quality and efficiency, as proposed by Hvenegaard et al. (2011) was not corroborated by the empirical evidence of our study. In fact, based on the bivariate analysis plots, improvements in quality and access do not necessarily imply efficiency deterioration, *i.e.* it is possible to improve quality of care with no efficiency sacrifice, in line with Chang et al. (2011), Ferrando et al. (2005), and Nayar and Ozcan (2008).

Clinical safety, appropriateness and timeliness of care, and access to health care services have shown to be significant drivers of efficiency. Appropriate and timely care services are, then, a priori more efficient, which is an expected finding. The higher the quantity of readmitted inpatients, the lower the technical efficiency, because the former are accounted only once (as single patients) for a larger amount of consumed resources. The same occurs for inpatients staying more than 30 days in hospital wards, as they are occupying one bed that could be available for other inpatients requiring care delivering, and thus consuming more resources than the appropriate without an increase of the volume of services (measured by the number of treated inpatients). Likewise, outpatient surgeries are usually cheaper than major surgeries, contributing to hospital cost containment and efficiency improvement (lower expenses and higher volume of treated patients). If outpatients are unnecessarily subject to major surgeries, then the hospital loses resources that could be better allocated in other priority areas. As appropriateness and timeliness, the better the access to health care services, the higher their technical efficiency. In the current study, access is measured by the amount of patients undergoing surgery or seen in first medical appointments, within the maximum ensured response time. These variables also reflect how fast the hospital is at answering to health care. The faster the response, the higher the quantity of treated patients (shorter waiting lists) and, keeping the resources nearly unchanged, the higher the technical efficiency. According to the GRA coefficients, the rate of surgeries within the maximum time defined by law seems to be the variable that contributes the most for this fact. Finally, patients' clinical safety is also associated with the hospitals' technical efficiency. Thus, hospitals that are inefficient on their resources management are prone to observe adverse clinical events (care complications), such as septicaemia.

Previous empirical evidence meet the main goals of public policies in the Portuguese health sector, and show the organisational best practices over caregivers that have been introduced aiming at improving efficiency, access, and quality of health care. In Portugal, the features of payment contracts have been reviewed in the past few years. In 2016, new payment and hospital activity contracting ways are expected to be one of the main drivers of the timely, effective, and efficient delivery of care services. New health policies introduced in 2016 put the patient and her/his family in the core of the National Health Service, and encourage good clinical and health governance. Contracts between hospitals (as corporate public entity, see Ferreira and Marques (2015)) and the Ministry of Health are characterised by a considerable number of features, including benchmark-based performance incentives for best practices, as well as penalties for contract breaching and medical and nursing care misconducts. These facts are likely to justify (at least, partially) the results found in prior subsections. Improving efficiency, effectiveness, and quality of health care will seemingly improve the population's health status (which becomes more productive) and positively impacts on health care providers management, because it avoids the hospital stays extension due to complications, reduces hospital septicaemia and blood infection cases, as well as the after-effects, the incidence of adverse events, the emergency demand, and the readmissions to hospital wards. Last but not the least, the introduction of both Integrated Responsibility Centres and Reference Centres within hospitals can also be responsible for the improvement of quality, access, and efficiency. Integrated Responsibility Centres, for instance, manage hospital services' resources via internal contracting mechanisms. This is an importance aspect, especially in highly differentiated hospitals, as it allows a timely and integrated response to care demand, decreasing waiting time and improving the profitability of assets.

Some limitations of our study can be identified. The GRA analysis, for instance, seems to be technically appealing as it is simple to compute, use and interpret; still, it may limit the application of DEA at some extent. By aggregating multiple indicators into quality-composites, it may be unclear what *quality* is in fact, and how managers can improve it in practice. Nonetheless, one must keep in mind that our current objective is to analyse the relationship (or trade-offs) between technical and quality within one sector playing a prominent role on society. The construction of quality-composites avoids the dimensionality problem in nonparametric models. The same can be said of using PCA to aggregate data with no meaningful loss of information. This is a desirable alternative given the substantial number of adopted variables. However, we remark that PCA does not distinguish between variance due to measurement noise or due to underlying signal variations. Thus, if data are noisy, PCA may represent the noise instead of data, which is another limitation of this study.

Another shortcoming is the reliance on the separability condition underlying the double-bootstrap. Alternatives could be the methods based on partial frontiers (order-m and order-α) Aragon et al. (2005) that easily handle with the current problematic. Nonetheless, they depart from the assumption that inputs contraction/ outputs expansion have to occur at the same rate, contrasting with the prerequisite that models should be nonradial.

6 Conclusions

This study discussed the potential link between quality and efficiency on secondary health care services. On the one hand, one may think that quality improvements require considerable amounts of investments, which may jeopardise the health care systems financial sustainability. On the other hand, hospital managers usually want to know how far they can improve the quality of provided services with no sacrifice of efficiency, or, in opposition, how they can constrain resources and increase the health care delivery services without quality deterioration. Although this topic has been previously discussed in a vast literature, there are some shortcomings limiting the validity of these findings. We sought to contribute to the literature by using econometric methods and multicriteria decision analysis tools, and then offering consistent and statistically solid results. Accordingly, efficiency and quality seem to evolve in the same direction, i.e. whenever one dimension is improved, the other is expected to improve as well. By way of explanation, investments on quality are not likely to negatively affect efficiency, on the contrary. Quality improvements are shown to increase technical efficiency.

Acknowledgements We would like to express their gratitude to three anonymous referees and the Editor (Prof. Victor Podinovski) who kindly and significantly have improved this paper's quality, clarity, and structure, due to their beneficial comments. The contents of this paper are our own responsibility. The usual disclaimer applies.

Funding This research was generously supported by the project 'hSNS: Portuguese public hospital performance assessment using a multi-criteria decision analysis framework' (PTDC/EGE-OGE/30546/2017), funded by the Portuguese Foundation for Science and Technology (Grant No. 02/SAICT/2017/30546). The first author also acknowledges the financial support by the same institution (Grant SFRH/BD/113038/2015).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Allin S, Grignon M, Wang L (2016) The determinants of efficiency in the canadian health care system. Health Econ Policy Law 11 (1):39–65

Andrulis DP, Brach C (2007) Integrating literacy, culture, and language to improve health care quality for diverse populations. Am J Health Behav 31(1):770–776

Aragon Y, Daouia A, Thomas-Agnan C (2005) Nonparametric frontier estimation: a conditional quantile based approach. Econom Theory 21:358–389

Arocena P, García-Prado A (2007) Accounting for quality in the measurement of hospital performance: evidence from Costa Rica. Health Econ 16(7):667–685

Bădin L, Daraio C, Simar L (2010) Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. Eur J Oper Res 201(2):633–640

Bădin L, Daraio C, Simar L (2012) How to measure the impact of environmental factors in a nonparametric production model. Eur J Oper Res 223(3):818–833

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092

Bhat VN (2005) Institutional arrangements and efficiency of health care delivery systems. Eur J Health Econ 6(3):215–222

Bottle A, Aylin P (2006) Mortality associated with delay in operation after hip fracture: observational study. BMJ 332(7547):947–951

Caballer-Tarazona M, Moya-Clemente I, Vivas-Consuelo D, Barrachina-Martinez I (2010) A model to measure the efficiency of hospital performance. Math Comput Model 52 (7–8):1095–1102

Chambers RG, Chung Y, Färe R (1996) Benefit and distance functions. J Econ Theory 70(2):407–419

Chambers RG, Chung Y, Färe R (1998) Profit, directional distance functions, and nerlovian efficiency. J Optim Theory Appl 98 (2):351–364

Chang H, Cheng MA, Das S (2004) Hospital ownership and operating efficiency: Evidence from taiwan. Eur J Oper Res 159 (2):513–527

Chang SJ, Hsiao HC, Huang LH, Chang H (2011) Taiwan quality indicator project and hospital productivity growth. Omega Int J Manag Sci 39(1):14–22

Charnes A, Cooper WW, Rhodes E (1979) Measuring the efficiency of decision-making units. Eur J Oper Res 3(4):201–209

Chowdhury H, Zelenyuk V (2016) Performance of hospital services in Ontario: DEA with truncated regression approach. Omega Int J Manag Sci 63:111–122

Chowdhury H, Zelenyuk V, Laporte A, Wodchis WP (2014) Analysis of productivity, efficiency and technological changes in hospital services in Ontario: How does case-mix matter? Int J Prod Econ 150:74–82

Clement JP, Valdmanis VG, Bazzoli GJ, Zhao M, Chukmaitov A (2008) Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion. Health Care Manag Sci 11(1):67–77

Dahl CM, Kongstad LP (2017) The costs of acute readmissions to a different hospital: does the effect vary across provider types? Soc Sci Med 183:116–125

Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Product Anal 24(1):93–121

Daraio C, Simar L (2007a) Advanced robust and nonparametric methods in efficiency analysis. Springer Science + Business Media, Inc., New York, USA

Daraio C, Simar L (2007b) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. J Product Anal 28(1–2):13–32

Deng J (1982) Control problems of grey systems. Syst Control Lett 1:288–294

Dervaux B, Leleu H, Minvielle E, Valdmanis V, Aegerter P, Guidet B (2009) Performance of French intensive care units: a directional distance function approach at the patient level. Int J Prod Econ 120:585–594

Dismuke CE, Sena V (2001) Is there a trade-off between quality and productivity? The case of diagnostic technologies in Portugal. Ann Oper Res 107:101–116

Donabedian A (1988) The quality of care: how can it be assessed? J Am Med Assoc 260(12):1743–1748

Donabedian A (2005) Evaluating the quality of medical care. Milbank Q 83(4):691–729

Ferrando A, Ivaldi C, Buttiglieri A, Pagano E, Bonetto C, Arione R, Scaglione L, Gelormino E, Merletti F, Ciccone G (2005) Guidelines for preoperative assessment: Impact on clinical practice and costs. Int J Qual Health Care 17(4):323–329

Ferreira DC, Marques RC (2015) Did the corporatization of Portuguese hospitals significantly change their productivity? Eur J Health Econ 16(3):289–303

Ferreira DC, Marques RC (2016a) Identifying congestion levels, sources and determinants on intensive care units: the Portuguese case. Health Care Manag Sci 21:348–375

Ferreira DC, Marques RC (2016b) Should inpatients be adjusted by their complexity and severity for efficiency assessment? Evidence from Portugal. Health Care Manag Sci 19(1):43–57

Ferreira DC, Marques RC (2016c) Malmquist and Hicks-Moorsteen productivity indexes for clusters performance evaluation. Int J Inf Technol Decis Mak 15(5):1015–1053

Ferreira DC, Marques RC (2017) A step forward on order-α robust nonparametric method: inclusion of weight restrictions, convexity and non-variable returns to scale. Oper Res. https://doi.org/10.1007/s12351-017-0370-1

Ferreira DC, Marques RC (2018) Do quality and access to hospital services impact on their technical efficiency? Omega Int J Manag Sci. https://doi.org/10.1016/j.omega.2018.07.010

Ferreira DC, Marques RC, Pedro MI (2016) Comparing efficiency of holding business model and individual management model of airports. J Air Transp Manag 57:168–183

Ferreira DC, Marques RC, Nunes AM (2017a) Economies of scope in the health sector: The case of Portuguese hospitals. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2017.09.044

Ferreira DC, Marques RC, Nunes AM, Figueira JR (2017b) Patients' satisfaction: the medical appointments valence in Portuguese public hospitals. Omega Int J Manag Sci. https://doi.org/10.1016/j.omega.2017.08.009

Ferrier G, Trivitt JS (2013) Incorporating quality into the measurement of hospital efficiency: a double DEA approach. J Product Anal 40(3):337–355

Ferrier GD, Valdmanis VG (2004) Do mergers improve hospital productivity? J Oper Res Soc 55(10):1071–1080

Fiscella K, Franks P, Gold MR, Clancy CM (2000) Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. J Am Med Assoc 283(19):2579–2584

Garavaglia G, Lettieri E, Agasisti T, Lopez S (2011) Efficiency and quality of care in nursing homes: an Italian case study. Health Care Manag Sci 14(1):22–35

Gholami R, Higón DA, Emrouznejad A (2015) Hospital performance: efficiency or quality? Can we have both with IT? Expert Syst Appl 42:5390–5400

Girginer N, Köse T, Uçkun N (2015) Efficiency analysis of surgical services by combined use of Data Envelopment Analysis and Gray Relational Analysis. J Med Syst. https://doi.org/10.1007/s10916-015-0238-y

Gok MS, Sezen B (2013) Analyzing the ambiguous relationship between efficiency, quality and patient satisfaction in healthcare services: the case of public hospitals in Turkey. Health Policy 111(3):290–300

Grossman M (1972) On the concept of health capital and the demand for health. J Political Econ 80(2):223–255

Guerrini A, Romano G, Campedelli B, Moggi S, Leardini C (2017) Public vs. private in hospital efficiency: Exploring determinants in a competitive environment. Int J Public Adm 41(3):181–189

Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, Hudson M (2002) What does 'access to health care' mean? J Health Serv Res Policy 7(3):186–188

Hadad S, Hadad Y, Simon-Tuval T (2013) Determinants of healthcare systemâ€™s efficiency in oecd countries. Eur J Health Econ 14(2):253–265

Helling DK, Nelson KM, Ramirez J, Thammy LH (2006) Kaiser permanente colorado region pharmacy department: Innovative leader in pharmacy practice. J Am Pharm Assoc 46(1):67–76

Hirschberg JG, Lloyd PJ (2002) Does the technology of foreign-invested enterprises spill over to other enterprises in china? an application of post-dea bootstrap regression analysis. In: Lloyd PJ, Zang XG eds. Modelling the Chinese Economy. Edward Elgar Press, London, UK

Hollingsworth B (2008) The measurement of efficiency and productivity of health care delivery. Health Econ 45:1107–1128

Hollingsworth B, Peacock SJ (2008) Efficiency measurement in health and health care. Routledge, New York, USA

Hvenegaard A, Arendt JN, Street A, Gyrd-Hansen D (2011) Exploring the relationship between costs and quality: Does the joint evaluation of costs and quality alter the ranking of Danish hospital departments? Eur J Health Econ 12(6):541–551

Khushalani J, Ozcan YA(2017) Are hospitals producing quality care efficiently? An analysis using Dynamic Network Data Envelopment Analysis (DEA). Socio-Econ Plan Sci 60:15–23

Kissick W (1994) Medicine's dilemmas: infinite needs versus finite resources. Yale University Press, New Haven, CT

Kontodimopoulos N, Papathanasiou ND, Tountas Y, Niakas D (2010) Separating managerial inefficiency from influences of the operating environment: an application in dialysis. J Med Syst 34(3):397–405

Kuo Y, Yang T, Huang G (2008) The use of grey relational analysis in solving multiple attribute decision-making problems. Comput Ind Eng 55(1):80–93

Laine J, Linna M, Häkinen U, Noro A (2005) Measuring the productive efficiency and clinical quality of institutional long-term care for the elderly. Health Econ 14(3):245–256

Lin S, Horng S, Lee B, Fan P, Pan Y, Lai J, Chen R, Khan M (2011) Application of grey-relational analysis to find the most suitable watermarking scheme. Int J Innovative Comput Inf Control 7(9):5389–5401

Lindlbauer I, Schreyögg J (2014) The relationship between hospital specialization and hospital efficiency: do different measures of specialization lead to different results? Health Care Manag Sci 17:365–378

Lyder CH, Wang Y, Mtersky M, Curry M, Kliman R, Verzier N, Hunt DR (2012) Hospital-acquired pressure ulcers: Results from the national medicare patient safety monitoring system study. J Am Geriatr Soc

Marques RC, Carvalho P (2013) Estimating the efficiency of Portuguese hospitals using an appropriate production technology. Int Trans Oper Res 20(2):233–249

Martini G, Berta P, Mullahy J, Vittadini G (2014) The effectiveness-efficiency trade-off in health care: The case of hospitals in Lombardy, Italy. Reg Sci Urban Econ 49:217–231

Matranga D, Sapienza F (2015) Congestion analysis to evaluate the efficiency and appropriateness of hospitals in Sicily. Health Policy 119(3):324–332

Mobley LR, Magnussen J (2002) The impact of managed care penetration and hospital quality on efficiency in hospital staffing. J Health Care Financ 28(4):24–42

Morán J, Granada E, Míguez JL, Porteiro J (2006) Use of grey relational analysis to assess and optimize small biomass boilers. Fuel Process Technol 87:123–127

Morey RC, Fine DJ, Loree SW, Retzlaff-Roberts DL, Tsubakitani S (1992) The trade-off between hospital cost and quality of care. Med Care 30(8):677–698

Navarro-Espigares JL, Torres EH (2011) Efficiency and quality in health services: a crucial link. Serv Industries J 31(3):385–403

Nayar P, Ozcan YA (2008) Data envelopment analysis comparison of hospital efficiency and quality. J Med Syst 32(3):193–199

Nayar P, Ozcan YA, Yu F, Nguyen AT (2013) Benchmarking urban acute care hospitals: efficiency and quality perspectives. Health Care Manag Rev 38(2):137–145

Olson DL, Wu D (2006) Simulation of fuzzy multiattribute models for grey relationships. Eur J Oper Res 175:111–120

Ozcan YA (2014) Health care benchmarking and performance evaluation: an assessment using data envelopment analysis (DEA). Springer-Verlag, New York, US

Perez ED (1992) Regional variation in vamc's operative efficiency. J Med Syst 16(5):207–213

Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH (2008) Poverty and access to health care in developing countries. Ann N Y Acad Sci 1136:161–171

Rego G, Nunes R, Costa J (2010) The challenge of corporatisation: the experience of Portuguese public hospitals. Eur J Health Econ 11 (4):367–381

Retzlaff-Roberts D, Chang CF, Rubin RM (2004) Technical efficiency in the use of health care resources: a comparison of oecd countries. Health Policy 69(1):55–72

Safdar KA, Emrouznejad A, Dey PK (2016) Assessing the queuing process using data envelopment analysis: An application in health centres. J Med Syst. https://doi.org/10.1007/s10916-015-0393-1

Schang L, Hynninen Y, Morton A, Salo A (2016) Developing robust composite measures of healthcare quality: ranking intervals and dominance relations for scottish health boards. Soc Sci Med 162:59–67

Shimshak DG, Lenard ML, Klimberg RK (2009) Incorporating quality into data envelopment analysis of nursing home performance: a case study. Omega Int J Manag Sci 37(3):672–685

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London, UK

Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Manag Sci 44(11):49–61

Simar L, Wilson PW (2007) Estimation and inference in two-stage semi-parametric models of production processes. J Econ 136:31–64

Simões P, Marques RC (2011) Performance and congestion analysis of the Portuguese hospital services. Cent Eur J Oper Res 19 (1):39–63

Singaroyan R, Seed CA, Egdell RM (2006) Is a target culture in health care always compatible with efficient use of resources? A cost-effectiveness analysis of an intervention to achieve thrombolysis targets. J Public Health 28(1):31–34

Sudore RL, Mehta KM, Simonsick EM, Harris TB, Newman AB, Satterfield S, Rosano C, Rooks RN, Rubin SM, Ayonayon HN, Yaffe K (2006) Limited literacy in older people and disparities in health and healthcare access. J Am Geriatr Soc 54(5):770–776

Sullivan N, Schoelles KM (2013) Preventing in-facility pressure ulcers as a patient safety strategy: a systematic review. Ann Intern Med 158:410–416

Tsang C, Aylin P, Palmer W (2008) Patient safety indicators: A systematic review of the literature. Centre for Patient Safety & Service Quality. Imperial College London. Available at: http://www1.imperial.ac.uk/resources/147E3ECA-1FD2-4AF8-BA34-08AAF1FBCAB1/psireportv3.pdf. Accessed 27 Feb 2018

Valdmanis VG, Rosko MD, Mutter RL (2008) Hospital quality, efficiency, and input slack differentials. Health Serv Res 43 (5):1830–1848

Varabyova Y, Blankart CR, Schreyögg J (2016a) Using nonparametric conditional approach to integrate quality into efficiency analysis: empirical evidence from cardiology departments. Health Care Manag Sci 20:565–576

Varabyova Y, Blankart CR, Torbica A, Schreyögg J (2016b) Comparing the efficiency of hospitals in italy and germany: nonparametric conditional approach based on partial frontier. Health Care Manag Sci 20:379–394

Varabyova Y, Blankart CR, Schreyögg J (2017) Integrating quality into the nonparametric analysis of efficiency: a simulation comparison of popular methods. Ann Oper Res 261:365–392

Worthington AC (2004) Frontier efficiency measurement in health care: a review of empirical techniques and selected applications. Med Care Res Rev 61(2):135–170

Yang J, Zeng W (2014) The trade-offs between efficiency and quality in the hospital production: some evidence from Shenzhen, China. China Economic Rev 31:245–256

Zhang N, Choi Y (2014) A note on the evolution of directional distance function and its development in energy and environmental studies 1997–2013. Renew Sustain Energy Rev 33:50–59

Zhou P, Ang BW, Wang H (2012) Energy and CO2 emission performance in electricity generation: a non-radial directional distance function approach. Eur J Oper Res 221(3):625–635