

Evaluation of efficiency in Colombian hospitals: An analysis for the post-reform period

Víctor Giménez^a, William Prieto^b, Diego Prior^a, Emili Tortosa-Ausina^{c,*}

^a Department of Business, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

^b Faculty of Economic and Administrative Sciences, Universidad Católica de Colombia, Bogotá, Colombia

^c Department of Economics, Universitat Jaume I, Castelló, Spain

ARTICLE INFO

Keywords:

Hospitals

Efficiency

Quality

Bad outputs

Global Malmquist-Luenberger index/JEL

Classifications:

C61

H51

I18

ABSTRACT

In this study we analyze the performance of 602 level 1 Colombian hospitals for the period 2009–2013. The analysis is carried out from both static and temporal perspectives in order to evaluate the evolution of total factor productivity (change in hospital performance) and its components throughout the period. The study also explores a question relevant not only to the Colombian health system, but to many others around the world, of whether primary care centers excessively refer patients to high-level hospitals, thereby negatively affecting the quality, efficiency, and effectiveness of all healthcare service provision. The results demonstrated that adjusted production (service provision) and levels of quality and referrals to higher-level hospitals could be improved, on average, by 44%. This increase in health service provision levels and their quality can be achieved by reducing personnel expenditure (by an average of 22%), expenditure on medicines (by 20%), and purchasing expenses (by 11%). The temporal analysis shows that total factor productivity (hospital performance change) worsened by 1% during the period, mainly due to the technological backlash experienced despite a slightly positive evolution in efficiency.

1. Introduction

In the early 1990s, under the principles of efficiency, universality, solidarity, comprehensiveness, unity, and participation, Law 100 (1993) ushered in the most significant reform to the healthcare system in Colombia's modern history. The fundamental objective of the new system was to ensure universal coverage in the healthcare system, particularly for the poor and vulnerable sectors of the population. In this way, the new regulation sought to gradually extend the system's coverage, establish free basic care and improve the organization of healthcare services in a decentralized manner, while also preventing inefficient use of healthcare resources.

The government has critical responsibilities to evaluate the results of the healthcare system in relation to the principles of efficiency and professional quality and suitability, and to create a policy for managing healthcare information. Under the law, efficiency is a priority objective, and the system must guarantee the effective provision of healthcare to the entire population through the use of technology, resources and available services. In turn, as regards quality and professional suitability, the statute states that the system's services must be user-focused through medical and technical appropriation that guarantees

compliance with quality standards endorsed by the scientific community within the healthcare sector. This situation therefore requires suitable evidence on the current status of efficiency and quality of public hospitals.

From an operational point of view, the regulations establish that all affiliates, without discrimination, will receive an integrated healthcare protection plan known as the Obligatory Healthcare Plan (OHP), which covers preventive care, medical-surgical care and essential medications. The Solidarity and Guarantee Fund (FOSYGA), affiliated to the Ministry of Health, is responsible for collecting contributions. The insurance carriers (EPS), as the administrative and funding bodies, collect the contributions, remitting to FOSYGA the difference between the premium they receive for each affiliate and the corresponding contributions from affiliates with formal and independent work contracts. The insurance premium the EPS receive from FOSYGA is a capitation payment unit (UPC) adjusted for age and gender, which is established annually by the National Council for Social Security in Healthcare. Affiliates are free to choose both their EPS and the healthcare promotion institution (IPS) subscribed to the EPS network services. In turn, the IPS may be either public entities (public hospitals), or mixed, private, community or charitable entities that provide healthcare services

* Corresponding author.

E-mail address: tortosa@uji.es (E. Tortosa-Ausina).

to affiliates. Care services are classified according to levels of complexity: level 1 covers outpatient care and inpatient care, level 2 offers specialized outpatient care, while levels 3 and 4 provide specialized inpatient care.

The Department of Industry and Trade has made several recommendations to improve the provision of insurance and healthcare services in the system.¹ One of these suggestions is to construct relative efficiency indicators for provision of insurance and healthcare services to promote a mechanism of competition by comparison. The present study responds to this recommendation by carrying out a temporal efficiency analysis for level 1 public hospitals for the period 2009–2013. Therefore, the aim of the study is to assess efficiency and productivity levels and temporal evolution of hospital performance in Colombia during a recent period corresponding to the post-reform (i.e., application of Law 100) era. The reason for choosing level 1 public hospitals is twofold. First, quantitative and qualitative information is more readily available for level 1 public hospitals than for other healthcare service providers, and is more frequent in the sample of public hospitals monitored by the Ministry of Health. Our database contains some interesting variables particularly related to the quality of the healthcare provision process, which should be considered as bad outputs, i.e., variables (such as waiting time or re-admissions within 20 days) whose decreases imply improvements in performance. The second reason is that performance of level 1 public hospitals may directly affect the efficiency and effectiveness further up the healthcare system, since they are generally the initial contact for a healthcare system user and, as such, they should be able to deal efficiently and effectively with many of the patients and avoid unnecessary referral to higher levels. In this line, we address the additional research question of whether too many patients are referred to higher levels from primary care levels, which could increase costs and reduce quality in the more complex hospitals. This is a common concern among healthcare managers. Finally, the analyzed period (2009–2013) was chosen due to the availability of consistent and comparable data.²

The literature on healthcare sector efficiency is extensive, has a variety of aims, and applies to different countries.³ Recent developments in the literature have extended productivity analysis to include minimization of bad outputs. In this line, Scheel [1], Jahanshahloo et al. [2], Ozcan et al. [3] and Hadi Vencheh et al. [4] propose alternatives within the data envelopment analysis framework [5] to include bad outputs. Scheel [1] suggests new radial efficiency measures,

assuming proportional changes in both good and bad outputs, while Jahanshahloo et al. [2] and Hadi Vencheh et al. [4] propose models with multiple outputs, including bad or undesirable outputs. However, to the best of our knowledge, this is the first study using a global Malmquist-Luenberger index to assess the temporal evolution of productivity, technological change and efficiency in the hospital sector.

Two particular issues are of special interest for our purposes: first, the trade-off between efficiency and quality and, second, the effects of the reform in the healthcare system. The trade-off between efficiency and quality in the provision of healthcare services is often considered relevant since pressures on production can indirectly affect quality provision. The literature reports evidence for and against this relationship. For instance, Prior [6] found that between 30% and 40% of Catalan hospitals upheld the postulates of Total Quality Management (that is, maximum improvement in desirable outputs simultaneously induces reduction in the undesirable attributes of bad quality). Navarro-Espigares and Torres [7] carried out a temporal DEA analysis using the Malmquist index to characterize the relationship between quality and efficiency in hospitals in a region of Spain from 1997 to 2004, finding a slight association between efficiency indicators and quality indicators. Yang and Zeng [8] three-stage DEA study, also implemented by estimating a Malmquist index to measure changes in productivity, efficiency and quality in hospitals in the city of Shenzhen (China) from 2006–2010, does not dismiss the possibility of a relationship between efficiency and quality, particularly in small and medium hospitals.

Several scholars have explored the effects of healthcare system reforms using non-parametric and parametric frontier models. Arocena and García Prado [9] and [10] report interesting comparisons of market and cooperative reforms incorporating hospital management incentives in both cases. Arocena and García Prado [9], using a nonparametric methodology to estimate a productivity index, determine how public hospitals in Costa Rica respond to reforms to the healthcare system implemented between 1997 and 2001. The results indicate an increase in productivity associated with increases in quality, promoted mainly by technical and scale efficiency rather than a regressive technological change. In their study, Hu et al. [10] found the reform had a more positive effect on efficiency for the inland Chinese provinces compared to those located on the coast. Both studies considered data with multiple products and also included bad outputs. Although the reform in Costa Rica is geared more toward a cooperative model, evidence suggests that the relationship between efficiency and quality of service responds appropriately to reforms that attempt to align their incentives with those of public hospital administration. This also appears to be confirmed by Alonso et al. [11], who compare hospitals with traditional administrations in Madrid. Their results suggest that what makes the difference is management, rather than the particular management model the hospitals adopt. Likewise, the studies by Matranga and Sapienza [12] and Fidler et al. [13] also incorporate bad outputs. Their results are congruent with the above evidence since the inefficiency of public hospitals may be attributed to congestion, purely technical inefficiency, and the need to align the incentives of hospital management to the reform objectives. In particular, Fidler et al. [13] showed how the combination of public resources and private management generated a new corporate culture with greater flexibility to correct inefficiencies by reducing size and generating economies of scale. More recently, Chen et al. [14] found that the changes introduced in the healthcare sector in China during the last thirty years have improved the cost efficiency of the whole system, especially after the public subsidies and medical insurance reform.

Turning to studies on the healthcare sector in Colombia, Sarmiento et al. [15] analyzed the efficiency of public hospitals in 2003, while McPake et al. [16] carried out a parametric estimation of time series and a graphed trend analysis together with qualitative methods to evaluate the impact of reform in five hospitals in Bogotá, monitored from 1996 to 1998. The study finds evidence of increased activities and

¹ Department of Industry and Trade. Problem of competence in the health sector in Colombia, presentation given at the 2nd annual meeting of the work group for trade and competence, June 2012.

² The issue as to the transfer of patients to levels of greater complexity by hospitals has to do with productivity increases in public hospitals. If we do not take into account quality indicators, hospitals transfer patients whose problems can be resolved in level 1 complexity. These transfers are made to higher levels of complexities with more expensive technologies, which increase the cost for a given level of production which, *ceteris paribus*, would erode efficiency. This is a result of keeping productivity levels to hold contract services with EPSS. In other words, by increasing the number of patients, productivity levels increase, but without improving quality services, this increment will reduce efficiency as long as it deviates health issues that can be solved in levels of service with lower costs. For this reason, hospitals are not interested in transferring a smaller number of patients because it affects productivity—for a given level of resources. In this way, fewer transfers imply losses in productivity with regard to other hospitals which are competitors for service provision. While users can resort to other hospitals, access to levels of greater complexity depend on the authorization of medical professionals in each hospital. For this same reason, increases in production which are not accompanied by improvements in quality are considered a bad output.

³ Some representative examples of this can be found in Nunamaker [49], Lewin [50], Harris et al. [51], Valdmanis et al. [52], Weng et al. [53], Harrison et al. [54], Dexter and O'Neill [55], Hollingsworth [56,57], Pina and Torres [58], Solà and Prior [46], Dalmau-Atarrodona and Puig-Junoy [59], Helmig and Lapsley [60], Field and Emrouznejad [61], Ersoy et al. [62], Sahin and Ozcan [63], Giokas [64], Ouellette and Vierstraete [65], Athanassopoulos and Gounaris [42], Gerdtham et al. [66], Björn et al. [67], Kirigia et al. [68], Ferrier and Valdmanis [69], Prior [6], Grosskopf et al. [70,71], Harrison et al. [54], Wang et al. [72], Cordero et al. [73], Martini et al. [74], O'Neill [75], Kalhor et al. [76], Ozcan [77] Chen et al. [14], Büchner et al. [78] or Jola-Sanchez et al. [79].

productivity following the implementation of Law 100 (1993) without relevant effects on quality of healthcare service provision. Similarly, Pinzón-Martínez [17] estimated nonparametric frontiers using a model which evaluates the performance of low-complexity public hospitals, finding that 70% of the institutions analyzed operate with an average efficiency coefficient of 72.2%. Tamayo and Maldonado [18] estimated technical efficiency in public hospitals for the period 2002–2005 with information from the hospital information system (“Sistema de Información Hospitalaria”, SIHO). The main contribution of their study is the alternative classification of five homogeneous groups, which differs from the usual classification by complexity of services.

The rest of this article is organized as follows. The next section (Section 2) describes the methodology used in this paper. The sample and variables selected as inputs, products and quality for the healthcare provision process are shown in Section 3. Results are discussed in Section 4, and finally, Section 5 presents the main conclusions, limitations and future research lines.

2. Methodology

2.1. Measuring efficiency and its evolution over time

Temporal efficiency studies often use the Malmquist index [19]. This index explains the change in total productivity of factors as a product of change in efficiency, or “catching up”, and technological change. Chung et al. [20] modified the Malmquist index in order to adapt it to cases of directional distance functions (DDF). These have been widely used in efficiency measurement studies, which incorporate the environmental impact of the units analyzed by looking at bad outputs of the production process [21–23]. The new index is known as the Malmquist-Luenberger index.

However, both indexes suffer from two problems [24,25]. First, their circularity is not assured. Circularity is when the change in productivity performance is explained through resulting changes in productivity in the different sub-periods it is comprised of. Secondly, there is the possibility of infeasibility in the calculation of cross-distance functions between the periods necessary to calculate it. In spite of it being a necessary and sufficient condition of technical change, being Hicks-neutral to ensure circularity [26], or that a determined data structure ensures the absence of feasibility problems [27], its empirical compliance is often difficult. Pastor and Lovell [25] proposed a modification of the Malmquist index to correct these two deficiencies, which they called the global Malmquist index. Oh [24] then adapted the global Malmquist-Luenberger index (henceforth GML) in a similar way.

This study uses the GML proposed by Oh [24] for the temporal analysis of productivity, which in our specific context refers to hospital performance. There are two reasons for choosing the GML to analyze performance of level 1 hospitals in Colombia. The first concerns with the existence of bad outputs to be minimized in the service provision process, especially related to quality variables. The second refers to compliance with circularity and feasibility implicit in that of the GML. The choice of the GML therefore complies with both the empirical and the theoretically desirable properties in any temporal study of hospital performance.

The description of the GML in the case of health (hospital) service provision in Colombia can be performed as follows. Let K hospitals with information available on their production systems for the years $t = 1 \dots T$. Since each hospital's service provision vector is composed of M outputs $y \in R_+^M$ and H bad outputs $b \in R_+^H$ through consumption of N inputs $x \in R_+^N$, then the production possibilities set is defined by:

$$P(x) = \{(y, b) \mid x \text{ can produce } (y, b)\} \quad (1)$$

The axioms that the above technology must meet are those commonly required by the theory of production [28]. The efficiency of any of the pertaining units at $P(x)$ can be measured through the following directional distance function, DDF [22,24,29]:

$$D(x, y, b) = \max(\beta \mid (y + \beta g_y, b - \beta g_b)) \quad (2)$$

The above DDF determines the maximum increase and reduction β attainable both in good and bad outputs, respectively, over the vector $g = (g_y, g_b)$ that defines the desirable directions for improvement for both types of outputs. In this study we use the vector of $M + H$ components $g = (y, b)$ as suggested by Chung et al. [20] and Oh [24].

The GML index for the years t and $t + 1$ is defined as [24]:

$$GML^{t,t+1}(x^t, y^t, b^t, x^{t+1}, y^{t+1}, b^{t+1}) = \frac{1 + D^G(x^t, y^t, b^t)}{1 + D^G(x^{t+1}, y^{t+1}, b^{t+1})} \quad (3)$$

where $D^G(x, y, b) = \max(\beta \mid (y + \beta g_y, b - \beta g_b) \in P(x))$ is the DDF defined over the global production possibilities set $P^G(x)$, i.e., a value of $GML^{t,t+1}$ greater than the unit represents an improvement in productivity between years t and $t + 1$, since the distance to the global frontier was greater in t than in $t + 1$. A value less than the unit for $GML^{t,t+1}$ represents a deterioration in the corresponding period.

Expression (3) can be decomposed in the following way:

$$\begin{aligned} GML^{t,t+1}(x^t, y^t, b^t, x^{t+1}, y^{t+1}, b^{t+1}) &= \frac{1 + D^G(x^t, y^t, b^t)}{1 + D^G(x^{t+1}, y^{t+1}, b^{t+1})} \\ &= \frac{1 + D^t(x^t, y^t, b^t)}{1 + D^{t+1}(x^{t+1}, y^{t+1}, b^{t+1})} \\ &\quad \times \left[\frac{1 + D^G(x^t, y^t, b^t)}{1 + D^t(x^t, y^t, b^t)} \right] \\ &= \frac{TE^{t+1}}{TE^t} \times \left[\frac{BPG_{t+1}^{t,t+1}}{BPG_t^{t,t+1}} \right] \\ &= EC^{t,t+1} \times BPC^{t,t+1} \end{aligned} \quad (4)$$

where $EC^{t,t+1}$ collects the technical efficiency change or catching up between the year t and $t + 1$. If $EC^{t,t+1} > 1$, there was an improvement in technical efficiency during the period. In other words, the unit is closer to its contemporary frontier in the year $t + 1$ than in t . A lower value is interpreted conversely. The term $BPC^{t,t+1}$ is a measure of the technological change in the period, i.e., it describes how the contemporary frontiers displaced during the period. Equation (4) shows that the expression for the calculation of $BPC^{t,t+1}$ is:

$$BPC^{t,t+1} = \frac{BPG_{t+1}^{t,t+1}}{BPG_t^{t,t+1}} = \left[\frac{1 + D^G(x^t, y^t, b^t)}{1 + D^t(x^t, y^t, b^t)} \right] \left[\frac{1 + D^G(x^{t+1}, y^{t+1}, b^{t+1})}{1 + D^{t+1}(x^{t+1}, y^{t+1}, b^{t+1})} \right] \quad (5)$$

where:

$$BPG_{t+1}^{t,t+1} = \frac{1}{1 + D^G(x^{t+1}, y^{t+1}, b^{t+1})} \quad (6)$$

In the above expression $BPG_{t+1}^{t,t+1}$ corresponds to the inverse of the quotient between distance to the frontier defined by $P^G(x)$ and to the contemporary frontier defined by $P^G(x)$ for the period $t + 1$. If $BPC^{t,t+1} > 1$ it is because the frontier in $t + 1$ is closer to the global frontier than in t , implying that technological progress has taken place. The case of $BPC^{t,t+1} < 1$ represents the opposite situation.⁴

Various methods can be used to calculate $D^G(x^t, y^t, b^t)$ and $D^t(x^t, y^t, b^t)$. In this study we use data enveloping analysis models [5] widely used in efficiency studies [30]. The calculation of $D^t(x^t, y^t, b^t)$ is made by solving the following linear program for each hospital analyzed under the assumption $g = (y, b)$ [31]:

⁴ The technical change component, apart from exhibiting progress or regress, can also be adapted to quantify the potential existence of biased technical change. In order to maintain the length of this paper in a reasonable dimension, we do not enter in this potential extension.

$$\begin{aligned}
& \text{Max } \beta \\
& \text{s. t. } \sum_{k=1}^K \lambda_k y_{km}^t \geq y_m^{ot} (1 + \beta) \quad m = 1 \dots M \\
& \quad \sum_{k=1}^K \lambda_k b_{kh}^t \leq b_h^{ot} (1 - \beta) \quad h = 1 \dots H \\
& \quad \sum_{k=1}^K \lambda_k x_{kn}^t \leq x_n^{ot} \quad n = 1 \dots N \\
& \quad \sum_{k=1}^K \lambda_k = 1 \\
& \quad \beta \geq 0; \lambda_k \geq 0 \quad k = 1 \dots K
\end{aligned} \quad (7)$$

where β is the maximum increase and reduction attainable simultaneously in the good and bad outputs, respectively; y_{km}^t represents output m of unit k in year t , b_{kh}^t the bad output h of unit or hospital k in year t , and x_{kn}^t the input n used by hospital k in year t ; y_m^{ot} , b_h^{ot} , x_n^{ot} are the observed levels of outputs, bad outputs and inputs, respectively, for the hospital evaluated in year t .

Analogously, for the calculation of $D^G(x^t, y^t, b^t)$, the following linear program is used:

$$\begin{aligned}
& \text{Max } \beta \\
& \text{s. t. } \sum_{k=1}^K \sum_{t=1}^{t+1} \lambda_k^T y_{km}^T \geq y_m^{ot} (1 + \beta) \quad m = 1 \dots M \\
& \quad \sum_{k=1}^K \sum_{t=1}^{t+1} \lambda_k^T b_{kh}^T \leq b_h^{ot} (1 - \beta) \quad h = 1 \dots H \\
& \quad \sum_{k=1}^K \sum_{t=1}^{t+1} \lambda_k^T x_{kn}^T \leq x_n^{ot} \quad n = 1 \dots N \\
& \quad \sum_{k=1}^K \sum_{t=1}^{t+1} \lambda_k^T = 1 \\
& \quad \beta \geq 0; \lambda_k^T \geq 0 \quad k = 1 \dots K
\end{aligned} \quad (8)$$

The main difference from (7) is that the frontier is determined by considering all the observations for the period.

2.2. Bipartite decomposition of the relative contributions to hospital performance

In accordance with the expressions detailed above, the global Malmquist-Luenberger index (*GML*) index, which captures changes in hospital performance, can be decomposed into technical change (*EC*) and best practice gap change (*BPC*). Apart from analyzing the contributions of each of these components makes to the overall hospital performance change between t (2009) and $t + 1$ (2013), as measured by *GML*, on average, we can also evaluate what the largest contributors are via an analysis of counterfactual distributions. To this end, we use nonparametric density estimation, based on kernel smoothing.

Specifically, we rewrite expression (4) above as follows:

$$gml^{EC \times BPC} = EC^{t,t+1} \times BPC^{t,t+1} \quad (9)$$

according to which $gml^{EC \times BPC}$ indicates that the change in hospital performance is obtained by successively multiplying its two components which, in turn, enables us to construct counterfactual distributions by sequentially introducing each component.

In particular, the counterfactual hospital performance change attributable to *changes in efficiency* isolates the effect on the distribution of changes due to efficiency only, assuming *BPC* does not contribute to the change in hospital performance (*gml*), and it corresponds to the following expression

$$gml^{EC} = EC^{t,t+1} \quad (10)$$

Continuing with the sequential decomposition, we would proceed as follows:

$$gml^{EC \times BPC} = EC^{t,t+1} \times BPC^{t,t+1} = gml^{EC} \times BPC^{t,t+1} \quad (11)$$

We can invert the order of the sequential decomposition. In such a case, the counterfactual hospital performance change attributable to best practice gap change would be:

$$gml^{BPC} = BPC^{t,t+1} \quad (12)$$

In this case we would be isolating the effect on the distribution of best practice gap changes only, assuming *EC* does not contribute to the change in hospital performance (*gml*). Hence expression (11) would

become:

$$gml^{BPC \times EC} = BPC^{t,t+1} \times EC^{t,t+1} = gml^{BPC} \times EC^{t,t+1} \quad (13)$$

We will refer to the decomposition in both expressions (11) and (13) as the bipartite decomposition of the relative contributions to the changes in the distribution of hospital performance.

As indicated above, we use kernel smoothing to estimate the densities, which involves selecting a kernel and a bandwidth (h).⁵ Of the choices available (e.g., Epanechnikov, triangular, Gaussian etc.), we chose the popular (Gaussian kernel). The impact on the final outcome is, however, much lower than that of the bandwidth, which smooths the number *bumps* under each curve—higher values of h tend to smooth more, low values of h tend to smooth less, providing more detail. Given the relevance of this decision, we consider both a *global* (same amount of smoothing at all data points, see Ref. [32] and *local* bandwidth (varying amount of smoothing dependent on the structure of the data at a given point, see Ref. [33]).

Counterfactual distributions are not common in the field of hospital performance measurement (or health and healthcare in general), although they are more widely used in other fields. We have followed Kumar and Russell [34] approach in the context of macroeconomic convergence, which combines the (deterministic) frontier production function literature [35] with the distribution dynamics approach to convergence analysis [36,37]. Kumar and Russell [34] contribution was soon followed by extensions of the model that factored in the contribution of human capital [38] and financial development [39] to productivity growth and convergence.

One of the main advantages of the bipartite decomposition suggested is that we focus on how the *entire* distributions of hospital performance evolve, instead of concentrating on few summary statistics such as the mean or standard deviation. Therefore, we can detect possible groups of hospitals which perform poorly or, at the opposite extreme, clusters of benchmarks to be imitated. Any summary statistic, such as the mean, would conceal this meaningful information.

3. Sample and variables

Three subtypes of level 1 hospitals were identified based on the services they provide. The first group consists of hospitals with facilities for outpatient consultations and emergency care only. Hospitals in the second group, in addition to the services provided by the first group, also have inpatient beds. Finally, the third group contains hospitals that, in addition to the above, have operating rooms for simple surgical procedures. These structural differences indicate that the production process is not totally homogeneous and, consequently, a joint evaluation does not seem advisable.

Table 1 reports the distribution of the data available by level 1 hospital subtype. The sample consists of 602 hospitals for which complete information was obtained. The tendency during the period studied was to reduce the number of hospitals with surgical capacity, and increase the number of hospitals with inpatient capacity. The table shows that some hospitals changed category during the period, which could make interpretation of results more difficult in these special cases. For this reason, and due to space restrictions, results will be presented only at aggregate level.

As indicated earlier, data were obtained from the information system on public hospital management, SIHO. It is one of the most complete available databases since it provides information on physical, human and financial resources and, simultaneously, it builds different output aggregates relevant for an analysis of efficiency. By law,⁶ it is mandatory that health care provision institutions report the necessary

⁵ The interested reader can refer to Silverman [80], Scott [81], Li and Racine [82] and, more recently, Henderson and Parmeter [83].

⁶ Article 52 of Law 812 of 2013, Colombia.

Table 1
Number of hospitals by type.

Year	Surgical	With hospitalization	Without hospitalization	Total
2009	76	450	76	602
2010	66	460	76	602
2011	60	470	72	602
2012	57	468	77	602
2013	49	476	77	602

and sufficient information that is established by the Comprehensive Health Care Information System. As a consequence, all public hospitals report complete health care service provision budgetary information to the Ministry of Health. In the period from 2009 to 2013, the records of the database show the existence of 949 public hospitals, of which the national level hospital and mental health hospitals are not taken into account. In this sense, the data on which this research is based contemplates all public hospitals during the period of analysis. Specifically, the distribution of the universe of hospitals corresponds to 802 level 1 hospitals, 121 level 2 hospitals and 26 level 3 hospitals. In the estimation of the DEA model, 602 level 1 complexity hospitals were used, which represents 75% of the total universe of level 1 public hospitals registered in the SIHO database.

3.1. Selection of inputs

A relevant question when evaluating efficiency in the hospital (health) sector is which inputs and outputs to use to represent the production process. A large number of operational variables have been used in both categories [40], as they often depend on available databases. The most commonly used input variables can be grouped into three main categories. The first category refers to investment in **capital**. Ozcan and Luke [41] proposed measuring capital investment with two indicators: the number of operational beds (as a measure of hospital size), and the complexity of services provided, measured as the number of diagnostic and special services offered exclusively by the hospital. The second input category concerns **personnel**. This variable has frequently been used in studies, either expressed in terms of physical units of available personnel, which are generally transformed to full time equivalent (FTE) persons in a complete working day, or in monetary terms. Studies normally differentiate personnel according to their functions: medical care (sometimes distinguishing between doctors and nurses), administrative or other functions. Examples of specification of personnel as an input can be found for instance in Athanassopoulos and Gounaris [42] and Ozcan and Luke [41]. The third common category of inputs in the literature is operational expenditures. This variable generally covers other expenditures that are not related to personnel, such as amortizations, supplies, and especially the purchase of goods and services necessary to provide the service [41].

The inputs selected for our study were the following:

- Total spending on purchase of goods and services (I_1). Expressed in thousand billions of Colombian pesos.
- Total spending on medicines (I_2). Expressed in thousand billions of Colombian pesos.
- Total monthly spending on administrative personnel (I_3). Expressed in thousand billions of Colombian pesos.
- Total monthly spending on medical personnel (I_4). Expressed in thousand billions of Colombian pesos.

The staff variables have been expressed in monetary terms because the information in physical units was not available in the database. However, this should not be a problem, since the differences in salaries between the different Colombian public hospitals are not high. In addition, this option has been used previously in the literature [43–45].

All these variables were deflated to be expressed in 2008 prices, thus preventing the price effect from distorting the efficiency measures.

3.2. Outputs

The most commonly used outputs in evaluating non-teaching hospitals are outpatients and inpatients. The main difficulty in computing the latter is not so much their number, but rather their complexity. For this reason, some type of homogenized measurement is generally associated with the complexity of each illness or condition. Several studies have also included service quality provided by hospitals. For instance, Solà and Prior [46] use nosocomial infections as a measure of technical quality. Nosocomial infections are contracted during hospital stays and are closely linked to poor sterilization and disinfecting of surgical instruments and equipment, among other causes. Other variables have also been used to measure quality, such as hospital re-admissions within a given period related to the same condition. In addition to variables associated with technical quality, Vuori [47] recommends using other variables more closely related to quality of service, one of which is patient satisfaction. Hospital quality would thus comprise a technical component as well as service quality. Note that from a methodological point of view, quality variables used in the hospital (health) sector often improve when their value is minimized and as such they should be considered as undesirable or bad outputs.

The inputs, outputs, and quality variables were obtained from the hospital information system (SIHO) of the Ministry of Health and Social Protection of the Republic of Colombia, following Tamayo and Maldonado [18]. The period of study for which information is available and comparable is 2009–2013.

Quantification of service provision levels is always complex in the case of hospitals. The difficulty lies in the wide variety of services provided and their varied levels of complexity. To overcome this problem, all the production will be summarized in a single variable. In this case, production can be added because information was available on the “relative value unit” (UVR). This is the criteria used to determine the complexity of the treated cases when the health care providers (IPS) bill their services to the insurance carriers (EPS). Combining all production into a single variable makes it easier to include other variables such as those related to quality without affecting the discriminatory capacity of the model. Therefore, the final chosen output variable was:

- **Weighted production (O_1)**. A sum of the weighted services provided by their UVR.⁷

Table 2 presents the descriptive statistics for the inputs and outputs selected.

The tendency for inputs I_1 and I_2 grew over the period studied, although growth was higher for purchases of goods and services, while

⁷ Public hospitals produce a heterogeneous variety of health services. In order to obtain an aggregate measurement of public hospitals' service provision, several methodologies have been proposed recently—particularly for the case of Colombia. Cortés-Martínez [84], Gómez Duque [85] and Sarmiento Gómez et al. [86] described a weighted production measurement that is based on creating equivalent relations among health services. For instance, if the equivalent measure is the cost of a surgery, all health services will be compared to this cost in order to get a relative weight among total production. According to Sarmiento Gómez et al. [86], the relative value unit was first calculated by the health and social security public institute, with the purpose of measuring different technical complexity levels among public hospitals production. This relative unit was made by a group of experts whose expertise is relevant as to set a specific value for each service produce by hospitals. In doing this, the experts take into account how complex is the service provision in terms of knowledge, skills, and technological inputs. Therefore, delivering a surgery have a higher weight than a medical consultation, given that the former requires a higher specialized set of skills and more complex technological inputs. Today, the relative value unit estimates are given by the Ministry of Health and Social Protection to public hospitals in order to calculate an aggregate production level. Public hospitals used this relative unit value to provide an estimate of aggregate production which later on is available in the public hospital database.

Table 2
Descriptive statistics for inputs and outputs.

Year	Statistic	Inputs				Outputs
		Spending on purchase of goods and services ^a (I_1)	Spending on medicines ^a (I_2)	Spending on administration personnel ^a (I_3)	Spending on assistance personnel ^a (I_4)	Weighted production (O_1)
2009	Mean	422,000.00	245,000.00	18,300.00	15,400.00	129,027.60
	Std. Dev.	560,000.00	344,000.00	162,000.00	11,400.00	189,931.20
2010	Mean	442,000.00	262,000.00	12,000.00	15,300.00	128,607.30
	Std. Dev.	641,000.00	345,000.00	10,200.00	11,800.00	192,601.20
2011	Mean	491,000.00	281,000.00	12,700.00	15,600.00	128,163.20
	Std. Dev.	684,000.00	393,000.00	12,000.00	11,600.00	194,068.00
2012	Mean	524,000.00	297,000.00	13,300.00	17,000.00	134,804.10
	Std. Dev.	744,000.00	426,000.00	10,800.00	14,500.00	216,702.80
2013	Mean	631,000.00	316,000.00	14,200.00	17,600.00	139,664.90
	Std. Dev.	920,000.00	499,000.00	13,600.00	14,100.00	220,145.90
Total	Mean	502,000.00	280,000.00	14,100.00	16,200.00	132,053.40
	Std. Dev.	723,000.00	406,000.00	73,100.00	12,800.00	203,018.80

^a In thousands billions of 2008 Colombian pesos.

the rising tendency was more moderated in the area of medicines. Personnel expenditure reflected an upward trend in spending on nursing staff, and a falling trend in administrative personnel expenses, despite showing higher average spending since 2010. The trend was therefore toward an increase in real terms of all inputs during the period. Production levels, however, stagnated until 2011, showing a moderate increase during the final two years of the period. Consequently, costs increased in the 2009–2011 period even though adjusted production did not.

We selected several quality variables to reflect performance from a technical perspective such as patient satisfaction:

- *Satisfied patients* (Q_1). Number of patients who claimed to be satisfied in the survey carried out by the Ministry of Health and Social Protection. For technical purposes, this variable was considered as an output hospitals should maximize, given the supplies used, together with adjusted production.⁸
- *Total wait in days* (Q_2). Total number of days patients had to wait for an appointment until.
- *Total minutes waiting in the emergency room* (Q_3). Total number of minutes waiting to be attended in the emergency room.
- *Re-admission within 20 days* (Q_4). Total number of patients re-admitted to hospital in a period of less than 20 days after discharge for the same condition.
- *Patients referred to higher levels* (Q_5). Total number of patients referred to higher levels.

The final four variables were considered as bad outputs, since the minimization of their value would lead to improved quality. Regarding the referred patients variable (Q_5), we assumed that, as with the value

for the rest of variables, hospitals which refer fewer patients to higher levels without negatively affecting the rest of the variables would function more effectively, thus reducing congestion at more specialized higher levels. This would contribute to a more efficient and effective management of resources in the whole healthcare system. On the other hand, high values of the total variable waiting time in days (Q_2) could be also interpreted as a high-quality level of the evaluated hospital, contrary to what is supposed here. However, the reason for considering it as an indicator to be minored is that, at the same value of the rest of the quality variables, shorter waiting times would be preferred by patients. Additionally, although in some of the municipalities there are different health providers, in most of them the choice is low or non-existent. Regarding the variable patients referred to higher levels (Q_5) it should be highlighted that access to the levels of higher complexity depends on the authorization of medical professionals in each hospital. Therefore, low levels in this variable should not be related to the fact that patients seek care directly at higher level medical facilities, except in cases of manifest severity. In addition, interviews with expert analysts in the Colombian health system revealed their concern for what they considered to be and excessive and unjustified, from a technical point of view, level of referred patients to higher level hospitals, which was causing both technical and cost inefficiencies.

A technical quality variable often used in the hospital efficiency literature is the number of nosocomial infections [46]. However, we did not include this variable in our study because level 1 hospitals do not perform the complex surgical procedures often associated with this type of infection, especially through the use of poorly disinfected instruments and equipment. Therefore, it is an appropriate variable for healthcare units with high numbers of surgical procedures, but not for primary care centers where surgical activity is low or non-existent. Table 3 shows the descriptive statistics for the quality variables considered in our case.

Both the number of satisfied patients and the total number of minutes waiting to be attended in the emergency room remained fairly stable during the study period. We found the same result in the total number of days waiting for an appointment, with the exception of 2010, in which it fell noticeably.

The number of re-admissions in less than 20 days for the same condition was also stable for the period, with upturns in 2010 and 2011. However, a clear incremental tendency in the number of patients referred to higher-level hospitals is observed from 2009 to 2013. Ideally, level 1 hospitals would only refer essential cases to higher-level hospitals, thus allowing them to operate as efficiently as possible, rather than using them to resolve cases that should be treated in primary care centers. It would therefore be interesting to contain or reduce this variable whenever it does not negatively impact on the other quality

⁸ According to the Act 1011 of 2006, each public hospital must report the number of satisfied patients to the Ministry of Health and Social Protection, which is the institution in charge of providing technical assistance for its measurement. The variable is obtained from a survey that is applied to patients that have received services in public hospitals, therefore it is applied to patients once they have been discharged. Although there is not a technical requirement to undertake a sample design when implementing a survey to measure patient satisfaction, there is an explicit requirement to obtain a measurement that compares perceptions of the services that have been received to the expectations that patients might have before receiving them in public hospitals (Act 1011, Art.32). This has been done in the specialized literature in Colombia [87–89] by identifying gaps between what it is expected by patients, and what patients perceived they have received in each of the following aspects of public hospitals services: (1) infrastructural conditions; (2) administrative protocols and processes; and (3) health outcomes. Because of the lack of a standard scientific norm, there is not certainty about having all public hospitals doing a sample design that includes a random selection of patients when they apply a satisfaction survey, before sending data to the Ministry of Health and Social Protection. However, they are compelled by law to use scientific criteria in order to develop a quality improvement plan, that is later on audited by the Ministry of Health and Social Protection.

Table 3
Descriptive statistics for the quality variables.

Year	Statistic	Satisfied patients ^a (Q ₁)	Total wait ^b (Q ₂)	Total waiting in emergencies' room ^c (Q ₃)	Re-admission within 20 days ^a (Q ₄)	Patients referred to higher levels (number of patients) ^a (Q ₅)
2009	Mean	2043.527	37,017.5	192,183.8	10.55814	2413.093
	Std. Dev.	5003.232	79,710.4	636,781.7	79.18977	7446.975
2010	Mean	1936.935	31,364.660	213,367.200	5.636213	2355.669
	Std. Dev.	3808.124	70,745.860	685,429.200	18.62527	5398.188
2011	Mean	1882.042	29,637.12	190,967	7.566445	2636.186
	Std. Dev.	3677.645	65,983.130	594,297.900	43.82092	5994.869
2012	Mean	2027.188	27,636.700	193,785.500	5.594684	2814.231
	Std. Dev.	4187.541	61,436.480	544874.100	16.89938	6356.719
2013	Mean	2257.995	25,421.08	189,033.1	6.036545	3359.571
	Std. Dev.	4339.927	52,834.26	439,126.7	22.24471	7056.597
Total	Mean	2029.537	30,215.41	195,867.3	7.078405	2715.750
	Std. Dev.	4228.322	66,822.700	585,888.200	43.18319	6497.871

^a Number of patients.

^b Days.

^c Minutes.

variables. However, an upward trend in patient referrals to higher levels is observed for the period. This fact, together with an increase in the weighted production during that period and the other variables remaining fairly stable, may reflect a low filtering and resolution of cases in primary care, which may be leading to an unnecessary number of cases referred to higher level hospitals.

4. Results and discussion

Efficiency levels were estimated for each year by calculating the directional distance function (*static analysis*). The changes in hospital performance during the period were also quantified (*temporal analysis*), and are explained, in turn, as a result of the product of efficiency change (or catch up) and technological change. The concept of efficiency change refers to whether hospitals are closer to or further away from observed *best practices*—which are determined by units (hospitals) lying on the frontier. The technological change, in contrast, refers to how these have moved over the time period, i.e., to shifts of the frontier.

4.1. Results for hospital efficiency

4.1.1. General trends

As explained above, three categories of hospitals were considered to evaluate efficiency. The procedure used was as follows: (i) to evaluate **type 1** hospitals we also included those from types 2 and 3 as possible benchmarks; (ii) to evaluate **type 2** hospitals, type 3 hospitals were also included; (iii) finally, hospitals in the **type 3** category were evaluated separately. The rationale behind this procedure is that type 3 hospitals are those with more complex activities, which could imply higher costs and, ultimately, worse efficiency results. Therefore, it makes little sense to compare them with hospitals whose activities are less complex—i.e., hospitals with more elementary production processes. This procedure was proposed by Banker [48]; but the idea has also been also considered from other perspectives.⁹

Estimated efficiency levels are presented in Tables 4 and 5, calculated by solving linear programming problem (7). Efficiency scores may take the value of zero as a minimum as long as it does not have an upper bound; the lower its value, the higher the efficiency level. A value of zero represents efficiency, and high values represent inefficiency. Their interpretation is the maximum proportion in which each hospital can improve observed production and quality levels without increasing its consumption of inputs. For example, the interpretation of the

Table 4
Mean efficiency levels.

Subtype	2009	2010	2011	2012	2013	Mean
Without hospitalization	0.405	0.390	0.428	0.341	0.326	0.377
With hospitalization	0.531	0.447	0.496	0.403	0.498	0.475
Surgical	0.319	0.191	0.327	0.230	0.215	0.260
Mean	0.489	0.411	0.471	0.379	0.453	0.441

Table 5
Mean efficiency scores by departments.

Department	# obs.	2009	2010	2011	2012	2013	Mean eff.
Antioquia	110	0.765	0.635	0.688	0.575	0.659	0.664
Arauca	2	0.282	0.092	0.127	0.198	0.241	0.188
Atlántico	11	0.163	0.115	0.064	0.154	0.293	0.158
Bogotá, D.C.	9	0.566	0.540	0.625	0.411	0.391	0.507
Bolívar	11	0.199	0.183	0.340	0.488	0.574	0.357
Boyacá	55	0.329	0.346	0.478	0.377	0.357	0.377
Caldas	22	0.521	0.431	0.550	0.429	0.486	0.483
Caquetá	5	0.237	0.187	0.124	0.268	0.276	0.218
Casanare	3	0.436	0.358	0.444	0.497	0.468	0.441
Cauca	15	0.249	0.138	0.140	0.064	0.095	0.137
Cesar	16	0.320	0.189	0.402	0.304	0.479	0.339
Chocó	1	0.696	0.508	0.527	0.461	0.581	0.555
Córdoba	9	0.194	0.184	0.180	0.176	0.227	0.192
Cundinamarca	34	0.301	0.249	0.327	0.285	0.391	0.311
Guaviare	1	0.385	0.000	0.149	0.205	0.000	0.148
Huila	30	0.515	0.486	0.430	0.451	0.477	0.472
La Guajira	11	0.220	0.177	0.221	0.270	0.422	0.262
Magdalena	20	0.257	0.209	0.261	0.158	0.401	0.257
Meta	11	0.611	0.276	0.424	0.263	0.344	0.384
Nariño	48	0.562	0.578	0.540	0.377	0.496	0.510
Norte de Santander	10	0.472	0.268	0.341	0.175	0.235	0.298
Putumayo	2	0.265	0.150	0.304	0.237	0.358	0.263
Quindío	8	0.505	0.465	0.444	0.422	0.470	0.461
Risaralda	12	0.440	0.336	0.360	0.214	0.205	0.311
Santander	62	0.641	0.521	0.548	0.397	0.453	0.512
Sucre	19	0.353	0.182	0.296	0.175	0.223	0.246
Tolima	31	0.433	0.447	0.499	0.357	0.512	0.450
Valle del Cauca	34	0.447	0.338	0.513	0.444	0.482	0.445
Mean	21.5	0.489	0.411	0.471	0.379	0.453	0.441

coefficient 0.319 in the table for year 2009 for surgery hospitals should be that this group of hospitals was able to increase its adjusted production level (O_1) and number of satisfied patients (Q_1) by an average of 31.90%, while simultaneously reducing the variables associated with quality (Q_2 to Q_5) by the same proportion. All these improvements should be possible without the need to use more inputs than those observed.

⁹ From a methodological perspective see, for instance, Battese et al. [90], Balaguer-Coll et al. [91], from an empirical point of view see, for instance, Tortosa-Ausina [92].

The average efficiency during the period was 0.441 (see [Tables 4 and 5](#)), which can be considered relatively high according to the interpretation described in the previous paragraph. From a management point of view, high levels of inefficiency imply there is leeway to increase the quantity and quality of services without consuming more inputs. Whether the weighted production depends on how many patients visit a center and with which conditions, an aspect relatively beyond hospital's control (although prevention campaigns can play an important role), the potential for improvement in the dimensions of quality (Q_1 to Q_5) seems clear, even assuming that the observed production and inputs levels are maintained.

A secondary aim of this research is to quantify avoidable referrals to higher level hospitals (measured by variable Q_5). This is a complex question, the response to which would probably require fieldwork in collaboration with medical professionals. However, the methodology used allows us to make an initial approximation. As indicated in the previous paragraphs, the results point to a potential average reduction of 44.1%. This high percentage indicates that the system is probably bearing higher costs than it actually should, as well as affecting effectiveness and quality further up the healthcare system.

Results by hospital type show that inefficiency is greater in centers with inpatient beds and without operating rooms (0.475, see [Table 4](#)). Therefore, the largest group in the sample is also the least efficient, a circumstance that holds throughout the entire period. The most efficient hospitals, with an efficiency coefficient of 0.260, are those that have operating rooms, which is also the case for all the years analyzed. These results seem to evidence for the presence of scope economies, as hospitals offering a more complete catalog of services appear to take advantage of the use of fixed inputs in a more consistent way. This situation is dependent on the environmental conditions as big hospitals located in urban areas, or closer to them, can take advantage of the potential scope economies.

As shown in [Table 5](#), the departments with the most efficient hospitals were Cauca (0.137) and Guaviare (0.148), followed by the Atlántico (0.158) and Arauca (0.188), while the least efficient were Antioquia (0.664), Chocó (0.555) and Santander (0.512). Bogotá D.C. also had a high level of inefficiency (0.507).

The differences among departments might be related to the existence of market and government failures. First of all, there is a higher concentration of public hospitals in urban areas in the departments whose economic performance is better—including lower unemployment rates. This asymmetric spatial allocation of public hospitals around the country could affect the spending level per hospital in each department which, in turn, increases inefficiencies. The market failure affects the provision of health services on less accountable levels related to adverse selection and moral hazard¹⁰ in affiliation due to incomplete information on health status, especially when there is not a well established network of health providers.¹¹ Therefore, departments with a

greater number of public hospitals with a less developed network of health provision have lower efficiency scores compared with other departments with fewer public hospitals with better networks of health provision. This would increase costs in health provision in each department for a given level of outputs. Then, if a public hospital is in a better developed network then it could be able to reduce the effect of market failures associated with moral hazard and adverse selection, improving efficiencies, in comparison with a public hospital located in less developed networks for a given level of outputs. This is particularly relevant for the departments of, for instance, Cauca, Guaviare, Atlántico, and Arauca which have better efficiency scores in comparison to Antioquia, Santander and Bogotá.

Second, there might be government failure when trying to correct market failures generated by asymmetric spatial concentration of public hospitals. After the regulatory change,¹² departments have a greater deal of autonomy in allocating public resources to improve coverage using demand subsidies. Nowadays local governments oversee and allocate subsidies to improve health coverage for low income individuals. This new mechanism of allocating public resources may have increased the interest of private parties to capture public interest of improving health provision. This affects efficiency scores regardless of both the number of hospitals, and how well they are connected in a network of health providers. Then, new policies aimed at fostering coverage might have increased the incentive of capturing local government institutions of private parties,¹³ boosting the spending of public hospitals due to corruption in public resources management for a given level of outputs. This seems to be case of the provinces of Casanare, Huila and Chocó which have according to 2013/14 transparency index a medium and high risk of corruption¹⁴ and have also lower levels of efficiency.

4.1.2. On the sources of inefficiencies: slacks

The methods used allowed us to measure the potentially achievable reductions for each input category. In other words, the reductions that could be achievable in each input additionally to the potential improvement for the good and bad outputs shown in [Tables 4 and 5](#). This information is obtained from the slacks in linear programming problem (7). Results are shown in [Table 6](#).

The items with the highest reduction potential (above 20% on average) are those associated with personnel and medicine. Reductions in the personnel category are 21.19% and 23.06% for administrative and medical care personnel, respectively, which must be interpreted with certain caution. The healthcare sector must have a permanent operating capacity to attend and deal with unpredictable events such as accidents, epidemics, and so on; this requirement can lead to excess resources on average, especially in personnel.

However, the data indicate a growing tendency to inefficiency in personnel costs for the period, for which it would be advisable to reduce these areas or, at the very least, limit their growth, since there are

¹⁰ Asymmetric information concerning present and future patients' health status is a major source of inefficiency for public hospitals, because it increases operational costs. One way in which this happens is moral hazard. According to Cutler and Zeckhauser [93], moral hazard causes both over-consumption and over-provision of health services, because there is an incentive to spend more (when offering a generous insurance), and because there is an incentive to manipulate consumption to induce a demand of services that is not required. Therefore, for a given level of health status, a public hospital would spend more resources than the necessary to achieve it. Moral hazard refers to a fundamental trade-off between risk sharing, and incentives to deviate from cost-efficient behavior. With more generous insurance programs to reduce the risk of having a negative outcome on health status, there are higher chances of over consumption and over service provision. These deviations are a source of inefficiency for public hospitals.

¹¹ Because there exists an uneven distribution of health providers in rural areas, which in turn reduces coverage and quality in health services provision, the Ministry of Health and Social Protection has supported an organization known as Network of Health Providers (NHP). According to the Ministry of Health and Social Protection [94], a NHP is an articulated set of either or both private and public health care providers located in a geographical area that is defined by the integrated model of health care provision in Colombia. As an organization, it shares resources to provide access to health care through

(footnote continued)

proper, effective, and continued health services. This type of organization has also been advocated by Act. 1438, which defines a NHP as a set of organizations that manages to set up agreements to provide an efficient, equitable, complete, and continuous health service. The concept is built upon a network framework developed by both the Pan American Health Organization (PAHO) and The World Health Organization (WHO) in 2010, according to which a NHP is a set of linked organizations willingly open to be audited in terms of resources management and health outcomes, whose main purpose is to share resources to provide equitable, and comprehensive health care services to a defined population [95].

¹² Law 100 in 1993.

¹³ We refer here to a non-productive illegal rent seeking activity that increases both private benefits and social costs. According to Lambsdorff [96], it is corruption when it trespasses laws or rules that have been set up to protect social benefits. It is usually conceived in a principal-agent model when the actions of the agent are meant to increase its own benefit, without considering the principal's benefit.

¹⁴ The measurement is known as an index that accounts for the transparency of public institutions. The measurement of 2013 is available at <http://indicedetransparencia.org.co/ITD/Gobernaciones>.

Table 6

Additional reductions achievable in inputs after improving the good and bad outputs, average 2009–2013 (%).

Year	Descriptive statistics	Spending on purchase goods and services (I_1)	Spending on medicines (I_2)	Spending on administration personnel (I_3)	Spending on assistance personnel (I_4)
2009	Mean	14.39	13.86	19.16	22.81
	Std. Dev.	21.99	22.86	24.10	25.27
2010	Mean	7.78	14.24	21.49	22.68
	Std. Dev.	17.54	21.59	25.36	26.13
2011	Mean	16.18	27.95	21.92	22.68
	Std. Dev.	24.31	29.31	25.21	24.94
2012	Mean	5.77	20.32	20.27	21.74
	Std. Dev.	14.95	25.95	23.28	24.86
2013	Mean	9.86	26.59	23.13	25.38
	Std. Dev.	19.22	27.43	23.38	26.63
Total	Mean	10.80	20.59	21.19	23.06
	Std. Dev.	20.24	26.25	24.30	25.59

actually considerable margins for reduction. An alternative option may be to use some of the excess personnel spending to offer new services, open new primary care centers to improve coverage, or to raise current quality levels.

As an illustration, there are thirteen hospitals in which spending on administrative personnel is higher than on nursing staff, whereas the opposite would seem more logical. In these cases, the potential to reduce spending on administrative personnel will be much higher than in other hospitals with similar levels of production and quality, but with fewer administrative personnel.

Regarding spending on medicines, we estimated a potential average reduction of 20.59%. Note that the potential reduction in this area increases considerably after 2011, almost doubling by 2013. This indicates a growth of inefficiency in managing medicine purchases. It would therefore be advisable to analyze the causes of this increase and introduce remedial measures into the system. Procurement of goods and services is the area showing the most moderate potential reduction (10.80%); no clear rising or falling tendency was identified and it is apparently the most controlled area.

Table 7 shows that the departments with the lowest potential for reducing their expenditure were Cauca (6.94%), the Atlantic (8.39%) and Risaralda (11.02%), suggesting that they managed their spending more efficiently. At the other extreme, the departments with the highest excess costs were Antioquia (29.99%), Chocó (40.97%) and the capital, Bogotá D.C. (37.82%).

The number of patients treated is often a variable relatively out of the control of an hospital, although it could be sometimes reduced by prevention and vaccination campaigns. Hence, it is interesting to quantify the maximum improvements that could be achieved in each of the quality variables, as shown in Tables 8 and 9. These values are obtained from the optimal values of the program 7. Specifically, for the variable satisfied patients (Q_1) it is obtained from the expression $y_m^{ot} (1 + \beta^*) + s_m^*$, where s_m^* is the slack variable of the constrain associated with the good output. For the rest of the quality variables, the maximum potential improvements are obtained from the expression $b_h^{ot} (1 - \beta^*) - s_h^*$, where s_h^* is the optimum value of the slack variable of the constrain associated with each bad output. The results show that during the 2009–13 period the quality variable with the greatest potential for improvement was patient satisfaction (Q_1), standing at 143.283% per year. The variables total wait (Q_2), total waiting in emergencies' room (Q_3), and patients referred to higher levels (Q_5) stand with a potential improvement between 47 and 50%. In this case, it should be highlighted that these quality variables are often associated with a more objective criteria than Q_1 confirming again the existence of considerable deficiencies in the quality levels of the health services

Table 7

Additional reductions achievable in inputs after improving the good and bad outputs by departments, average 2009–2013.

Department	Spending on purchase goods and services (I_1)	Spending on medicines (I_2)	Spending on administration personnel (I_3)	Spending on assistance personnel (I_4)	Mean
Antioquia	13.86	26.24	24.46	42.69	29.99
Arauca	19.29	15.79	26.00	7.71	15.30
Atlántico	8.10	9.74	16.76	3.67	8.39
Bogotá, D.C.	20.08	45.51	44.29	39.62	37.82
Bolívar	13.17	19.58	23.63	31.06	23.70
Boyacá	4.14	15.15	21.27	22.54	17.13
Caldas	2.69	36.09	32.82	9.52	18.13
Caquetá	22.96	2.30	24.08	30.91	22.23
Casanare	23.56	16.94	22.64	19.33	20.36
Cauca	13.02	11.95	3.01	3.36	6.94
Cesar	13.07	10.59	19.84	21.09	17.14
Chocó	25.66	4.36	52.57	61.13	40.97
Córdoba	9.53	4.65	27.84	25.61	18.65
Cundinamarca	8.07	27.79	3.31	13.50	13.24
Guaviare	30.00	7.99	29.13	32.61	26.47
Huila	11.02	32.40	13.79	9.43	15.21
La Guajira	23.38	6.02	15.38	18.12	16.21
Magdalena	9.83	10.96	14.55	26.35	17.61
Meta	14.50	8.53	20.79	13.48	14.16
Nariño	17.19	39.92	14.56	13.04	19.55
Norte de Santander	9.79	6.17	11.52	26.17	15.96
Putumayo	36.04	0.00	17.31	17.09	17.51
Quindío	0.00	27.26	14.98	14.39	14.20
Risaralda	0.89	24.35	21.34	4.26	11.02
Santander	5.64	16.22	29.42	27.07	21.08
Sucre	9.72	11.16	12.05	15.75	12.89
Tolima	18.71	9.80	11.99	15.10	14.14
Valle del Cauca	3.83	27.19	27.10	17.75	18.72
Mean	10.80	20.59	21.19	23.06	19.74

Table 8

Total annual improvement achievable in quality (%) after improving the good and bad outputs by year, average.

Year		Satisfied patients	Total wait	Total waiting in emergencies' room	Re-admission within 20 days	Patients referred to higher levels
		(Q ₁)	(Q ₂)	(Q ₃)	(Q ₄)	(Q ₅)
2009	Mean	185.338	53.469	57.926	36.678	53.337
	Std. Dev	361.826	34.979	35.192	44.161	33.614
2010	Mean	108.335	49.709	43.414	23.514	43.409
	Std. Dev	371.118	33.933	32.344	33.719	30.938
2011	Mean	116.280	52.350	49.847	27.007	49.368
	Std. Dev	246.259	31.411	31.679	36.157	29.999
2012	Mean	120.790	41.766	38.699	25.707	41.612
	Std. Dev	455.697	28.228	28.414	34.272	28.601
2013	Mean	195.672	51.710	48.120	28.515	49.240
	Std. Dev	442.896	32.295	32.587	39.288	31.433
Total	Mean	145.283	49.801	47.601	28.284	47.393
	Std. Dev	384.470	32.504	32.739	37.960	31.237

provided by the hospitals. The number of patients re-admitted within 20 days (Q_4) could improve by approximately 28%. Since this variable is probably the most directly related to the technical quality of care it is not a negligible improvement despite being the quality variable with the least potential for improvement. The detailed analysis by years does not reveal any remarkable trend (see Table 8). The results by department (see Table 9), show that the departments with a better performance in average levels of quality have been Guaviare, Cauca and Arauca all of them with potential improvements lower than 25%. On the other hand,

Table 9

Total improvement achievable in quality (%) after improving the good and bad outputs by department, average 2009–2013.

Department	Satisfied patients	Total wait	Total waiting in emergencies' room	Re-admission within 20 days	Patients referred to higher levels	Mean
	(Q ₁)	(Q ₂)	(Q ₃)	(Q ₄)	(Q ₅)	
Antioquia	183.799	70.939	68.556	46.585	67.090	87.394
Arauca	18.781	23.260	21.052	23.106	29.606	23.161
Atlántico	151.087	24.461	22.659	8.272	20.294	45.355
Bogotá, D.C.	119.056	72.375	57.139	41.147	75.032	72.950
Bolívar	265.767	42.215	38.706	22.139	39.582	81.682
Boyacá	75.529	42.749	25.830	9.152	41.331	38.918
Caldas	206.491	57.450	48.026	49.897	51.120	82.597
Caquetá	99.926	33.845	26.837	16.344	21.636	39.718
Casanare	55.007	46.928	46.877	46.558	44.052	47.884
Cauca	18.205	15.346	17.135	20.741	18.880	18.061
Cesar	155.950	35.610	47.707	27.400	37.378	60.809
Chocó	96.941	55.462	55.462	45.302	57.924	62.218
Córdoba	251.539	45.697	39.938	28.559	34.132	79.973
Cundinamarca	47.053	29.490	23.241	20.052	19.212	27.810
Guaviare	14.763	22.647	14.763	14.763	14.763	16.340
Huila	78.667	49.687	48.658	24.073	52.587	50.734
La Guajira	28.437	29.605	38.161	16.455	26.982	27.928
Magdalena	165.633	29.567	32.955	20.250	27.637	55.208
Meta	265.419	41.151	42.369	21.759	40.002	82.140
Nariño	202.635	54.638	52.008	7.264	53.130	73.935
Norte de Santander	41.214	31.765	43.009	7.862	33.481	31.466
Putumayo	83.256	31.611	26.947	0.000	26.276	33.618
Quindío	112.427	56.581	55.187	44.782	59.245	65.644
Risaralda	117.062	46.951	32.869	45.321	35.186	55.478
Santander	139.788	55.498	55.266	23.893	54.351	65.759
Sucre	117.440	27.242	37.536	13.104	30.921	45.249
Tolima	128.248	49.277	57.566	45.971	47.341	65.681
Valle del Cauca	115.955	51.683	50.847	39.285	50.187	61.591
Mean	145.283	49.801	47.601	28.284	47.393	63.672

four departments (Bolívar, Meta, Caldas and Antioquia) have the worst results with average potential increases of quality over 80%.

4.2. Temporal analysis

Table 10 reports the results for the evolution of total factor productivity (i.e., change in hospital performance), defined in Section 2 as

Table 10

Total factor productivity, efficiency and technological changes. Mean values.

Hospital Type	2009–2010	2010–2011	2011–2012	2012–2013	Mean
<i>Surgical</i>					
Efficiency Change	1.079	0.928	1.057	0.970	1.008
Technological Change	0.903	1.070	0.960	1.007	0.985
Total Factor Productivity	0.967	0.992	1.013	0.971	0.986
<i>With Hospitalization</i>					
Efficiency Change	1.074	0.979	1.087	0.958	1.024
Technological Change	0.906	1.019	0.932	1.074	0.983
Total Factor Productivity	0.971	0.987	1.003	1.012	0.993
<i>Without Hospitalization</i>					
Efficiency Change	1.044	0.977	1.069	1.009	1.025
Technological Change	0.898	0.986	0.984	1.021	0.972
Total Factor Productivity	0.922	0.951	1.031	1.002	0.976
Efficiency Change (Mean)	1.071	0.972	1.081	0.966	1.022
Technological Change (Mean)	0.905	1.021	0.942	1.059	0.982
Total Factor Productivity (Mean)	0.964	0.983	1.008	1.005	0.990

the global Malmquist-Luenberger index (*GML*), along with the decomposition between efficiency change (*EC*) and technological change (also referred to as best-practice change, *BPC*).¹⁵

A value higher than unity means that hospital performance (total factor productivity) improved in the period considered, whereas a lower value indicates deterioration in its performance. Focusing on the components of performance change, the catching up component (efficiency change, *EC*) shows how a hospital's proximity to the efficient good practices frontier changed for the period. A value greater than one means that during the period the hospitals were closer to the best observed practices. Technological change (best practice change, *BPC*) measures frontier shifts, which serves as a reference for measuring the efficiency of hospitals off the efficient frontier. Once again, a value greater than one indicates an improvement or, in other words, that best practices were able to achieve more output with the same or less input. A value lower than one implies the opposite.

The lower right cell in Table 10 shows that the hospitals' performance change (productivity change) in the period was 0.99. This means that the productivity for the period worsened, on average, by 1%. The most deterioration, by 2.4%, was seen in hospitals with no inpatient capacity. An analysis of the determining components of productivity reveals that catch up was slightly positive in the period. The index has a value of 1.022, but the best practices (technological change) worsened, showing a coefficient of 0.982. As a result, the changes in efficiency were due in part to a worsening of efficient units, which led to a drop in the total factor productivity for the period. Nevertheless, a detailed analysis shows that for the periods 2011–2012 and 2012–2013, productivity tentatively started on the road to recovery after an initially poor performance. Taking into account that the static analysis revealed notable levels of inefficiency, and the temporal analysis showed small

¹⁵ As pointed out in the theoretical section of this paper, remember that the technical change described here can signal both progress or regress, but it quantifies all kinds of biased or neutral estimations of technical change.

Table 11
Change in hospital performance and its decomposition, Colombian departments, average (2009–2013).

	Efficiency change	Technological change	Change in hospital performance
Antioquia	1.027	0.988	1.002
Arauca	1.013	1.016	1.017
Atlántico	0.992	0.994	0.966
Bogotá, D.C.	1.043	1.003	1.037
Bolívar	0.950	1.009	0.954
Boyacá	1.018	0.959	0.955
Caldas	1.017	0.973	0.977
Caquetá	1.004	0.992	1.000
Casanare	0.999	0.992	0.988
Cauca	1.041	1.004	1.035
Cesar	0.988	0.977	0.954
Chocó	1.021	0.978	0.992
Córdoba	1.010	0.993	0.985
Cundinamarca	0.999	0.988	0.978
Guaviare	1.103	0.935	1.041
Huila	1.019	0.965	0.971
La Guajira	0.974	0.985	0.953
Magdalena	0.992	1.007	0.982
Meta	1.078	0.976	1.034
Nariño	1.030	0.979	0.989
Norte de Santander	1.058	0.973	1.014
Putumayo	0.992	1.016	1.000
Quindío	1.006	0.965	0.972
Risaralda	1.054	0.960	1.002
Santander	1.055	0.970	1.007
Sucre	1.052	0.994	1.035
Tolima	1.011	1.004	0.999
Valle del Cauca	1.010	0.979	0.976
Mean	1.022	0.982	0.990

changes in productivity, we can deduce that there is no clear improvement trend in productivity and in the components that contribute to an improvement in performance for level 1 hospitals; rather, inefficiency levels are consolidated.

The departments with higher increases in total factor productivity were Guaviare, with 4.1%, Sucre and Cauca with 3.5%, and the capital Bogotá, D.C. (3.7%) (see Table 11). Note also that these three departments and the capital are the only ones where total factor productivity and its two components improved. This means not only that efficiency improved, but also that the frontier area with which they were compared improved as well. Guaviare and Cauca also appeared in the list of more efficient departments in the static analysis. The temporal analysis confirmed that they also have the most active productivity improvement. The case of Bogotá D.C., the country's capital, is particularly notable for its economic importance and population concentration. Although the static analysis showed room for improvement in efficiency results, the evolution during the period of productivity was positive, although much remains to be done to improve efficiency. The departments with the worst evolution were La Guajira, Bolívar, Cesar, and Boyacá—all with a drop in the total factor productivity of more than 4%.

As mentioned in section 3.2, Q_2 and Q_5 are variables that can be interpreted in a contrary way to the one proposed in this article, as well as being a non-standard choice of variables in the literature. For this reason, a re-estimation of the model was carried out to check the robustness of the results before changes in the specification of these variables. Specifically, the calculations have been carried out again, eliminating them. The two variables have been eliminated at the same time instead of eliminating them one at a time because the impact on the results will always be greater the more variables are modified as it is a non-parametric technique such as the DEA based on linear programming. The results obtained after this estimation (see Table 12) shows very small global differences in the results, both in the Total Factor Productivity as well as and in its explanatory components. It is

also appreciated that the results have not changed hardly in quantity or interpretation. The only exception is for the 2012–2013 period, in which contradictory results are obtained with the new estimation with respect to the initial one. Therefore, we consider that the results obtained originally can be considered as robust given the small changes observed after the removing simultaneously both variables.

4.3. Bipartite decomposition of the contributions to performance change

The results corresponding to the bipartite decomposition explained in Section 4.3 are reported in Fig. 1. They refer to the changes in hospital performance for the entire period, i.e., between 2009 and 2013.¹⁶ Fig. 1 is split into four sub-figures, those in the upper and lower panels corresponding to the two directions of the sequential order of the analysis. Specifically, the upper panel (sub-Fig. 1 a, corresponding to the contribution of efficiency change, and 1.b, corresponding to the contribution of best practice gap change) refers to the decomposition in Equation (11), and the lower panel (sub-Fig. 1 c, corresponding to the contribution of best practice gap change, and 1.d, corresponding to the contribution of efficiency change) represents the decomposition in Equation (13).

As indicated in Section 2.2, the analysis was carried out using kernel smoothing (Gaussian kernel), and considering several bandwidth choices, both to acknowledge this rich literature and also to increase the robustness of the results. Therefore, whereas results in Fig. 1 were obtained using a global (direct plug-in) bandwidth [32], in Fig. 2 we also report findings considering a bandwidth that smooths locally—i.e., depending on the data in each region of the distribution [33].

The density in Fig. 1 a points to greater disparity in the contributions of efficiency to hospital performance change between 2009 and 2013 (i.e., to the global non-radial Malmquist index). This is the counterfactual hospital performance change attributable to changes in efficiency only (gml^{EC}), i.e. we isolate the effect of changes in efficiency. The remarkable heterogeneity is revealed by the long tails of the density (particularly the upper tail), as well as the existence of several bumps (corresponding to more hospitals). However, given that probabilistic mass tends to concentrate more tightly around unity, the final average effect is slightly positive, as revealed by the vertical solid line in Fig. 1 a (which is the average of gml^{EC} between 2009 and 2014).

The upper panel in Fig. 1 displays the contribution of best practice change (Fig. 1b). However, in this case its effect is not isolated, since we combine it with the effect of efficiency change ($gml^{EC \times BPC}$). The contribution of best practice change (technological change) is, on average, detrimental, since the average hospital performance change—represented by the dashed vertical line in Fig. 1 b—shifts leftwards, leading to an overall negative performance change. The negative contribution of best practice change is due to a high number of hospitals in the (0.5, 1.0) interval, revealed by the high probabilistic mass. In contrast, several hospitals also excel in terms of best practice change, since there is also probabilistic mass in the (1.0, 2.0) interval, but it is much lower than that corresponding to efficiency change. Fig. 2a and 2 b report the results corresponding to the local bandwidths, pointing in the same direction and, therefore, corroborating the results, although the trends are much smoother.

The lower panels of both Fig. 1 (global bandwidth) and 2 (local bandwidth) report results analogous to the upper panel, but first isolating the best practice change effect. In the case of the global bandwidth (Fig. 1c and 2 c), the densities are tighter compared to the effect of isolating the efficiency change, indicating fewer disparities among hospitals in terms of best practice gap change. However, the existence of probabilistic mass in the vicinity of 2 (Fig. 1c) indicates that some (very few) hospitals excel in this particular regard. All in all, the results

¹⁶ Results for each of the sub-periods analyzed are not reported here due to space restrictions, but are available upon request.

Table 12Total factor productivity, efficiency and technological changes without Q₂ and Q₅ (robustness analysis). Mean values.

Hospital Type	2009–2010	2010–2011	2011–2012	2012–2013	Mean
<i>Surgical</i>					
Efficiency Change	1.107	0.912	1.096	0.977	1.023
Technological Change	0.894	1.085	0.941	0.983	0.976
Total Factor Productivity	0.976	0.986	1.026	0.960	0.987
<i>With Hospitalization</i>					
Efficiency Change	1.081	0.975	1.071	1.021	1.037
Technological Change	0.928	1.029	0.945	0.967	0.967
Total Factor Productivity	0.991	0.994	1.003	0.982	0.992
<i>Without Hospitalization</i>					
Efficiency Change	1.258	0.984	1.069	1.009	1.080
Technological Change	0.887	0.968	0.990	0.982	0.957
Total Factor Productivity	1.025	0.946	1.050	0.983	1.001
Total Efficiency Change (Mean)	1.106	0.968	1.074	1.014	1.041
Total Technological Change (Mean)	0.918	1.029	0.950	0.971	0.967
Total Factor Productivity (Mean)	0.994	0.987	1.012	0.979	0.993

presented in the upper and lower panels of both Figs. 1 and 2 would suggest that the paths followed by different hospitals to achieve particular levels of performance differ greatly, and the efficiency gains of some hospitals contribute to offset the overall negative impact on performance of best practice change.

5. Conclusions

After the series of reforms introduced in the year 1993, this paper assessed the evolution of the efficiency in the Colombian public health sector. More specifically, one of our main interests was to determine to what extent the excessive amount of patients referred by primary care centers to high level hospitals influenced their level of efficiency, having into account the existent capacity to provide services with good quality.

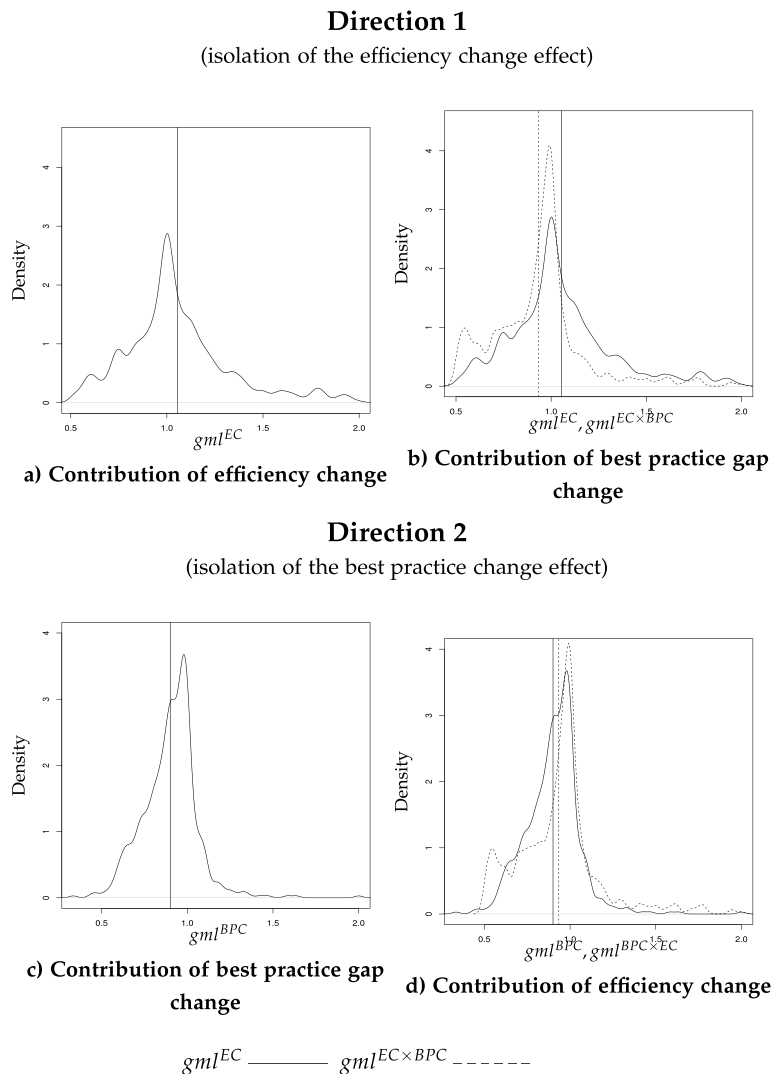
The static analysis of productivity showed an average level of efficiency of 0.441 during 2009–2013. The highest levels of efficiency were achieved by hospitals with operating rooms, while the highest level of inefficiency was found in hospitals with inpatient beds but no operating rooms. In addition, the inputs with the highest reduction potential were those associated with personnel and medicines. Likewise, the results also suggest a reduction in patient referrals to more complex hospitals of 44.1%. The public hospital system may therefore be bearing more costs than it should, which also affects the effectiveness and quality further up the health system. In turn, the temporal analysis established that productivity for the period fell by 1%, with a notable reduction for hospitals with no inpatient capacity. Our evaluation of the components of productivity determined that the best practices frontier worsened, although the detailed analysis for the periods 2011–2012 and 2012–2013 showed a slight recovery. Considering the elevated levels of inefficiency identified, as well as the small changes in productivity found in the temporal analysis, there is no clear tendency toward improvement in either productivity or its components, but rather, a consolidation of inefficiency levels in the public hospitals studied.

The analysis by geographical location showed that the most efficient hospitals are located in the departments of Cauca and Guaviare, while the most inefficient are in the departments of Antioquia and Chocó. Likewise, hospitals with a lower potential reduction in spending are Cauca and the Atlantic, while the departments with the highest excess in costs are located in the department of Chocó. The evolution of productivity shows that the hospitals in the departments of Guaviare and Cauca are the same as in the static analysis, and are therefore the most active in improving productivity. The hospitals with the poorest evolution are located in the departments of La Guajira, Bolívar, Cesar and Boyacá. Finally, the hospitals located in Bogotá, the capital, have clear potential to improve their efficiency levels, even when they register improvements in their productivity evolution.

With reference to Law 100 (1993), the results reflect deficient performance in public hospital efficiency. Taking into account that the reform was designed to trigger competition in order to achieve universal coverage based on interaction among insurers, healthcare service providers, and healthcare system users, the organizational autonomy the new social state companies (ESEs) gives to public hospitals has not had a conclusive effect in consolidating a positive tendency in productivity and efficiency. Compared with the growth in demand created by replacing supply subsidies due to the payment of capitation units to insurers, public hospitals continue to face difficulties in their personnel and medicine expenditures, items in which the results reveal potential for considerable improvements in resource management. On this point, and echoing the healthcare statute of 2015, it is important to highlight the relevance of institutional environment factors such as a clearly defined national pharmaceutical policy (Art. 23) coordinated with public healthcare policy (Art. 20) and the obligation to guarantee efficient availability of healthcare services in marginalized areas (Art. 6 and Art. 24) aimed to provide favorable conditions for public hospital management. Likewise, internal management tools and efficiency indicators must be in place in accordance with the challenge of market competition, which include standard institutional practices within the framework of a quality model for service aimed at the user. Suitable information systems must be created for hospital administrations, and any profits should be reinvested in improving structural capacity, with the resulting positive effects on hospital efficiency. This is congruent with related findings in the literature calling for guaranteed contractual designs in which the objectives of a market reform are linked to the aims of public sector hospital administration.

There is an asymmetrical spatial concentration of public hospitals in the country which amplifies the negative effects of having a great number of hospitals with weak networks of health providers, and weak local governments with a high and medium risk of corruption. A department with a great number of public hospitals poorly connected in a network of health providers, and with private interests capturing public institutions that allocate demand subsidies to low income individuals, will likely have a lower level of inefficiency for a given level of outputs. This seems to be the case of departments of Antioquia, Santander and Bogotá which have a greater number of hospitals, and the case of Chocó, Huila and Casanare that have a relative high risk of corruption, compared to the others department in the sample. An extension of this research will test how the existence of public health networks and lack of public institutions transparency allocating demand subsidies affect efficiency scores.

For the policy making, useful implications can be deduced from these results. Firstly, compared with the prevalent situation before the reform, the new regulatory framework tried to increase coverage, quality and efficiency by introducing more market competition. The



Notes: All figures contain densities estimated via kernel smoothing for the different components of the bipartite decomposition in expression (9), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in hospital performance (*gml*). The vertical lines in each plot represent the average for each component of the decomposition. Densities were estimated using a Gaussian kernel and the Sheather and Jones (1991) plug-in bandwidth (global bandwidth).

Fig. 1. Kernel density plots of the bipartite decomposition of hospital performance improvement, DEA, 2009–2013 (global bandwidth). Notes: All figures contain densities estimated via kernel smoothing for the different components of the bipartite decomposition in expression (9), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in hospital performance (*gml*). The vertical lines in each plot represent the average for each component of the decomposition. Densities were estimated using a Gaussian kernel and the Sheather and Jones [32] plug-in bandwidth (global bandwidth).

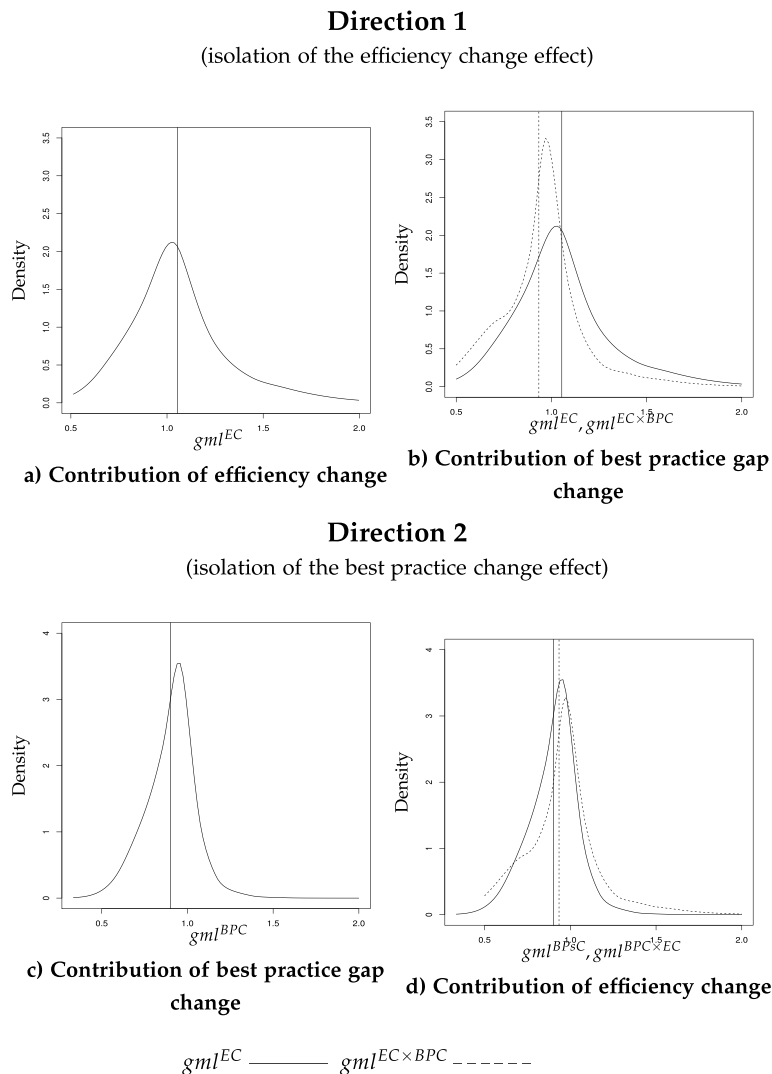
problem is that the supposed increase in competition has not been completely achieved because of the lack of a good design model of control of the EPSs, as the poor quality in the service does not imply any penalty in the incomes recognized for these organisms. Moreover, the excessive concentration in specific areas has limited in a substantial way the real competition. Summing up, there are issues to be solved from the design of the contracts as well as from the geographical reach of the organisms to enhance the level of competition and, as a results of this, the improvement of the efficiency.

The main limitation of this work is that it analyzes only level 1 of Colombian public hospitals. A useful line of future research would be to expand the sample to include higher level and private hospitals so as to provide a more transversal analysis of the Colombian healthcare sector. Specifically, a network-DEA model could be used to assess the effect of inefficiency at one level on higher levels. An additional extension of this research could also include network effects and corruption risks on

efficiency score estimation to control for asymmetric spatial concentration of health providers around the country.

Acknowledgements

This paper is part of a wider research project founded by the Inter-American Development Bank (IDB) entitled *Impactos de los obstáculos institucionales para incrementar la eficiencia, eficacia y transparencia de la gestión de la inversión pública en Colombia* (project reference ATN/Fr12361-CO). The authors gratefully acknowledge financial support and collaboration from the IDB and the Ministry of Health and Social Protection of the Republic of Colombia, which provided access to the database used in this research. We also thank the helpful comments by Mario Loterszpil, which contributed to improve the overall quality of the article.



Notes: All figures contain densities estimated using kernel smoothing for the different components of the bipartite decomposition in expression (11), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in educational performance (*gml*). Densities were estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel was chosen.

Fig. 2. Kernel density plots of the bipartite decomposition of hospital performance improvement, DEA, 2009–2013 (local bandwidth). Notes: All figures contain densities estimated using kernel smoothing for the different components of the bipartite decomposition in expression (11), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in educational performance (*gml*). Densities were estimated using local likelihood methods [33], and a Gaussian kernel was chosen.

References

- [1] Scheel H. Undesirable outputs in efficiency valuations. *Eur J Oper Res* 2001;132(2):400–10.
- [2] Jahanshahloo GR, Hadi Vencheh A, Foroughi AA, Kazemi Matin R. Inputs/outputs estimation in DEA when some factors are undesirable. *Appl Math Comput* 2004;156(1):19–32.
- [3] Ozcan YA, Merwin E, Lee K, Morrissey JP. Benchmarking using DEA: the case of mental health organizations. *Operations research and health care*. Boston: Springer US; 2005. p. 169–89.
- [4] Hadi Vencheh A, Kazemi Matin R, Tavassoli Kajani M. Undesirable factors in efficiency measurement. *Appl Math Comput* 2005;163(2):547–52.
- [5] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *Eur J Oper Res* 1978;2(6):429–44.
- [6] Prior D. Efficiency and total quality management in health care organizations: a dynamic frontier approach. *Ann Oper Res* 2006;145(1):281–99.
- [7] Navarro-Espigares JL, Torres EH. Efficiency and quality in health services: a crucial link. *Serv Ind J* 2010;31(3):385–403.
- [8] Yang J, Zeng W. The trade-offs between efficiency and quality in the hospital production: some evidence from Shenzhen, China. *China Econ Rev* 2014;31:166–84.
- [9] Arocena P, García Prado A. Accounting for quality in the measurement of hospital performance: evidence from Costa Rica. *Health Econ* 2007;16(7):667–85.
- [10] Hu H-H, Qi Q, Yang C-H. Analysis of hospital technical efficiency in China: effect of health insurance reform. *China Econ Rev* 2012;23(4):865–77.
- [11] Alonso JM, Clifton J, Díaz-Fuentes D. The impact of New Public Management on efficiency: an analysis of Madrid's hospitals. *Health Pol* 2015;119(3):333–40.
- [12] Matranga D, Sapienza F. Congestion analysis to evaluate the efficiency and appropriateness of hospitals in Sicily. *Health Pol* 2015;119(3):324–32.
- [13] Fidler AH, Haslinger RR, Hofmarcher MM, Jesse M, Palu T. Incorporation of public hospitals: a “Silver Bullet” against overcapacity, managerial bottlenecks and resource constraints? *Health Pol* 2007;81(2–3):328–38.
- [14] Chen Z, Barros CP, Hou X. Has the medical reform improved the cost efficiency of Chinese hospitals? *Soc Sci J* 2016;53(4):510–20.
- [15] Sarmiento A, Castellanos W, Nieto A, Alonso C. Análisis de eficiencia técnica de la red pública de prestadores de servicios dentro del sistema general de seguridad social en salud. *Revista Gerencia y Políticas de Salud* (Pontificia Universidad Javeriana) 2006;5(11):70–95.
- [16] McPake B, Yepes FJ, Lake S, Sánchez LH. Is the Colombian health system reform improving the performance of public hospitals in Bogotá? *Health Pol Plann* 2003;18(2):182–94.
- [17] Pinzón-Martínez MJ. Medición de eficiencia técnica relativa en hospitales públicos de baja complejidad mediante la metodología Data Envelopment Analysis (DEA) vol. 245. Archivos de Economía República de Colombia. Departamento Nacional de Planeación; 2003.
- [18] Tamayo N, Maldonado A. Estudio integral de eficiencia de los hospitales públicos [Technical report] 2007.
- [19] Caves DW, Christensen LR, Diewert WE. The economic theory of index numbers and

- the measurement of input, output, and productivity. *Econometrica* 1982;50(6):1393–414.
- [20] Chung YH, Färe R, Grosskopf S. Productivity and undesirable outputs: a directional distance function approach. *J Environ Manag* 1997;51(3):229–40.
- [21] Färe R, Grosskopf S, Noh D-W, Weber W. Characteristics of a polluting technology: theory and practice. *J Econom* 2005;126(2):469–92.
- [22] Sueyoshi T, Goto M. Should the US clean air act include CO2 emission control? Examination by data envelopment analysis. *Energy Pol* 2010;38(10):5902–11.
- [23] Watanabe M, Tanaka K. Efficiency analysis of Chinese industry: a directional distance function approach. *Energy Pol* 2007;35(12):6323–31.
- [24] Oh D-h. A global Malmquist-Luenberger productivity index. *J Prod Anal* 2010;34(3):183–97.
- [25] Pastor JT, Lovell CAK. A global Malmquist productivity index. *Econ Lett* 2005;88(2):266–71.
- [26] Balk BM. Scale efficiency and productivity change. *J Prod Anal* 2001;15(3):159–83.
- [27] Xue M. Note: ranking DMUs with infeasible super-efficiency DEA models. *Manag Sci* 2002;48(5):705–10.
- [28] Färe R, Grosskopf S, Pasurkajr C. Environmental production functions and environmental directional distance functions. *Energy* 2007;32(7):1055–66.
- [29] Luenberger DG. New optimality principles for economic efficiency and equilibrium. *J Optim Theor Appl* 1992;75(2):221–64.
- [30] Emrouznejad A, Parker BR, Tavares G. Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. *Soc Econ Plann Sci* 2008;42(3):151–7.
- [31] Mandal SK, Madheswaran S. Environmental efficiency of the Indian cement industry: an interstate analysis. *Energy Pol* 2010;38(2):1108–18.
- [32] Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *J Roy Stat Soc B* 1991;53(3):683–90.
- [33] Loader CR. Local likelihood density estimation. *Ann Stat* 1996;24(4):1602–18.
- [34] Kumar S, Russell RR. Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence. *Am Econ Rev* 2002;92(3):527–48.
- [35] Färe R, Grosskopf S, Lovell CAK. *Production frontiers*. Cambridge: Cambridge University Press; 1994.
- [36] Quah DT. Empirical cross-section dynamics in economic growth. *Eur Econ Rev* 1993;37:426–34.
- [37] Quah DT. Galton's fallacy and tests of the convergence hypothesis. *Scand J Econ* 1993;95(4):427–43.
- [38] Henderson DJ, Russell RR. Human capital and macroeconomic convergence: a production-frontier approach. *Int Econ Rev* 2005;46(4):1167–205.
- [39] Badunenko O, Romero-Ávila D. Financial development and the sources of growth and convergence. *Int Econ Rev* 2013;54(2):629–63.
- [40] O'Neill L, Rauner M, Heidenberger K, Kraus M. A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Soc Econ Plann Sci* 2008;42(3):158–89.
- [41] Ozcan Y, Luke RD. A national study of the efficiency of hospitals in urban markets. *Health Serv Res* 1993;27(6):719–39.
- [42] Athanassopoulos A, Gounaris C. Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications. *Eur J Oper Res* 2001;133(2):416–31.
- [43] Chirikos TN, Sear AM. Measuring hospital efficiency: a comparison of two approaches. *Health Serv Res* 2000;34(6):1389.
- [44] Hofmarcher MM, Paterson I, Riedel M. Measuring hospital efficiency in Austria—a DEA approach. *Health Care Manag Sci* 2002;5(1):7–14.
- [45] Wan TT, Lin BY-J, Ma A. Integration mechanisms and hospital efficiency in integrated health care delivery systems. *J Med Syst* 2002;26(2):127–43.
- [46] Solà M, Prior D. Measuring productivity and quality changes using Data Envelopment Analysis: an application to Catalan hospitals. *Financ Account Manag* 2001;17(August):219–45.
- [47] Vuori HV. El control de calidad en los servicios sanitarios. Masson S.A., Barcelona: Conceptos y metodología; 1991.
- [48] Banker R. Use of categorical variables in data envelopment analysis. *Manag Sci* 1986;32(12):1613–27.
- [49] Nunamaker TR. Measuring routine nursing service efficiency: a comparison of cost per patient day and Data Envelopment Analysis models. *Health Serv Res* 1983;18(2):183–208.
- [50] Lewin AY. Comments on “Measuring routine nursing service efficiency: a comparison of cost per patient day and Data Envelopment Analysis”. *Health Serv Res* 1983;18(2):206–8.
- [51] Harris J, Ozgen H, Ozcan Y. Do mergers enhance the performance of hospital efficiency? *J Oper Res Soc* 2000;51(7):801–11.
- [52] Valdmanis VG, Rosko MD, Mutter RL. Hospital quality, efficiency, and input slack differentials. *Health Serv Res* 2008;43(5p2):1830–48.
- [53] Weng S-J, Wu T, Blackhurst J, Mackulak G. An extended DEA model for hospital performance evaluation and improvement. *Health Serv Outcome Res Meth* 2009;9(1):39–53.
- [54] Harrison JP, Coppola MN, Wakefield M. Efficiency of federal hospitals in the United States. *J Med Syst* 2004;28(5):411–22.
- [55] Dexter F, O'Neill L. Data Envelopment Analysis to determine by how much hospitals can increase elective inpatient surgical workload for each specialty. *Anesth Analg* 2004;99(5):1492–500.
- [56] Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Manag Sci* 2003;6(4):203–18.
- [57] Hollingsworth B. The measurement of efficiency and productivity of health care delivery. *Health Econ* 2008;17(10):1107–28.
- [58] Pina V, Torres L. Methodological aspects in efficiency evaluation of public hospitals. *Financ Account Manag* 1996;12(1):21–36.
- [59] Dalmiau-Atarrodona E, Puig-Junoy J. Market structure and hospital efficiency: evaluating potential effects of deregulation in a national health service. *Rev Ind Organ* 1998;13(4):447–66.
- [60] Helmig B, Lapsley I. On the efficiency of public, welfare and private hospitals in Germany over time: a sectoral data envelopment analysis study. *Health Serv Manag Res* 2001;14(4):263–74.
- [61] Field K, Emrouznejad A. Measuring the performance of neonatal care units in Scotland. *J Med Syst* 2003;27(4):315–24.
- [62] Ersoy K, Kavuncubasi S, Ozcan YA, Harris II JM. Technical efficiencies of Turkish hospitals: DEA approach. *J Med Syst* 1997;21(2):67–74.
- [63] Sahin I, Ozcan YA. Public sector hospital efficiency for provincial markets in Turkey. *J Med Syst* 2000;24(6):307–20.
- [64] Giokas DI. Greek hospitals: how well their resources are used. *Omega* 2001;29(1):73–83.
- [65] Ouellette P, Vierstraete V. Technological change and efficiency in the presence of quasi-fixed inputs: a DEA application to the hospital sector. *Eur J Oper Res* 2004;154(3):755–63.
- [66] Gerdtham UG, Löthgren M, Tambour M, Rehnberg C. Internal markets and health care efficiency: a multiple-output stochastic frontier analysis. *Health Econ* 1999;8(2):151–64.
- [67] Bjørn E, Hagen TP, Iversen T, Magnussen J. The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA efficiency scores 1992–2000. *Health Care Manag Sci* 2003;6(4):271–83.
- [68] Kirigia JM, Emrouznejad A, Sambo LG, Munguti N, Liambila W. Using Data Envelopment Analysis to measure the technical efficiency of public health centers in Kenya. *J Med Syst* 2004;28(2):155–66.
- [69] Ferrier GD, Valdmanis VG. Do mergers improve hospital productivity? *J Oper Res Soc* 2004;55(10):1071–80.
- [70] Grosskopf S, Margaritis D, Valdmanis V. Comparing teaching and non-teaching hospitals: a frontier approach (teaching vs. non-teaching hospitals). *Health Care Manag Sci* 2001;4(2):83–90.
- [71] Grosskopf S, Margaritis D, Valdmanis V. Competitive effects on teaching hospitals. *Eur J Oper Res* 2004;154(2):515–25.
- [72] Wang BB, Ozcan YA, Wan TTH, Harrison J. Trends in hospital efficiency among metropolitan markets. *J Med Syst* 1999;23(2):83–97.
- [73] Cordero JM, Alonso-Morán E, Nuño-Solinis R, Orueta JF, Arce RS. Efficiency assessment of primary care providers: a conditional nonparametric approach. *Eur J Oper Res* 2015;240(1):235–44.
- [74] Martini G, Berta P, Mullahy J, Vittadini G. The effectiveness–efficiency trade-off in health care: the case of hospitals in Lombardy, Italy. *Reg Sci Urban Econ* 2014;49:217–31.
- [75] O'Neill L. Multifactor efficiency in Data Envelopment Analysis with an application to urban hospitals. *Health Care Manag Sci* 1998;1(1):19–27.
- [76] Kalhor R, Amini S, Sokhanvar M, Lotfi F, Sharifi M, Kakemam E. Factors affecting the technical efficiency of general hospitals in Iran: data Envelopment Analysis. *J Egypt Publ Health Assoc* 2016;91(1):20–5.
- [77] Ozcan YA. Efficiency of hospital service production in local markets: the balance sheet of U.S. medical armament. *Soc Econ Plann Sci* 1995;29(2):139–50.
- [78] Büchner VA, Hinz V, Schreyögg J. Health systems: changes in hospital efficiency and profitability. *Health Care Manag Sci* 2016;19(2):130–43.
- [79] Jola-Sanchez AF, Pedraza-Martinez AJ, Bretthauer KM, Britto RA. Effect of armed conflicts on humanitarian operations: total factor productivity and efficiency of rural hospitals. *J Oper Manag* 2016;45:73–85.
- [80] Silverman BW. *Density estimation for statistics and data analysis*. London: Chapman and Hall; 1986.
- [81] Scott DW. *Multivariate density estimation: theory, practice, and visualization*. New York: Wiley; 1992.
- [82] Li Q, Racine JS. *Nonparametric econometrics: theory and practice*. Princeton and Oxford: Princeton University Press; 2007.
- [83] Henderson DJ, Parmeter CF. *Applied nonparametric econometrics*. Cambridge, MA: Cambridge University Press; 2015.
- [84] Cortés-Martínez AE. La economía de la salud en el hospital. *Rev Geren Polít. Salud* 2010;9(19):138–49.
- [85] Gómez Duque A. Un método rápido para el cálculo de costos hospitalarios. *Rev Fac Med* 1996;44(2):63–9.
- [86] Sarmiento Gómez A, Castellanos Aranguren W, Nieto Rocha A, Alonso Malaver CE. Análisis de eficiencia técnica de la red pública de prestadores de servicios. *Rev Geren Polít. Salud* 2006;5(11):71–95.
- [87] Cabrera-Arana GA, Londoño-Pimienta JL, Bello-Parías LD. Validación de un instrumento para medir calidad percibida por usuarios de hospitales de Colombia. *Rev. Salud Públ.* 2008;10(3):443–51.
- [88] Delgado-Gallego ME, Vázquez-Navarrete ML, de Moraes-Vanderlei L. Calidad en los servicios de salud desde los marcos de sentido de diferentes actores sociales en Colombia y Brasil. *Rev. Salud Públ.* 2010;12(4):533–45.
- [89] Numpaque-Pacabaque A, Rocha-Buelvas A. Modelos SERVQUAL y SERVQHOS para la evaluación de calidad de los servicios de salud. *Rev Fac Med* 2016;64(4):715–20.
- [90] Battese GE, Rao D, O'Donnell CJ. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *J Prod Anal* 2004;21(1):91–103.
- [91] Balaguer-Coll MT, Prior D, Tortosa-Ausina E. Output complexity, environmental conditions, and the efficiency of municipalities. *J Prod Anal* 2013;39(3):303–24.
- [92] Tortosa-Ausina E. An alternative conditioning scheme to explain efficiency differentials in banking. *Econ Lett* 2004;82:147–55.
- [93] Cutler DM, Zeckhauser RJ. *The anatomy of health insurance*. Culyer AJ, Newhouse JP, editors. *Handbook of health economics*, vol. 1. Amsterdam/New York: Elsevier;

2000. p. 563–643 (Part A), chapter 11.
- [94] Ministry of Health and Social Protection. Redes Integrales de prestadores de servicios de salud. Lineamientos para el Proceso de Conformación, Organización, Gestión, Seguimiento y Evaluación. Lineamientos, Ministerio de Salud y Protección Social. Bogotá D.C: Dirección de Prestación de Servicios y Atención Primaria; 2016.
- [95] Pan American Health Organization (PAHO)/World Health Organization (WHO). Redes Integradas de Servicios de Salud. Conceptos, Opciones de Política y Hoja de Ruta para su Implementación en las Américas. La Renovación de la Atención Primaria de Salud en las Américas 4. Washington, D.C: Organización Panamericana de Salud (PAHO); 2010.
- [96] Lambsdorff JG. Corruption and rent-seeking. *Publ Choice* 2002;113(1):97–125.