Socio-Economic Planning Sciences 73 (2021) 100798

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: http://www.elsevier.com/locate/seps

Public-private partnerships in health care services: Do they outperform public hospitals regarding quality and access? Evidence from Portugal

D.C. Ferreira [*], R.C. Marques

CERIS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

ARTICLE INFO

Keywords:
Health policy
Public-private partnerships
Public hospitals
Social performance
Benefit of doubt

ABSTRACT

Public-private partnerships (PPPs) are widely spread long-term arrangements between governments and strategic private partner(s). One of their objectives is to reduce the financial pressure on the public treasury with regard to new investments. PPPs have been employed within the health care sector which, in turn, carries a huge social burden. In Portugal, for instance, PPPs in health care concern bundling hospital infrastructure and clinical services management. Notwithstanding the need to ensure sustainability and efficient use of hospital resources, it is clearly compulsory to guarantee that patients receive appropriate and timely care, with maximum security, and equitable manner. Still, little or even no attention has been paid in the literature to the clinical response capacity of PPP hospitals and to the populism arguing that these entities have a lower social performance than typical public hospitals. This study uses robust benchmarking methodologies alongside recent data about Portuguese hospitals (FY2012-FY2017) to demystify this idea and to demonstrate that, actually, PPP hospitals can deliver health care services with social performance levels at least as good as public hospitals.

1. Introduction

PPPs are long-term arrangements signed by a government or a public agency – on behalf of the public sector – and a private partner – commercial enterprises, consortia, nongovernmental organisations, religious groups, …–, who will be responsible for managing part of or the whole infrastructure, like a hospital [1]. According to Ref. [2], "*PPPs are connected with infrastructure projects and are institutional arrangements for cooperation expressed through the establishment of new organisational units. In the world of infrastructure projects, PPPs are also seen as financial models that enable the public sector to make use of private finance capital in a way that enhances the possibilities of both the elected government and the private company*". In general, PPPs are characterised by: (*i*) demanding forecasts subject to great uncertainty – given the long lasting nature of these projects, demand predictions become practically impossible to accomplish [3]; (*ii*) incomplete, sometimes rigid, contracts [4]; (*iii*) large investments that commonly cannot be afforded by governments (or they do not want to take the full responsibility) [5]; (*iv*) shared risk, rewards, and roles between both public and private partners [6]; (*v*) public control (at some degree) over the project [7]; and (*vi*) large sunk investments, subject to opportunistic behaviours by contracted and contractors [8].

The concept of PPP has been widely discussed in several sectors, although its potentialities and disadvantages are not widely known. The preconceived notion that such partnerships result in ruinous contracts for the public purse has, on its genesis, the State's inability, first and foremost, to take advantage of the competitive potential of the private strategic partner and its higher efficiency in risk management; second, to increase its bargaining power vis-à-vis the private contractor, resulting in poor contract management and thus leading to numerous renegotiations that often harm the public contractor (State). In addition, the effect of optimism biasing resulting from too high expectations on demand, consequently, generates considerable rewards to the private partner. Nevertheless, there are several reasons for such a partnership. It is widely acknowledged that both deadlines and budgets for public works often skim to values well above what was initially expected. Since the private partner is paid a fixed price, and it aims to maximise its profit, the skidding risk is minimised if the work is delivered to it. In this way, part of the scarce statutory resources are released to other types of applications. Moreover, innovative solutions are possible, and active and dynamic management of infrastructures and services is ensured; thus, a more rigorous selection of projects becomes possible.

It is widely accepted that firms, either public or private, must be financially, socially, and environmentally sustainable. Ensuring

* Corresponding author. CERIS, Instituto Superior Técnico, Universidade de Lisboa, Portugal, Avenida Rovisco Pais, 1, 1049-001, Lisboa, Portugal.
E-mail addresses: diogo.cunha.ferreira@tecnico.ulisboa.pt (D.C. Ferreira), rui.marques@tecnico.ulisboa.pt (R.C. Marques).

https://doi.org/10.1016/j.seps.2020.100798
Received 15 April 2019; Received in revised form 26 December 2019; Accepted 12 January 2020
Available online 23 January 2020
0038-0121/

sustainability means to perform well with respect to a set of targets or, in other words, to be both efficient and effective. Efficiency measures the degree of outputs produced given a certain level of inputs, whereas effectiveness states the degree of fulfilled goals initially proposed. In certain cases, these two concepts may overlap. For instance, hospitals aim at providing the best care services [9], improving the quality of life of patients the best possible way (social performance), at the minimum waste (financial performance). Being socially efficient means successfully treating all seen patients. For example, the ratio of *patients treated with success per admitted patient* can be a measure of social efficiency. Hospitals' social goals include enhancing the quality of life and satisfaction of patients, without adverse events within the treatment period or after discharge. Therefore, social efficiency and social effectiveness are, here, comparable concepts, which hereinafter will be replaced by the generalised term *social performance*.

Ref. [10], quoted by Ref. [11], defined social performance as "*a business organisation's configuration of principles of social responsibility, processes of social responsiveness, and policies, programs, and observable outcomes as they relate to the firm's societal relationships*". Meanwhile, Ref. [12] defined *social responsibility* as the capacity of organisations to meet their fundamental goals, including to fulfil the stakeholders' interests. The State, on behalf of the community (citizens), is a major stakeholder of both public and PPP hospitals, and its interest lies mainly into the improvement of the citizens' health status. On the one hand, citizens wish better health levels because the latter are likely to enhance the quality of life. On the other hand, governments aim healthier populations because the latter are *a priori* more productive. For this reason, social responsibility and social performance are two interrelated concepts: the hospital (or any other health care provider) can only be a good socially performer if it meets (most of all) its stakeholders' interests.

Although most of hospital PPPs are based on infrastructure and hard facilities management, as in the UK model, in some cases the clinical management is also included into the arrangement. Examples include the hospital PPPs in Portugal (first wave model) and Spain (Alzira model) [13]. point out two questions about PPP projects: first, *do they provide value for money to the public sector?* and second, *how effective can be the inclusion of clinical management into the PPP?*. Obviously, these two questions are related to the societal importance of these entities. If quality of the delivered care cannot be periodically and reliably monitored, then clinical services should not be provided by the private partner; otherwise, adverse events on patients' health status may occur.

Quality of delivered health care services is unquestionably a determinant of their social performance. According to Refs. [14,15], quality is related to the infrastructure, the available technology and equipment, the actions undertaken by clinical and non-clinical staff within the healing process, the patients' safety, and the final outcomes – *i.e.*, whether the patient lives or dies after admission, and in the former case, whether she/he is positively satisfied and is likely to follow medical recommendations after discharge. Although one could think that success of treatments is only a matter of quality, the access to appropriate and timely services plays also a central role on patients' health care status improvement. For instance, Ref. [16] concluded that the patients' travel distances from home to hospital (a direct measure of access) have impact on the health care provider's social efficiency, especially if pathological and critical conditions are accounted for. Thus, access is part of the social performance and should not be disregarded from its definition. These two concepts – quality and access – constitute the core of this research as their proxies were considered variables – either desirable or undesirable – related to the process of care and used to construct composite indexes that characterise the social performance of the (Portuguese undifferentiated) hospitals.

This study aims to analyse if public hospitals and PPPs have different social performance levels, as measured by a composite index of the two already mentioned main hospital' social performance dimensions: quality and access. Composites are also measures of (in)efficiency, and weights used to construct those indexes are optimised through the widely spread linear programming tool PPP [17,18], under a nonradial framework that allows the simultaneous minimisation of the undesirable variables and maximisation of the desirable ones, at different rates – an approach that is closer to the real world than the usually employed radial techniques. The PPP-based model used to construct composite indexes is usually named *Benefit of Doubt*. PPP and models alike have been widely employed in the health care literature [19,20] as these models construct a piecewise frontier where efficient hospitals are placed and inefficient entities are projected following a predefined path. Targets for the inefficient ones are, then, the linear combination of Pareto efficient observations. Part of the PPP success lies on its flexibility, the lack of strong assumptions about the efficient frontier shape, its ability of handling multiple variables, and its empirical nature [21].

This kind of (benevolent) approaches has been widely employed for performance assessment in a broad range of fields, including project management and health care services. Regarding the latter, PPP-based models remain as the most employed ones for efficiency (and performance, in general) computation [22,23]. Among these studies, most are devoted to hospitals' managerial efficiency and to the effects of policies on their performance. According to Ref. [23] and under certain circumstances, public provision of care seems to be more efficient than private. For example, Ref. [24] used a bootstrapped integration of a Malmquist index and a nonconvex partial frontier-based PPP to study if corporatisation of public Portuguese hospitals did meaningfully introduce productivity and efficiency gains. This reform over public health care providers intended to introduce private management tools into the public sector. However, those authors concluded that it had no positive effects concerning the providers' efficiency. Opposite conclusions were drawn by Ref. [25] regarding the German public hospitals, which means that concluding by the negative or positive impact of a reform over performance should be cautious as it usually depends on the exogenous environment in which providers are inserted. In other words, introducing private management tools in the public sector, including the creation of PPPs, did not always improve its managerial performance. Furthermore, accounting for different environment is compulsory in any field, and particularly the health care.

Note that most, if not all, of those benchmarking studies have been committed to the measurement of resources utilisation to deliver care services (total number of treated inpatients and outpatients), regardless of other – perhaps more important – factors like quality and access, *i.e.*, the social performance of health care providers has been frequently disregarded from studies evaluating health policies and project management, despite its great importance for the society. This study is pioneer on comparing groups of hospitals in terms of social performance. A case study regarding the Portuguese NHS was used for such a purpose. The considerable growth of PPPs worldwide, especially in the health care sector, has brought several questions related to the moral hazard effect that would be expected from private companies operating and managing structures with an enormous social responsibility, such as the hospitals. The main research questions here are then: *Can PPPs provide equally or better care services than public hospitals, given the fact that them both belong to the NHS and must deliver tendentiously free and universal care to any citizen demanding it? Is the pursuit of profit, naturally present in private companies, imposing a trade-off on the social performance?* Answering these questions with some solid benchmarking methods has a potential impact on future related projects – and their management – as it may influence policy decisions and the public opinion.

To compare the two groups of hospitals, one may nonparametrically (empirically) create two group-specific frontiers and a metafrontier enveloping both groups. Gaps between each frontier provide precious information about the relative position of each group regarding the theoretically feasible production technology. Evaluating these gaps requires powerful statistical tools. Standard statistical tests are not appropriate in this framework given the efficiency scores nature and their serial correlation. Furthermore, hospitals' performance must be adjusted by the exogenous conditions under which they operate. For these reasons, this study adopts a conditional subsampling algorithm as a simplification of a powerful three-cycle Monte Carlo tool that has been recently proposed in the literature. Such an algorithm provides a large set of bootstrapped samples of efficiency scores adjusted for both bias and environment, and which can be used for statistical testing purposes. Dynamic effects can also be accounted for by using that algorithm. Finally, it is worth to mention that the used rich database contains some gaps of information. This research constructs two possible scenarios: first, hospitals with at least one gap of information were removed from the sample; second, intervals of admissible values were attributed to gaps and two models were constructed – one optimistic and another pessimistic – following an augmented interval PPP especially designed to account for the nonradial inefficiency sources. It is the first time that those two groups of hospitals are compared using such a solid technique.

The results and conclusions drawn by this research are important for: (*i*) *policy makers*, who usually make the decision on starting a new hospital PPP project, analyse proposals, and decide which is the best alternative, (*ii*) *citizens*, as they should become more informed, (*iii*) *hospital managers and clinical staff*, as they should search for the best practices in the field to improve their own performance, either with respect to their own group of similar entities or regarding the theoretically possible production technology, (*iv*) *operational researchers*, as innovative and solid tools are described and developed in this study, and, last but not least, (*v*) *project managers*, who should evaluate new proposals based not only on financial indicators but also on social performances of entities playing in a market with a huge social importance.

2. PPPs in health care: An overview

This section presents an overview regarding PPPs in health care. The first subsection provides the main models used to classify PPPs, whereas the second subsection brings some lessons from past research.

2.1. Different models of hospital PPP

PPPs in health care have been widely employed all over the world, especially in Europe and in the United States. These partnerships can range from a simple franchising, aimed to privately manage an existing hospital, to a full-service provision at all levels of care, basically characterised by integrating the hospital and the primary care centres to provide care to the community. An example of the latter is the Ribera Salud/Alzira companies. A major feature of these models regards the bundling of clinical services into the contract signed by the public sector and the strategic private partner. The most common model, following the British private finance initiatives, attributes the management of the infrastructure (including building itself) and of the ancillary services to the private partner, and keeps the provision of care under the umbrella of the national health services. For instance, countries like France, Germany, Italy, and, of course, the UK have adopted this type of PPP model [26]. In opposition, experiences in some Mediterranean countries (Portugal and Spain) have included the management of clinical services (*e.g.*, staff, clinical devices, drugs) and soft facilities (*e.g.*, laundry) into the hospital PPP contract [27].

Other classifications of hospital PPP contracts exist, though. For instance, the European Union have classified PPPs in health care using eight different types of models, ranging from a simple accommodation model – either accounting or not for IT, soft facilities management, and supply, installation and maintenance of some medical devices (most of the private finance initiatives are inserted into the accommodation model) –, to the full-service provision of all levels of care – as happens with the Alzira model. Other models within this classification include joint ventures (with separate infrastructure and clinical services) – the so-called Twin-Special Purpose Vehicle model, as in the Portuguese case –, and the full-service provision of secondary care by the same private entity – as in some German hospital chains [28].

One of the objectives (and eventually a benefit) of PPPs is the risk transfer (and sharing) from the State to the private partners. Hence, when the public responsibility decreases (alongside the degree of public sector risk), the private responsibility increases. In terms of scale and scope of PPPs, Ref. [29] ranked the different models from high to low risk for the public sector: (*i*) service/management contracts (contracting out), (*ii*) design and build, (*iii*) design, build, and maintain, (*iv*) lease, operate, and maintain, (*v*) design, build, operate, and maintain, (*vi*) build, own, operate, and transfer, (*vii*) concessions, (*viii*) build, own, and operate, and (*ix*) divestitures/full privatisations.

Note that neither the contracting out nor the divestiture/full privatisation schemes are within the scope of PPPs.

2.2. Some lessons from past research

There are several literature revisions reporting the international experience with PPP in health care. Some remarkable ones include Refs. [13,30,29,31], to name a few. This subsection does not intend to be an exhaustive review of published studies on this field. Rather, one aims to bring the main conclusions and lessons drawn by the aforementioned studies, which, in turn, were devoted to that goal.

Ref. [29] observed that the number of published papers on PPPs has exponentially grown since the early 90's. Based on more than 1,400 articles, those authors define a useful conceptual map relating three connected themes: *policy* – efficiency, access to idiosyncratic resources, appropriateness, risk management, and financial evaluation –, *practice* – stakeholders alignment, incentives, inter-organisational governance mechanisms, inter-project learning, and knowledge management –, and *outcomes*. In short, if well managed, hospital PPPs can take advantage of private strategic skills, including innovation, know-how, and managerial/technical efficiency, as well as public stakeholders, to deliver high quality secondary health care services to the citizens. Clear benefits would be expected, including jobs creation and incentives to innovative solutions and competitive environments, prone to induce efficiency enhancements on the sector. The authors also claim that future work is much needed to evaluate to what extent both infrastructure and clinical services management should be vertically bundled.

After a comprehensive revision of the literature on PPPs in health care services and infrastructure, Ref. [13] observed (*i*) the absence of transparency, informed debates, and accountability, (*ii*) the existence of unreliable, questionable, or even opportunistic assumptions underlying some PPP procurements – *e.g.*, the so-called and well-known optimist bias related to the overestimated demand for a service –, (*iii*) the potential to introduce inequalities, hospital's capacity downsizing, and skimming effects (ethical challenge), (*iv*) the trend to plan hospital PPPs in terms of financial rather than population needs, and (*v*) overall, a value for money that is skewed in favour of PPPs.

In short, less risky projects should be adopted instead of PPPs with high risk. When clinical services are bundled in the PPP procurement, some challenges can be identified, as concluded by those authors. Indeed, monitoring health care related dimensions is not as easy as in other utilities sectors. Hospitals are complex structures and defining all dimensions capturing the scope and quality of all of their services is an Herculean task. Besides, the high risk associated with dubious forecasts of both health technology development and citizens demand are also prone to force renegotiations, which may turn the PPP more costly than the public traditional procurement. Still, the authors argue that more flexibility should be given to the private partner to guarantee that potential synergies between infrastructure and clinical management are properly exploited. Although more research is compulsory in this field, the vertical bundling approach observed in Portugal, Spain, and certain German hospital chains has win-win potential.

Ref. [30] confirmed the need for a deeper and effective statutory regulation, to ensure quality and avoid moral hazard. The authors, after revising 18 published studies, identified nine success factors for health care PPPs: stewardship, coordination, regulation, incentives, capacity to support partnership, monitoring and evaluation, high-level support and buy-in, harmonisation and alignment, and innovation. The evaluation of PPP outcomes and processes is necessary to identify best practices, as well as poor performers, perverse incentives, and other unintended consequences.

Ref. [31] clustered 46 studies on health care PPPs into six main groups depending on their focus: effectiveness, benefits, country overview, public interest, partners, efficiency, and a miscellaneous group containing the studies that could not be classified within any other cluster. 19% of analysed studies regard effectiveness of PPPs. The authors observe that the introduction of private partners into the public sector, namely into the provision of health care, shifts the government's role from operator to regulator, especially when the public interest may be jeopardised and, therefore, the accountability, the stakeholders' involvement, the transparency, and the symmetry of information are mandatory. More intensive government participation in these partnerships is also required, including the designing phase. Moreover, it was observed that a protocol to evaluate the PPP effectiveness is missing, simply because there is no evidence about it in the literature. According to Ref. [31], there are eight main aspects to account for when effectiveness is going to be assessed: (*i*) relationship between public and private partners, (*ii*) nature of the partnership, (*iii*) financial arrangements, (*iv*) government policy, (*v*) outcomes and process of care dimensions, (*vi*) equity, and (*vii*) potential weaknesses of the analysis.

Those authors conclude that a deeper research on the effectiveness of PPPs should be made prior new investments.

A straightforward conclusion arising from the previous analysis is that, despite the broad literature on the field, the social performance of hospital PPP was neither assessed nor compared to other secondary health care providers. A plausible explanation may lie in the difficulty of describing and quantifying the social performance and related dimensions (quality and access). Because questions concerned to the effectiveness of these entities remain unanswered, the present research is deemed to be compulsory. Next section operationalises this issue under a solid mathematical benchmarking framework.

2.3. PPPs in health care: The Portuguese experience

The Portuguese NHS was created in 1979 and aims at providing universal and tendentiously free care services to all citizens demanding for them. The NHS is under the responsibility of the Portuguese Ministry of Health, who funds health care providers (primary care centres, hospitals, local health units, continuing care units) through prospective contracts. Funds are collected via taxes to the population; thus, the Portuguese NHS is a Beveridge system. Public expenditures with health care represent a considerable share of the GDP: currently, about 6% of the GDP regards public health care expenses. Fig. 1 portrays the evolution of both GDP and public health care expenditures-to-GDP, since FY2002, in Portugal. Until FY2010, the evolution of those expenses watched a positive trend, usually above the GDP growth. Nonetheless, the period FY2011–FY2014 was characterised by the Bailout Programme and the Memorandum of Understanding, which had imposed an

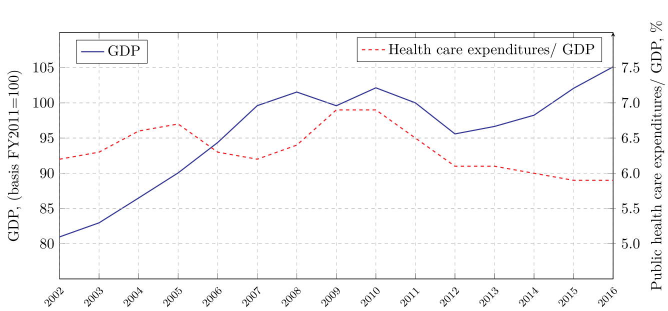

Fig. 1. GDP and public health care expenditures/GDP evolution, FY2002–FY2016 (Portugal).

expressive expenses reduction, including the health care related ones. After the crisis period, health care expenses observed nominal increases, still at smaller rates when compared with the pre-crisis period. In FY2016, public expenses with health care reached €,061 *per capita*, nearly 66% of total current expenses. Given that, Portugal remains as one of the highest spenders with health care in the European Union and the OECD [13].

Probably due to the weight that expenses with health care have on State accounts, the successive Portuguese governments have implemented a PPP model in the secondary health care provision sector (hospitals). This model lies between the conservative Private Finance Initiative-based British model and the Spanish (Alzira) model, see Subsection 2.1. The Portuguese hospital PPP model, known as *first wave*, is based on the DBFOT framework – design, build, finance, operate, and transfer –, and uses two long lasting contracts between private partners and the public sector: one for the infrastructure and maintenance activities (which lasts for 30 years), and another for the clinical management and soft facilities (10 year-long). This is the main difference between the Portuguese and the British model of hospital PPPs. Main features of the Portuguese *first wave* of hospital PPPs are described in Table 1. The idea underlying the Portuguese model is that synergies arising from bundled infrastructures and clinical management can be exploited [7], diminishing the *risk of interface friction*.

Currently, there are four hospital PPPs in Portugal, operating under the *first wave* framework: (*i*) Hospital de Cascais, (*ii*) Hospital de Braga, (*iii*) Hospital de Vila Franca de Xira, and (*iv*) Hospital Beatriz Ângelo, Loures. Although hospital PPPs belong to the Portuguese NHS, they are not funded using the same contracting terms as public hospitals. For instance, the private partner responsible for managing the infrastructure and hard facilities is paid based on some pre-defined infrastructure availability criteria. According to Ref. [13], the Portuguese *first wave* model has clear advantages because a new tender for clinical provision is

Table 1
Main features of the Portuguese first wave of hospital PPPs. Adapted from Refs. [13,7,32,33].

	Private partner responsible for the infrastructure management	Private partner responsible for the clinical management	Objectives
Model	DBFOT, including clinical services		To increase the risk transfer to the private partner
Contract issues	To design, build, and preserve (manage) the infrastructure and hard facilities (ancillary services)	To manage clinical staff and deliver health care to all citizens	To isolate risk by entity
Duration	30 years	10 years (maximum 30 years)	Different investment depreciation periods, according to the assets' life cycle
Risk	Construction, availability, infrastructure quality, operating costs related to hard facilities management	Availability of resources such that access to health care is not compromised – according to the contracted production –, quality of delivered care, operating costs associated with soft facilities management	Maximum risk transfer
Payment	Based on services availability – penalties for service failures	Based on contracted production – penalties for low quality	N.A.

launched every 10 years, thus poor performers will likely to be replaced. Those authors argue that this process provides incentives to competition between candidates and the good performance of incumbents, as well as it mitigates potential renegotiations (which, in turn, are more likely for 30 year-long contracts). Still, one must be aware that the optimism bias is usually present. This bias is linked to overestimation of demand for health care, which imposes compensations to the private partner(s) when the predicted demand is not observed. The literature commonly refers the skimming effect as a potential outcome of this kind of partnerships: chronic or very demanding inpatients may be avoided/transferred to another facility, because the private partner supports the risk of treating a number of patients above 110% of the contracted volume. Finally, health care has watched a huge development in terms of technologies, drugs, and other medicines; this evolution is prone to imply renegotiations between the public sector and the private partner responsible for the clinical management [33]. Hence, Portuguese health policy makers are now more sceptical with regard bundling both infrastructure and clinical services under the private management. Indeed, although no new hospital PPP has been constructed after those four cases, the *second wave* foresees that new projects will strictly follow the British hospital PPP model. For these reasons, comparing public hospitals with PPPs in terms of their social performance is meaningful for new PPP projects, for renegotiations, for new tenders, for policy makers, and for the citizens in general.

3. Case study: The Portuguese experience

This section presents the case study to determine if PPPs do exhibit better social performance levels than the remaining public hospitals. The case study presents and describes the main variables used as well as some first insights about the social performance of two groups of hospitals: PPPs and public hospitals.

3.1. Sample

This study uses a sample of 2,660 observations of hospitals, of which 2,380 correspond to public hospitals (single hospitals, hospital centres, and local health units) and 280 to hospital PPP arrangements. The sample is homogeneously distributed by five years, since January, FY2013, to October, FY2017. The Portuguese NHS is composed of primary and secondary health care providers. The latter are constituted by singular hospitals, hospital centres (horizontal merging of at least two singular hospitals), local health units (vertical merging of a singular hospital and some primary health care centres on the former vicinity), PPPs, oncology centres, maternities, and psychiatric hospitals. Although data regarding local health units contains information about both the hospital and the primary care centres, the social performance related variables regard only the secondary care level, making those units comparable with the remaining public general hospitals from the NHS. Meanwhile, oncology centres, maternities, and psychiatric hospitals were disregarded from the analysis because they are specialised entities and their production technologies are different from general hospitals [34].

3.2. Desirable and undesirable process variables

This study evaluates whether there is a meaningful gap between public hospitals and PPPs in terms of their social performance. The latter is computed using information of quality and access to health care services, which is made available by the official database (http://benchmarking.acss.min-saude.pt/), maintained by the (Portuguese) Central Administration of Health Systems.

The choice of quality variables follows here three main criteria: data availability for all the sample, the literature review, and the Donabedian's approach [15]. The latter categorises quality into three main interrelated categories: structural quality (infrastructure and equipment), process quality (actions undertaken by clinical staff to improve the patients' health status and to keep them safe from adverse events), and outcomes (the final effect of care on patients' quality of life) [35]. The utilisation of outcomes as the measure of quality is often objectionable [36,37] because they are difficult to measure and depend on a wide set of external and nondiscretionary events, including the health status prior hospitalisation and the follow-up of medical indications and prescriptions after discharge. For that reason, most researchers seem to prefer using health care process related quality indicators [38,39], although in some cases they are used alongside outcome measures [40]. Meanwhile, process quality and outcomes can be classified in terms of care appropriateness and timeliness, and of clinical safety. The former regards the capacity of caregivers to deliver the appropriate and timely care services at the same time medical/nursing procedures strictly follow evidence-based guidelines [41]. Patients' clinical safety is the ability of preventing complications during care that may lead ultimately to the patients' death. Of course, while appropriateness and timeliness of services are mostly a matter of the hospitals' administration and staff and their capacity of correctly handling the need for care, the clinical safety also depends on the (co) morbidities, pathologies, syndromes, diseases, and disorders of patients. Usually, one is not able to disclose if a variable is a measure of care appropriateness or of clinical safety, because these two concepts are interrelated. For instance, the in-hospital mortality for low levels of SOI can measure the (lack of) clinical safety because the hospital and its staff could not provide the appropriate (and probably timely) care services. Likewise, it is sometimes difficult to classify quality variables in terms of process of care or the outcomes because the latter is partially the result of the former.

According to a comprehensive literature review, the following are common quality related variables: failure to rescue and risk-adjusted in-hospital mortality, (acquired) infections due to medical care, readmissions, caesarean rate, patients' satisfaction, postoperative respiratory failure and sepsis, and inappropriate discharges; see Refs. [42,43,44,45,40,46,47,48,38,39,49,50,51,52,53,54]. Except for satisfaction, which is a desirable outcome, the remaining are undesirable. In some cases, authors opt for transforming variables to turn them desirable, *e.g.*, mortality is often replaced by its reverse, or by survival. Such a transformation is not advisable in most of benchmarking nonparametric frameworks, because several models are neither units nor translation invariant. Moreover, survival and mortality are not complementary concepts because the population at risk is not the same [55]. For instance, the population at risk associated with mortality is the amount of patients who received treatments within the hospital, whereas the population at risk for survival regards the patients under severe conditions and for whom the probability of death is high. For these reasons, this study clusters process of care related variables into desirable and undesirable, and handles them accordingly.

Like quality, access to health care services is a complex concept. Good outcomes in health care are, in general, the result of a mix of good clinical practices (process of care), universal/free/timely access to the appropriate service(s), and the patients' own conditions prior admission. According to Refs. [56,57], there are four main access dimensions: availability of services (*e.g.*, resources per 100 inhabitants), personal barriers and acceptability (*e.g.*, education and health literacy), financial barriers (*e.g.*, co-payments and transportation costs), and organisational barriers (*e.g.*, waiting lists).

Fig. 2 presents the 11 process of care related variables chosen by this research. The variables choice was based on their availability for a considerable number of entities (although in some cases missing data were replaced by ranges, *vide infra*), a literature review (including some inpatient quality and safety indicators defined by the North American Agency for Healthcare Research and Quality, https://www.ahrq.gov/), and the dimensions evaluated by the Portuguese government in arrangements with hospitals. Quality is clustered in terms of care appropriateness and clinical safety, whereas access is divided by timeliness of

- Process of Care
  - Quality
    - Care Appropriatness
      - g_1^+, Outpatient surgeries per 100 potential outpatient procedures,
      - g_5^-, Readmissions within 30 days after discharge, per 100 discharges,
      - g_6^-, Inpatients staying more than 30 days per 100 inpatients,
    - Clinical Safety
      - g_7^-, Decubitus ulcer cases per 1,000 inpatients,
      - g_8^-, Catheter related bloodstream infections per 1,000 inpatients,
      - g_9^-, Postoperative pulmonary embolism/deep vein thrombosis per 100,000 inpatients,
      - g_{10}^-, Postoperative septicaemia cases per 100,000 inpatients,
  - Access
    - Timeliness
      - g_2^+, First nonurgent medical appointments within the maximum (legislated) guaranteed time, per 100 first nonurgent medical appointments,
      - g_3^+, Hip surgeries in the first 48h after fracture per 100 hip surgeries,
      - g_{11}^-, Average delay before surgery (in days),
    - Services Availability
      - g_4^+, Hospital beds per 100 standard inpatients.

Fig. 2. Process of care: variables. Note: g^+ identifies the desirable variables while g^- identifies the undesirable ones.

care and services availability. Although the population at risk and its features are indicators of broader or narrower access to health care services, they were included as exogenous variables to homogenise the environment under which hospitals operate. Basic statistics for the chosen variables are in Table 2.

This study considers three variables to characterise care appropriateness: outpatient surgeries on potential outpatient procedures, readmissions after discharge and within an adequate range of time, and hospital stays beyond appropriate delay. The first dimension is desirable because there is no meaning on subject inpatients to major surgeries and hospitalisation when minor (outpatient) procedures are the most appropriate services. Small rates of g_1^+ indicate lack of appropriate care as well as the overuse of resources, as major surgeries are costlier than the minor ones, and decrease the access of other patients to the internment service because beds become unduly occupied. Meanwhile, readmissions represent undesirable process of care related variables, because they can result from poor care, unstable therapy at discharge, inadequate postdischarge care, and lack of patients' preparedness [58]. As readmissions, long stays increase the probability of other diseases appearance and reveal lack of coordination with other health care providers and health care levels [59]. Overall and on average, nearly 73% of all potential minor surgeries were fulfilled between FY2012 and FY2017, whereas 8% of discharges were readmitted after (at the most) 30 days and 3% of inpatient cases are long stays.

Clinical safety is here evaluated through four undesirable process dimensions: pressure ulcers and wound dehiscence in bedridden [60, 61], and the potentially lethal postoperative catheter related bloodstream infections, pulmonary embolisms and deep vein thromboses (or iatrogenic pneumothorax), and septicaemia cases [62,63,64]. Secure hospitals have low or even null clinically adverse events like ulcers on bedridden and infections within the postoperative period and until discharge; in opposition, unclean and non-disinfected facilities and clinical material are likely to produce that kind of unsafe events. Mortality as the ultimate consequence of unsafe environments, especially regarding the non-severe cases, would be also an interesting variable representing the lack of patients' clinical safety, but given data constraints this variable was disregarded from the analysis.

Services timeliness and availability are two main branches of access to health care. The faster the services and the more resources available to those who need them, the better the access to those services. Since resources are limited in hospitals but no (Portuguese) citizen can be excluded from public health care (regardless of her/his ability to pay), the former are said to be rivalrous, which means that barriers to access are inherently present. The first one, and perhaps the most common, is the average delay (or waiting time) for a service. This study assumes the waiting time before a surgery and after admission to the hospital ward, as proxy for the overall waiting time in the hospital and its lack of resources (physical, staff) to perform the surgeries within the scheduled time. The more time the inpatient stays in the hospital, the lower the number of possible further admissions. The number of first nonurgent medical appointments within the (legislated) maximum guaranteed response time per first appointment is another measure of timeliness. Indeed, Portuguese citizens may enter the hospital by two ways: via query appointment request made by the general practitioner (in primary care) who is as a gatekeeper, or via emergency service. Low rates of g_2^+ indicate that the hospital is unable to meet the legislation and to provide timely demanded speciality consultations. Likewise, the ability of orthopaedic surgeons to deliver hip fracture surgeries within 48h after admission is also an indicator of timely services provided by the hospital [65,66]. Finally, the availability of services is measured by the number of beds per 100 inpatients; it is frequently pointed out as proxy for resources per citizen demanding them.

According to Table 3, Pearson's correlation coefficients between the different process variables are small or even non-statistically significant at the level 1% of significance. This means that the association degree between the process of care related variables is weak and each one of them brings new and non-redundant information into the model. Furthermore, hospitals' clinical activity cannot, in general, be explained by a single or a small set of process variables, as entities performing well within one dimension can be bad players in another.

3.3. *Some first insights about the social performance of the two groups of hospitals*

Table 2 also provides the results of two statistical tests employed to compare the performance of each group (public hospitals and PPPs) in

Table 2
Process of care related variables: Basic statistics. Note: g^+ identifies the desirable variables while g^- identifies the undesirable ones.

	g_1^+	g_2^+	g_3^+	g_4^+	g_5^-	g_6^-	g_7^-	g_8^-	g_9^-	g_{10}^-	g_{11}^-
Overall sample											
Mean, $\hat{\mu}(g)$	73.29	76.46	49.22	17.16	7.95	3.16	1.48	0.07	142.38	745.26	0.86
Std. Deviation, $\hat{\sigma}(g)$	14.93	12.44	25.49	18.58	2.18	1.20	2.67	0.30	231.30	961.36	0.38
CV (%)[a]	20.37	16.27	51.79	108.26	27.45	37.95	180.75	447.61	162.45	129.00	43.57
min(g)	0.00	0.00	0.00	4.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max(g)	99.90	100.00	100.00	70.00	22.22	8.77	24.73	3.61	1612.90	8080.81	3.89
Public hospitals, A											
Mean, $\hat{\mu}(g)$	74.94	75.88	48.54	17.28	8.01	3.22	1.24	0.07	135.56	678.30	0.90
Std. Deviation, $\hat{\sigma}(g)$	13.33	12.80	25.51	19.45	2.25	1.23	2.01	0.30	226.82	880.45	0.36
CV (%)	17.79	16.87	52.55	112.57	28.12	38.18	161.81	451.09	167.32	129.80	39.89
min(g)	34.51	0.00	0.00	4.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max(g)	99.90	100.00	100.00	70.00	22.22	8.77	19.23	3.61	1612.90	5882.40	3.89
Public-private partnerships, B											
Mean, $\hat{\mu}(g)$	58.75	81.59	55.26	16.13	7.37	2.69	3.55	0.06	202.64	1337.10	0.51
Std. Deviation, $\hat{\sigma}(g)$	19.76	6.93	24.76	7.39	1.28	0.79	5.46	0.26	261.93	1369.00	0.32
CV (%)	33.64	8.50	44.81	45.84	17.31	29.36	154.12	414.76	129.26	102.38	63.66
min(g)	0.00	48.57	0.00	7.46	4.48	0.90	0.00	0.00	0.00	0.00	0.11
max(g)	83.54	100.00	100.00	69.38	9.88	5.27	24.73	1.48	993.38	8080.80	1.87
Student's t-test for means											
p-value[b]	0.00	0.00	0.01	0.52	0.00	0.00	0.00	0.90	0.00	0.00	0.00
Δ,[c] lower bound	13.50	−8.07	−11.56	−2.39	0.23	0.30	−2.80	−0.05	−111.02	−838.17	0.33
Δ, upper bound	18.88	−3.37	−1.87	4.70	1.06	0.75	−1.81	0.06	−23.13	−479.40	0.47
T statistic	11.81	−4.77	−2.72	0.64	3.06	4.53	−9.20	0.13	−2.99	−7.21	11.44
Best performer	A	B	B	similar	B	B	A	similar	A	A	B
Kruskal-Wallis nonparametric test for distributions											
p-value[d]	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.86	0.00	0.00	0.00
Best performer	A	B	B	similar	B	B	A	similar	A	A	B

[a] $CV(g) = 100 \cdot \hat{\sigma}(g)/\hat{\mu}(g)$ is the coefficient of variation, which measures heterogeneity among observations.
[b] p-value regards the null hypothesis of equal means and equal unknown variances using the two-sample t-test. Small p-values, say $p < 0.01$, imply rejecting the null hypothesis.
[c] Δ is confidence interval on the difference of the population means.
[d] p-value regards the null hypothesis of similar distributions.

Table 3
Pearson's correlation coefficients for process of care related variables. Italic entries identify non-statistically significant coefficients, at the 1% significance level.

g_1^+	g_2^+	g_3^+	g_4^+	g_5^-	g_6^-	g_7^-	g_8^-	g_9^-	g_{10}^-	g_{11}^-	
g_1^+	1	−0.1272	0.1866	*0.0570*	*0.0308*	−0.1493	−0.0945	*−0.0580*	*−0.0721*	−0.1552	0.1668
g_2^+	1	*−0.0577*	*0.0113*	0.1419	0.1447	0.0869	*0.0045*	*0.0603*	*0.0721*	0.1976	
g_3^+	1	*−0.0302*	*−0.0605*	−0.2317	0.0988	*−0.0038*	*−0.0392*	*−0.0266*	−0.1152		
g_4^+	1	*0.0378*	*−0.0357*	*−0.0163*	*−0.0351*	*−0.0623*	*0.0014*	*0.0082*			
g_5^-	1	0.1461	*−0.0244*	*0.0118*	*0.0383*	0.1369	0.1376				
g_6^-	1	0.1164	0.1559	0.1514	0.1776	0.4233					
g_7^-	1	*0.0584*	0.1333	0.2431	*−0.0444*						
g_8^-	1	*0.0329*	*0.0657*	*0.0497*							
g_9^-	1	0.1904	*0.0500*								
g_{10}^-	1	*0.0205*									
g_{11}^-	1										

each analysed dimension. p-values, confidence intervals on the difference of means (Δ), T statistics, and the best performer (group) per dimension are presented. Whenever p associated with the two-sample t-test is large, say $p > 0.01$, Δ contains the zero value and the null hypothesis cannot be rejected, which means that there is no evidence that the two groups have different means. This occurs, for instance, with the dimensions g_4^+ and g_8^-. That is, both public hospitals and PPPs exhibit similar performance in terms of beds per standard inpatient and catheter related bloodstream infections. Both tests (Student's t and Kruskal-Wallis) present identical outcomes, which may result from the considerable size of each cluster. In view of that, it becomes clear that:

(C_1) The two groups have similar performances regarding services availability, as measured by the number of beds per 100 standard inpatients;
(C_2) PPPs outperform public hospitals regarding the timeliness of care provision;
(C_3) Public hospitals outperform PPPs with respect to the clinical safety;
(C_4) There is no clear evidence supporting the hypothesis that one group outperforms the other concerning health care appropriateness.

These inconsistent findings suggest, then, the utilisation of a benchmarking tool to optimise the weights associated with each quality and access dimension to obtain a composite index that allows disclosing which group has the better social performance. Prior evidence disregards the target population of each hospital and, hence, the external conditions the latter faces; thus, the fact that one group outperforms the other in one particular dimension results from unfair and non-adequate comparisons.

3.4. Environmental variables

Exogenous environment plays a central role on hospital performance, as external operational conditions tend to influence the behaviour of health care providers. In fact, and according to Ref. [67], these conditions explain a significant part of hospitals inefficiency that cannot be explained by the case-mix (or case-load) of patients treated. In certain circumstances, environment can even replace the widely spread case-mix index to adjust efficiency scores by heterogeneity. In view of that, this study uses 14 quantitative continuous variables reflecting the population characteristics, including age, mortality, literacy, access to health care, and purchasing power, to account for the demographics and epidemiology under which hospitals operate; see Fig. 3 for details. Some of the presented exogenous variables are self-explanatory. Variables were chosen based on the Grossman's model [68] for health care services demand – propensity of individuals to use services, income, and need for care –, availability of data, and variables' own meaning. Table 4 provides some basic statistics on the 14 exogenous variables that are used to characterise the targeted population at risk treated by each hospital.

4. Prior considerations before running the performance model (s)

4.1. Basics about the performance assessment model(s)

Using a set of n observations corresponding to hospitals belonging to two clusters, A and B, with sizes n_A and n_B respectively, one may use a benchmarking model to construct a metatechnology frontier (or, simply, metafrontier). The efficiency of observations against this metafrontier is denoted by $\theta^{A\cup B}$. Likewise, one may estimate a frontier per cluster, A or B, as well as the efficiency against that frontier, denoted by θ^A or θ^B respectively.

In the present case we have adopted a non-radial directional PPP model with four desirable variables, g^+, and seven undesirable variables, g^-, and the directional vector of Ref. [69], through a framework known as *Benefit of Doubt* [70,71]. This framework is commonly used with a radial model that does not account for all inefficiency sources; that is why the non-radial is adopted here because of its higher flexibility. The model returns a set of slacks, $\mathscr{S}^+$ and $\mathscr{S}^-$ for desirable and undesirable variables. The development of the mathematical model can be seen in the (online) Appendix A.1.[1] An efficiency score, θ, can be constructed from those slacks, see Equation (A.2).

An important issue on the current study was the missing data for a meaningful set of observations. It is not unusual that some data may be missing from large databases. To solve this serious problem, several alternatives have been proposed in the literature. First, one may opt for removing entities verifying a considerable number of gaps [72]. This operation may impact on final efficiency scores because benchmarking methods are usually sensitive to the sample size. In several cases, and if gaps are not numerous, then they can be replaced by estimates, which introduces an error source that can be larger or smaller depending upon the used estimation method. Furthermore, estimations may introduce serious instability into the model [73]. Second, one may use blank entries to replace desirable variables and very large quantities in the place of undesirable [74]. By doing so, one can reduce the interference of missing data on efficiency assessment, but performance scores are not comparable among different entities because they are no longer evaluated using the same basis (the exact same variables). Last but not the least, one may adopt an interesting approach that regards the interval PPP alternative: it consists of transforming unknown variables by means of ranges. Such a model was proposed by Ref. [75] and details can be seen in Appendix A.2. At the end, this approach returns a confidence interval for efficiencies, see Equation (A.11).

Although widespreadly used with no major cautions, nonparametric benchmarking methods are purely deterministic and any deviation from the empirical frontier is assumed as inefficiency. Furthermore, no statistical properties are offered by those methods. That is, one cannot apply standard statistical tests and still expect reliable outcomes. To avoid these flaws, we applied bootstrap (subsampling) to efficiency scores, and correct them by bias. In short, bootstrap tries to mimic the true data generating process. One bootstrap alternative is the subsampling of N units from the set of n hospitals of the original sample, such that $N < n$. The subsampling procedure runs $\mathscr{B}$ times, which should be large. At the end, one gets $\mathscr{B}$ different efficiency estimates per hospital, from which one may get a bias-corrected score, θ, and a confidence interval for the αth quantile of the efficiency distribution, $CI_\alpha(\theta)$. Appendix A.3 (online) presents some mathematical details on subsampling.

Given the available sample, this study clusters the hospitals into two groups: A, for public hospitals ($n_A = 2,380$), and B, for PPPs ($n_B = 280$). Note that these clusters have very different sizes, which could constitute a pitfall on social performance assessment because nonparametric methods are dimensionality-sensitive. However, subsampling corrects this problem because it selects a finite and equal number of entities, N, for both clusters and for each iteration, making them comparable (*i.e.*, the dimensionality problem vanishes). In this case, $N = 38$ ensures that both clusters and the metatechnology are comparable. Appendix A.6 (online) explains how this value of N was determined.

Using the subsampling efficiency estimates one may perform some statistical tests to evaluate whether hospital PPPs do outperform public hospitals in terms of quality and access or not. Let us define the following null hypothesis:

Null Hypothesis, H_0: PPPs and public hospitals exhibit similar social performance levels.

We consider three ratios $T^{(w)}$, with $w \in \{-1, 0, +1\}$, that are applied to each one of the $\mathscr{B}$ subsampling based sets of efficiency estimates, $T_b^{(w)}, b = 1, \ldots, \mathscr{B}$, and compared with the same ratios obtained for the non-bias corrected initial set of efficiency scores, $T_{\text{obs}}^{(w)}$. This comparison allows us to construct three distinct statistical tests: left-tailed,

Operational Environment:

- z_1, Population Density, Inhabitants per km^2,
- z_2, Elderly Rate, Elderlies (>65y.o.) per 100 inhabitants,
- z_3, Youth Rate, Youngsters (<15y.o.) per 100 inhabitants,
- z_4, Dependence index, Elderlies and youngsters per 100 inhabitants at the working age,
- z_5, Death Rate, Deaths per 100 inhabitants,
- z_6, Child Mortality Rate, Mortality rate under-5 per 100 live births,
- z_7, Elderly Rel. Mortality Rate, Mortality rate over-65 per 100 deaths,
- z_8, Crude Birth Rate, Births per 100 inhabitants,
- z_9, Stillbirth Rate, Fetal deaths per 100 deaths under-5,
- z_{10}, Illiteracy Rate, Illiterates per 100 inhabitants,
- z_{11}, Sec./Tertiary Education Rate, Share of inhabitants with secondary or higher education,
- z_{12}, Inhabitants per Doctor,
- z_{13}, Inhabitants per Pharmacist,
- z_{14}, Purchasing Power Parity.

Fig. 3. Operational environment: variables.

[1] Please, go to https://drive.google.com/drive/folders/1Qy497caJOoXLwc4AjD1QFQOFza097SlV?usp=sharing.

Table 4
Operational environment: basic statistics.

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	z_{11}	z_{12}	z_{13}	z_{14}
Mean $\hat{\mu}(z)$,	456	21	14	33	11	3	90	8	33	7	29	316	997	94
Std. Deviation $\hat{\sigma}(z)$,	537	4	2	8	3	1	2	1	9	4	6	127	224	14
CV (%) [a]	118	20	13	25	28	41	3	18	29	59	21	40	22	15
min(z)	14	15	10	22	7	1	83	6	10	3	18	162	627	72
max(z)	2,428	30	17	49	16	7	94	10	50	18	40	737	1,517	125

[a] $CV(z) = 100 \cdot \hat{\sigma}(z)/\hat{\mu}(z)$.

right-tailed, and two-tailed tests, which return p-values ($p_1^{(w)}, p_2^{(w)}, p_3^{(w)}$). The null hypothesis above is rejected whenever any of these p-values is smaller than a significance level, $\alpha \in (0, 1)$, for at least two out of the three values of $w \in \{-1, 0, +1\}$. Details in Appendix A.5 (online).

Another criticism associated with performance assessment is that benchmarking models very often disregard the conditions under which hospitals operate (operational environment). This fact may introduce significant bias on results and conduce to misleading conclusions. Recently Ref. [21] proposed a three-cycle Monte-Carlo algorithm that accounts for the unbalancedness problem and, simultaneously, searches for comparable units in both clusters using a set of comparability characteristics, z. These ones may include variables such as the scope of services provided by the hospital or the population demographics and epidemiology. According to those authors, the algorithm avoids the problem of endogeneity, which can be very serious. Their algorithm was especially formulated to deal with the well-known Malmquist productivity index, but it can be simplified and adapted to the present case. Indeed, one can couple the subsampling procedure with their algorithm, and obtain simultaneously bias- and environment-adjusted efficiency scores. Appendix B (online) simplifies and presents the algorithm of Ref. [21]: Algorithm 1. It is worth to mention that it takes advantage of the interdependence of exogenous non-discretionary variables z by including the covariance matrix, Σ, associated with them. As a matter of fact the Pearson's correlation coefficients for these exogenous variables are large and statistical significant for a considerable number of couples of exogenous variables (see Table C2, Appendix C, online), meaning that the joint probability distribution associated with those couples is not likely null and justifying, by consequence, the adoption of the concentration matrix Σ^{-1} to construct the overall probability density function, see Algorithm 1 in Appendix B (online). By applying the concentration (precision) matrix, one avoids the assumption that variables z are independent when it can be untrue. In line with subsampling which is integrated in Algorithm 1, the latter's final outcome is a set of $\mathscr{B}$ bias- and environment-corrected efficiency scores that may result either from (a) the projection of the hospitals on their own technology based frontier, or (b) the projection of the hospitals on the metafrontier. Applying the aforementioned p-values over the estimates resulting from Algorithm 1 allows us, then, to conclude whether the null hypothesis underlying our research question deserves to be rejected in the light of statistical evidence.

4.2. Different models for robustness assessment

The procedure elicited by Algorithm 1 (Appendix B, online) is performed to conduct a comparative analysis of both public hospitals and PPPs against their own technology (frontier) and the metatechnology (metafrontier). One problem associated with large databases is that some of the entries may be missing, and there is no consensual approach to solve this problem. Instead of either estimating missing data, which could introduce serious deviations to efficiency estimates, or replacing desirable/undesirable process variables by zero/very large numbers, one may opt for employing two different scenarios:

Scenario I. Cases with missing data are removed from the dataset. This results on a sample of 1,161 hospitals, operating within FY2015 (January) and FY2017 (October), and from which $n_B = 118$ are observations regarding PPPs. This subsample has only 44% of the original sample size. The number of entities to be drawn by Algorithm 1 can be $N = 38$, which fulfils the condition of comparability between two samples.

Scenario II. Missing data is replaced by suitable values to achieve intervals of efficiency estimates (based on Table 2). In this case, the (sub) sample size remains, $N = 38$. It is useful to compare results with the ones achieved by scenario I because removing units from the sample can be a pitfall on the benchmarking exercise as PPP-like models measure relative efficiency and, then, are sensitive to the sample size.

It is worth mentioning that the choice of $N = 38$ for both scenarios enables direct results comparisons because frontiers are constructed using samples with the same size and the same number of variables. Furthermore, $N = 38 > \max\{s \cdot m, 3 \cdot (s + m)\} = 33$, *i.e.*, the parameter N mitigates simultaneously two main bias sources related to the dimensionality problem.

For each scenario, one may construct two different models, whether considering dynamic effects or not. As explained in Appendix B (online), dynamic effects are accounted for if more than one year is considered to create the probability density function used to draw comparable units in the Monte-Carlo procedure. This means that hospitals can be benchmarked against best practices from other (contiguous) years. Static models assume bandwidths strictly lower than 1. Meanwhile, dynamic effects are included by enlarging such bandwidths beyond 1. This study assumes any bandwidth ranging between 1 and 2 for dynamic models. The static model applied to the scenario $a = I, II$, is denoted by S_a, whereas the dynamic model is D_a. Given that environment is assumed to be stable, the dynamic model benchmarks hospitals against themselves in different moments, which means that this model is useful to study whether hospitals have improved their efficiency/productivity or not.

4.3. Model solving

Models have been implemented using the high performance Matlab® v2015b, and the optimisation package Ibm Ilog Cplex® Optimisation Studio, v12.6.3. Matlab® was used to run Algorithm 1 (Appendix B), whereas Cplex® was called to solve the linear programming problems associated with it.

Findings

Do PPPs outperform public hospitals regarding quality and access? Evidence from Portugal

5.1. Main results

This subsection presents the main findings of the current research. It starts by studying the social performance of public hospitals and PPPs relatively to their own technologies. Then, the metatechnology is constructed by merging both samples; technology gaps are determined using the bootstrap-based methodology proposed by Algorithm 1 (Appendix B, online). Those gaps are used to answer to the question: *Do PPPs socially outperform public hospitals?*.

Table 5 and Fig. 4 provide some basic statistics regarding the social performance of both public hospitals and PPPs with respect to their own-group frontier and to the metatechnology. Results for the different scenarios (static *vs* dynamic models, pessimistic *vs* optimistic models) are presented. This table contains data concerning the expected value of both performance composite(s) and metatechnology ratio: $\mathbb{E}(\theta) = \int_0^1 \theta \cdot f(\theta) \ d\theta$, where $f(\theta)$ is the density associated with the cumulative distribution function of θ, $\mathscr{F}$. The 90% confidence intervals, $CI_{90\%}$, and the frequency of good social performers, $\mathscr{F}(\tau = 0.95)$, are also provided. Table 5 is complemented by Fig. 5, which displays the 10% trimmed means and the associated 90% confidence intervals with respect to the scenario *I*. Table 5 is also complemented by Fig. 5 that presents the Gaussian based kernel density functions as well as the survival functions for both groups of hospitals. The density estimates were used to construct the function, *f*, and to compute the expected values presented in Table 5. Survival functions, $\mathscr{R}$, and the cumulative distribution ones are complementary, *i.e.*, $\mathscr{R} = 1 - \mathscr{F}$. For instance, according to the model S_I, the probability of a public hospital presenting a social performance score (composite) higher than or equal to 0.80 is nearly 50%, whereas less than 25% of them are socially efficient regarding their own-group technology (frontier).

As mentioned earlier, the two clusters of hospitals are compared based on their own gaps regarding the metatechnology. The last three columns of Table 5 provide the expected metatechnology ratios as well as the 90% confidence intervals associated with each model/scenario. Meanwhile, Tables 6 and 7 present the *p*-values with respect to the null hypothesis, H_0, and to its three alternatives, for both scenarios *I* and *II*, respectively. If, for a given significance level α, the *p*-values are above it, then there is no statistical evidence supporting the rejection of the null hypothesis. This study considers $\alpha = 5\%$ (or 0.05), which is a widely spread assumption.

5.2. Findings discussion

Using the nonparametric Kruskal-Wallis and Friedman tests, as well as the parametric two-sample Student's *t*-test, one may observe that social performance scores' distributions usually depend on the model's features; *i.e.*, composite indexes are sensitive to assumptions like the presence of dynamic effects and data precision. Importantly, dynamic effects seem to be present, likely resulting from productivity gains or losses within the coidered period. This means that these effects should always be accounted for and never disregarded. Despite the time correction made by the Monte-Carlo cycle imposed by Algorithm 1

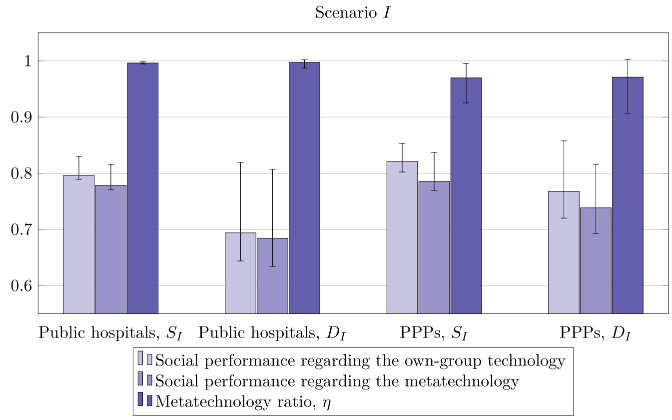

Fig. 4. Social performance of both public hospitals and PPPs regarding their own-group technology and the metatechnology (Scenario *I*): 10% trimmed means and 90% confidence intervals.

(Appendix B, online), it seems to be incorrect to consider that hospitals can only be benchmarks against best practices found in the past (or in the future), as better practices can be found if productivity gains or losses are meaningful. Although not statistically significant, differences can be observed by the smaller expected composite indexes and the broader 90% confidence intervals, when compared dynamic models with the static ones.

Consider the scenario *I*. On average, PPPs are slightly better socially performers than public hospitals, with a difference ranging from 2 to 6%, depending on the dynamic nature of the model. That is, the efficiency spread of PPPs is smaller than the one associated with public hospitals, and with respect to their own-group technology. The same does apply regarding the metatechnology. However, the differences are not meaningful in the statistical sense and, at least at the 10% significance level, because the 90% confidence intervals (associated with those two clusters of health care providers) overlap within a considerable range of efficiencies. Furthermore, the gap between each cluster's frontier and the metafrontier is, on average, small and does not go beyond 7.5% (in the case of PPPs). The gap associated with PPPs seems to be slightly larger when compared with public hospitals, which is corroborated by quantities $\mathscr{F}(\tau)$ for $\tau = 0.95$ (frequency of best performer). Once more, statistical evidence does not allow rejecting the null hypothesis and both clusters of hospitals seem to be at equal distances to the overall technology. In other words, the metafrontier is

Table 5
Social performance of public hospitals and PPPs: Some basic statistics.

			Performance regarding the own-group technology				Performance regarding the metatechnology				Metatechnology ratio			
Model/scenario		Cluster[a]	$\mathbb{E}(\theta)$[b]	$\min\{CI\}$[c]	$\max\{CI\}$[d]	$\mathscr{F}(\tau)$[e]	$\mathbb{E}(\theta)$	$\min\{CI\}$	$\max\{CI\}$	$\mathscr{F}(\tau)$	$\mathbb{E}(\eta)$	$\min\{CI\}$	$\max\{CI\}$	$\mathscr{F}(\tau)$
S_I		A	0.7961	0.7894	0.8303	24.3528	0.7784	0.7705	0.8159	23.3941	0.9963	0.9945	0.9983	92.8092
		B	0.8210	0.8022	0.8531	20.3390	0.7855	0.7689	0.8369	16.9492	0.9698	0.9251	0.9958	77.9661
D_I		A	0.6939	0.6439	0.8193	11.9847	0.6841	0.6340	0.8068	11.8888	0.9973	0.9872	1.0021	92.6174
		B	0.7679	0.7202	0.8576	16.9492	0.7385	0.6927	0.8160	10.1695	0.9711	0.9064	1.0027	71.1864
S_{II}	Pessimistic	A	0.7173	0.7116	0.7492	46.0970	0.6926	0.6787	0.7224	44.1249	0.9910	0.9861	1.0000	88.7017
		B	0.5745	0.5152	0.6468	35.0000	0.5085	0.4908	0.6681	32.1429	0.9638	0.9533	1.1659	59.6429
	Optimistic	A	0.7977	0.7909	0.8243	61.1750	0.7871	0.7789	0.8142	59.9836	0.9994	0.9983	1.0000	92.9334
		B	0.7565	0.7429	0.7772	56.0714	0.7208	0.7118	0.7726	54.2857	0.9584	0.9173	0.9987	70.7143
D_{II}	Pessimistic	A	0.5688	0.5041	0.7628	31.6352	0.5391	0.4657	0.7336	30.1150	0.9741	0.9209	0.9984	84.8398
		B	0.4036	0.1979	0.7125	13.9286	0.3188	0.1766	0.4954	13.9286	0.9599	0.9463	1.5345	32.5000
	Optimistic	A	0.6573	0.6134	0.7616	39.8110	0.6438	0.5991	0.7506	38.4552	0.9958	0.9857	1.0018	90.7560
		B	0.6139	0.5364	0.7794	32.8571	0.5670	0.5000	0.7211	30.0000	0.9202	0.7929	0.9899	50.3571

[a] A – Public hospitals; B – PPPs.
[b] Expected value of the social performance indicators.
[c] Lower bound of the 90% confidence interval.
[d] Upper bound of the 90% confidence interval.
[e] $\mathscr{F}(\tau)$ is the probability of getting hospitals whose social performance is, at least, equal to τ, here defined as $\tau = 0.9500$.

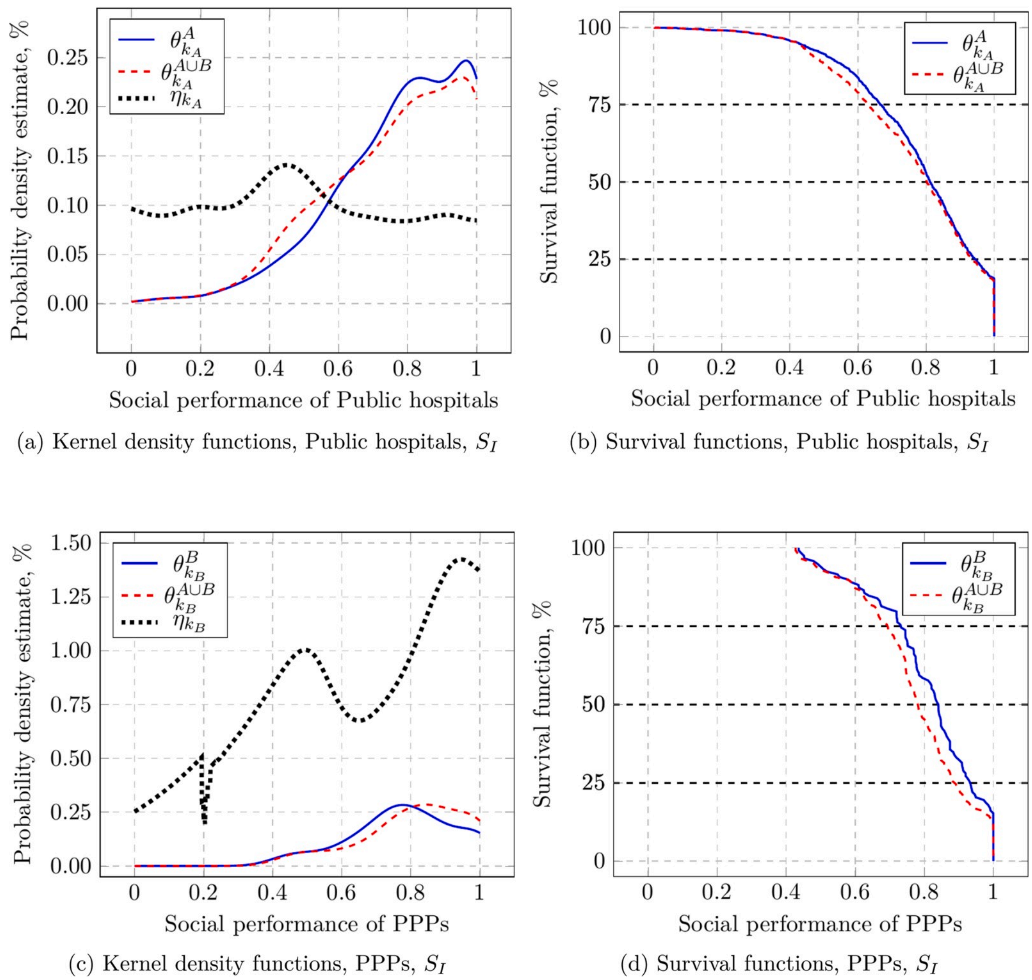

Fig. 5. Nonparametric Gaussian kernel based density and survival functions for both public hospitals and PPPs, model S_I.

Table 6
Statistical analysis (*p*-values) to compare the gaps of both groups of hospitals regarding the metafrontier – Scenario *I* (no missing data).

	$w=-1$		$w=0$		$w=+1$	
	S_I	D_I	S_I	D_I	S_I	D_I
$p_1^{(w)}$	0.2270	0.4260	0.2270	0.4420	0.4010	0.4460
$p_2^{(w)}$	0.7730	0.5740	0.7730	0.5580	0.5990	0.5540
$p_3^{(w)}$	0.4540	0.8520	0.4540	0.8840	0.8020	0.8120

composed of benchmarks of both PPPs and public hospitals. This conclusion is reinforced by the three bootstrap-based statistical tests, as all of the *p*-values associated with both S_I and D_I are larger than 0.10; *i. e.*, there is not enough evidence that $T_b^{(w)} \neq T_{\text{obs}}^{(w)}$ for any order $w \in \{-1, 0, 1\}$ and $b \in \{1, \ldots, \mathscr{B}\}$, for $\mathscr{B} = 1,500$ iterations. That is, the null hypothesis cannot be rejected and both groups – PPPs and public hospitals – exhibit similar social performances regarding what is theoretically achievable.

If dynamic effects play a central role on social performance assessment, then so does data precision. In general, results are sensitive to the sample size, to entered data, and ultimately to the ranges assumed to replace gaps. Through scenario *II*, one uses a sample that is 2.29 times larger than the sample length associated with scenario *I*. Because in

Table 7
Statistical analysis (*p*-values) to compare the gaps of both groups of hospitals regarding the metafrontier – Scenario *II* (optimistic and pessimistic models).

		$w=-1$		$w=0$		$w=+1$	
		S_{II}	D_{II}	S_{II}	D_{II}	S_{II}	D_{II}
Optimistic	$p_1^{(w)}$	0.4620	0.4370	0.4950	0.4310	0.6790	0.4850
	$p_2^{(w)}$	0.5360	0.5570	0.4980	0.5470	0.3210	0.4990
	$p_3^{(w)}$	0.9240	0.8740	0.9900	0.8620	0.6420	0.9700
Pessimistic	$p_1^{(w)}$	0.6810	0.7070	0.6260	0.1260	0.4660	0.1390
	$p_2^{(w)}$	0.0930	*0.0240*	0.0840	0.3520	0.3660	0.5600
	$p_3^{(w)}$	0.1860	0.0480	0.1680	0.2520	0.7320	0.2780

scenario *II* hospitals can search for benchmarks in a broader range of possibilities (when compared with scenario *I*), it is understandable that performance levels are, in general, smaller. Although this would constitute a pitfall when comparing results between scenarios in most of the empirical cases, this study has overcome it by using $N = 38$ within the subsampling procedure of Algorithm 1, which is equally consistent among scenarios. *Viz.*, results are comparable among different scenarios and models. According to Table 5 and the model S_{II}, both pessimistic and optimistic cases return different social performance scores, for both groups of hospitals. Pessimistic-based social performance levels are, of course, smaller than the optimistic ones. This happens regarding both own-group and metatechnology frontiers. However, when dynamic effects are considered, differences between pessimistic and optimistic cases become meaningless in the light of statistical evidence, as 90% confidence intervals tend to overlap. By means of model D_{II}, the expected own-group performance associated with clusters A – public hospitals and B – PPPs belong to the ranges $[0.5041, 0.7628]$ and $[0.1979, 0.7794]$, respectively. Note the considerable range associated with the dynamic social performance of PPPs, which seems to indicate that these entities are expected to be either very ineffective or moderately effective. Likewise, when the metatechnology is assumed, those ranges become $[0.4657, 0.7506]$ and $[0.1766, 0.7211]$, respectively. This means that, when both frontiers are compared and the metatechnology ratio is assessed, no meaningful differences between clusters can be found. Overall, PPP benchmarks have a margin of, at the most, 5.37% to improve their social performance with respect to the what is theoretically possible (metatechnology), whereas such a quantity slightly increases to 7.91% for public hospitals best practices. Once again, and as predicted by both confidence intervals and the *p*-values presented in Table 7, public hospitals and PPPs present similar social performances. Note that, in some cases, the upper bound associated with the metatechnology ratio, η, can be larger than 1. This would not be expected under deterministic frameworks, but the same cannot be said for the stochastic ones, which include the Algorithm 1.

The absence of a considerable difference between social performance levels of both groups can be partially explained by the competition in hospital services. For instance Ref. [76] argued that private hospitals, including PPPs, compete on prices and quality of services for contracts with insurance companies or health subsystems. Furthermore, they try attracting patients covered by health insurances based on criteria like perceived clinical quality and access to health care, namely as measured by services availability. Therefore, PPPs are not expected to be outperformed by public hospitals in social terms.

The fact that both groups of hospitals exhibit similar social performances levels means neither these groups are good societal performers neither that inefficiency (or ineffectiveness) sources are the same for them both. As seen in Table 5, there is a considerable room for improvement in terms of social performance of the two groups of hospitals. This is a common issue of all adopted models, although the margin for enhancement is larger in pessimistic models, especially the dynamic one. However, the Benefit-of-Doubt-based composites cannot be seen as factors imposing equal reduction/expansion of undesirable/desirable criteria, in a radial sense, because the model underlying the composites computation is nonradial. In fact, the PPP-based Benefit of Doubt model maximises shortage/excess slacks by optimising dual variables (virtual weights), such that performance of the hospital under scrutiny should be the best possible (PPP is a benevolent technique).

This means that each hospital has its own inefficiency sources. According to the worst scenario (dynamic pessimistic model), public hospitals can increase their desirable criteria, g_ℓ^+, by 23, 29, 9, and 4 pp. (on average) for $\ell = 1, \ldots, 4$, and reduce their undesirable criteria, g_ℓ^-, by 22, 16, 3, 1.5, 4, 5, 14 pp. (on average), for $\ell = 5, \ldots, 11$. Likewise, hospital PPPs have room to enhance their desirable criteria into 26, 33, 10 and 5 pp. (on average) and their undesirable criteria into 18, 4, 1.7, 4.4, 6, and 16 pp. (on average). Note that the average improvements are similar for both groups of hospitals, which results from the fact that they are equidistant from the metatechnology. Note also that these results do not translate the conclusions made in Subsection 3.3 about Table 2, which, as aforementioned, are most likely flawed as they rely solely on statistical tests that ignore the heterogeneous environment under which hospitals deliver health care to citizens. Unfortunately, given the considerable amount of powerful techniques as the presented one, most of studies and reports insist on comparing groups of entities through weak statistical tests that mostly provide unreliable biased outcomes.

6. Implications for policy makers, regulators, citizens, hospital managers, clinical staff, academics, and project managers

These results have strong implications for policy makers, citizens, hospital managers and clinical staff, operational researchers and other academics, and importantly project managers.

(R_1) POLICY MAKERS AND REGULATORS. The decision on starting a new PPP project is usually up to policy makers. Given the social concerns (including the potential moral hazard) associated with the entry of private partners into the NHS, policy makers would be tempted to avoid new projects, to revoke existing ones, and/or to bring into the public sphere the private arrangements that will finish in the short-run. For instance, the Cascais' hospital PPP is managed by two interrelated private partners: one for the infrastructure and ancillary services, and the second associated with the clinical activity. The contract between the State and the clinical-related private partner of this PPP is about to finish, and the Portuguese Government is struggled with three possibilities: a contract renewal with the same private partner, a new contract with another private entity (after public tender), or transform the private clinical management into a public one – mostly by, perhaps populist and uninformed, political pressures. Apparently, results drawn by this study may help on this decision (and others later on): since public hospitals and PPPs exhibit, in general, similar social performances – in terms of timely, equitable, appropriate, and safe health care – there is no reason to believe that changing the private/public nature of the clinical management is a good idea. The third option would introduce considerable switchingcosts, making it an unsustainable/inefficient/ineffective alternative for the exchequer. However, the fact that the two groups present similar levels of social performance does not mean that they are good performers. In fact, considerable margins for improvement were observed. As an example, the rate of hip surgeries in the first 48 h after hip fracture would ideally be equal to 100%, but one may observe that, on average, hospitals are far from that objective – and PPPs outperform public hospitals in this dimension. Penalties and rewards should be strengthened in contracts, and attributed to the worst and to the best practices in the field, respectively. The regulator(s) also plays a central role here because it must receive social performance-based indicators from both public hospitals and PPPs, analyse them to check for inconsistencies/missing data/unreliable data, and compare them within a benchmarking exercise like the one executed in this study.

(R_2) CITIZENS. There has been a (perhaps uninformed) political campaign aimed at biasing public opinion about PPPs in the health sector. One of the most used arguments is, precisely, that these entities do not provide the same level of quality and access as public hospitals, since the former are managed by private partners whose ultimate objective is to maximise profit, in detriment of the social dimension. Evidence provided in this study suggests that this is not necessarily true as the demand for profit does not seem to compete with the provision of better health care services. It is important to keep citizens informed based on empirical evidence, because their judgements may play a central role on national policy.

(R_3) HOSPITAL MANAGERS AND CLINICAL STAFFS. Delivering equitable, timely, appropriate,and safe care services to all citizens, regardless of their willingness to pay, is the mission of any health care provider should it be financed via taxes according to the Beveridge system (or alike). In the same sense, searching for better practices operating in similar conditions is a must-do critical strategy that should be carried out by both hospital managers and clinical staff. Given the results here presented, the clinical staff working in one of those two groups of hospitals should be able to copy good (better) practices observed within both groups, i.e., the benchmarks belonging to the metafrontier that can be either PPPs or public hospitals.

(R_4) OPERATIONAL RESEARCHERS AND OTHER ACADEMICS. This study provides interesting and important mathematical tools that can be used in a broad range of fields. Measuring performance of firms has been a common practice as efficiency, effectiveness, and sustainability are important parts of their strategies. To improve performance, one must compare itself against better practices in the field, and nonparametric benchmarking techniques have been widely employed for such a goal. This study assembles some of the most important, recent, and solid techniques to provide bias- and environment-corrected composite indexes, which, by their nature, are neither affected by data noise/imprecision nor by the sample heterogeneity.

(R_5) PROJECT MANAGERS. In general, project managers are concerned about project financial sustainability and viability, time-scale, investments, funding, risk assessment, and the definition and monitoring of outcomes, including the quality of deliverables. Indeed, most of projects associated with hospital PPPs do not end with the building construction; instead, the former go beyond that and must ensure that all stakeholders' concerns are met. Because health care carries a huge societal weight, the results found in this research point towards the convergence between the project manager and the remaining stakeholders' interests. In view of that, it seems that there is no conflict of interests, and value for money is being created.

7. Summary, limitations, and future work

The present study investigates whether public hospitals can exhibit better social performance levels than PPP hospitals. With strong implications for policy makers, project managers, hospital managers and workforce, and for the public opinion, empirical evidence is peremptory: public hospitals are not better social performers than PPPs, *i.e.*, the former usually do neither deliver timelier nor safer nor more appropriate health care services than the latter. However, both groups exhibit considerable social inefficiency levels. The potential for improvement is larger within dimensions related to care appropriateness – rate of minor surgeries per potential outpatient procedure, rate of readmissions after 30 days of discharge, rate of inpatients staying more than a month in the hospital ward –, and timeliness – rate of first medical appointments within the maximum legislated guaranteed time, and average delay before surgery.

This study provides a powerful tool for performance assessment when only indicators (rather than volume measures) are available, when there are missing/unreliable/biased data, and when heterogeneity among hospitals (or any other group of entities) is considerable. The framework here detailed can be employed to any scenario, *e.g.*, to rank projects based on a set of indicators reflecting their financial/economic, environmental, and social sustainability.

Although important under the societal point of view, the results here achieved are naturally not definitive and, for that reason, they must be compared with others constructed *a posteriori*, which, in turn, should account for more quality and access related variables. This study has considered only process of care-related dimensions, but structural quality (facility-related organisational attributes) and outcomes (overall effects of care) are also important ones. For instance, patients' satisfaction is a relevant outcome as it is strictly related to the treatments' successfulness [77]. Another important outcome refers to the in-hospital avoidable death rate, usually measured by the mortality of low/medium severity groups of diagnosis. If a patient is classified within a low severity group and dies anyway, then the quality of the health care provider is certainly poor and requires critical improvements. Quality and access to health care services are two complex and interrelated terms. For this reason, evaluating the social performance of a hospital requires the definition of dimensions related to them both. This study has considered only four access-related variables, but many more could be used instead, particularly to characterise the services availability: *e.g.*, full-time equivalent nurses or doctors per 1,000 standard patients or inhabitants, inpatient bed occupancy rate, operating theatre capacity utilisation, average delay in inpatient services (days per inpatient), waiting time for a medical appointment, and waiting lists' length, to name a few. Considering these new variables, which were disregarded from the current research due to meaningful gaps of information,one should recompute the social performance scores for both public and PPP hospitals, and compare them following the same procedure. This exercise would complement the analysis performed by the current research.

Social performance is a pillar of the NHS hospitals governance. Nonetheless, ensuring their financial sustainability, as well as the contractor's, is also relevant for the public accounts. In view of that, the two groups of hospitals analysed in this research should be benchmarked and compared using financial inputs and raw outputs (as delivered services, disregarding social effects). Eventually, a bias-/heterogeneity-corrected input-oriented PPP model should be sufficient for this purpose. These results must, then, be compared with the social performance Benefit-of-Doubt-based scores to test whether any potential trade-off between social and financial/economic goals of these institutions. It is worth mentioning at this point, as hint, that such a trade-off is unlikely: PPP hospitals are deemed as more technically efficient than public hospitals; still, no evidence supports the hypothesis that public hospitals are more effective (generically, exhibit better social performance levels) than the former group. This means that no link between technical efficiency – as proxy for the financial sustainability – and social performance is expected. Private management-related practices can deliver health care services at least as good as the public providers. Hence, the heart of the matter should lie mainly in maximising the value of money for the public purse.

Finally, it should be mentioned that the reproducibility of these conclusions may be limited to the cases where clinical services are included into the hospital PPP (bundled) arrangement, as the Portuguese first wave, the Spanish Alzira models, and the full-service provision health care model in some German hospital chains. In more conservative (unbundled) frameworks, like the British one, only the infrastructure and the ancillary services are privately managed, whereas clinical services are delivered by a public entity. This means that, in these cases, hospital PPPs are not expected to be different from the remaining public hospitals, although there may be an interface problem given potential conflict of interests of the two main players.

Declaration of competing interest

Authors declare that they have no conflict of interest.

Acknowledgement

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.seps.2020.100798.

References

[1] Woodson TS. Public private partnerships and emerging technologies: a look at nanomedicine for diseases of poverty. Res Pol 2016;45(7):1410–8.
[2] Hodge GA, Greve C. Public-private partnerships: an international performance review. Publ Adm Rev 2007;67(3):545–58.
[3] Cruz CO, Marques RC. Flexible contracts to cope with uncertainty in public-private partnerships. Int J Proj Manag 2013;31(3):473–83.
[4] Marques RC, Berg S. Revisiting the strengths and limitations of regulatory contracts in infrastructure industries. J Infrastruct Syst 2010;16(4):334–42.
[5] Law Company Practical. Public private partnerships: issues and considerations. URL, http://us.practicallaw.com/3-504-9995; 2013.
[6] Mota J, Moreira AC. The importance of non-financial determinants on public-private partnerships in europe. Int J Proj Manag 2015;33(7):1563–75.
[7] Cruz CO, Marques RC. O Estado e as Parcerias Público-Privadas. Lisbon, Portugal: Sílabo; 2012.
[8] Hong H, Shum M. Increasing competition and the winner’s curse: evidence from procurement. Rev Econ Stud 2002;69(4):871–98.
[9] Tucker AL. The impact of operational failures on hospital nurses and their patients. J Oper Manag 2004;22(2):151–69.
[10] Wood DJ. Corporate social performance revisited. Acad Manag Rev 1991;16(4): 691–718.
[11] Kokko S, Lukkarinen J, Mark-Hebert C. Public-private partnerships and social performance in the forest industry. Soc Bus 2014;4(1):15–31.
[12] Brandão C, Rego G, Duarte I, Nunes R. Social responsibility: a new paradigm of hospital governance? Health Care Anal 2013;21:390–402.
[13] Cruz C, Marques RC. Integrating infrastructure and clinical management in PPPs for health care. J Manag Eng 2013;29(4):471–81.
[14] Donabedian A. The quality of care: how can it be assessed? J Am Med Assoc 1988; 260(12):1743–8.
[15] Donabedian A. Evaluating the quality of medical care. Milbank Q 2005;83(4): 691–729.
[16] Bernet PM, Moises J, Valdmanis VG. Social efficiency of hospital care delivery: frontier analysis from the consumer’s perspective. Med Care Res Rev 2011;68(1): 36S–54S.
[17] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 1984;30(9):1078–92.
[18] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. Eur J Oper Res 1979;3(4):201–9.
[19] Berenguer Gemma, Iyer Ananth V, Yadav Prashant. Disentangling the efficiency drivers in country-level global health programs: an empirical study. J Oper Manag 2016;45:30–43.
[20] Jola-Sanchez Andres F, Alfonso J, Pedraza-Martinez, Bretthauer Kurt M, Britto Rodrigo A. Effect of armed conflicts on humanitarian operations: total factor productivity and efficiency of rural hospitals. J Oper Manag 2016;45:73–85.
[21] Ferreira DC, Marques RC. Malmquist and Hicks-Moorsteen productivity indexes for clusters performance evaluation. Int J Inf Technol Decis Making 2016;15(5): 1015–53.
[22] Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. Health Care Manag Sci 2003;6(4):203–18.
[23] Hollingsworth B. The measurement of efficiency and productivity of health care delivery. Health Econ 2008;45:1107–28.
[24] Ferreira DC, Marques RC. Did the corporatization of Portuguese hospitals significantly change their productivity? Eur J Health Econ 2015;16(3):289–303.
[25] Lindlbauer I, Winter V, Schreyögg J. Antecedents and consequences of corporatization: an empirical analysis of German public hospitals. J Publ Adm Res Theor 2016;26(2):309–26.
[26] Barlow J, Koberle-Gaiser M. The private finance initiative, project form and design innovation: the UK’s hospitals programme. Res Pol 2008;37(8):1392–402.
[27] McKee M, Edwards N, Atun R. Public-private partnerships for hospitals. Bull World Health Organ 2006;84(11):890–6.
[28] A.M.A. Oliveira. reportPublic-private partnerships and public hospitals in Portugal: evidence from delivery data. Master’s thesis, Instituto Superior Técnico, University of Lisbon, Portugal.
[29] Roehrich JK, Lewis MA, George G. Are public-private partnerships a healthy option? A systematic literature review. Soc Sci Med 2014;113:110–9.
[30] Kula N, Fryatt RJ. Public-private interactions on health in South Africa: opportunities for scaling up. Health Pol Plann 2014;29(5):560–9.
[31] Torchia M, Calabrò A, Morner M. Public-private partnerships in the health care sector: a systematic review of the literature. Publ Manag Rev 2015;17(2):236–61.
[32] Pita Barros P, Martinez-Giralt X. Contractual design and PPPs for hospitals: lessons for the Portuguese model. Eur J Health Econ 2009;10(4):437–53.
[33] Barros PP. Economia da Saúde: conceitos e Comportamentos. Portugal: Edições Almedina, Coimbra; 2009.
[34] Ferreira DC, Marques RC, Nunes AM. Economies of scope in the health sector: the case of Portuguese hospitals. Eur J Oper Res 2017. https://doi.org/10.1016/j.ejor.2017.09.044.
[35] Berwick D, Fox DM. “Evaluating the quality of medical care”: Donabedian’s classic article 50 years later. Milbank Q 2016;94(2):237–41.
[36] Davies HTO, Crombie IK. Assessing the quality of care. BMJ 1995;311:766.
[37] Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. BMJ 1995;311: 793–6.
[38] Navarro-Espigares JL, Torres EH. Efficiency and quality in health services: a crucial link. Serv Ind J 2011;31(3):385–403.
[39] Nayar P, Ozcan YA. Data envelopment analysis comparison of hospital efficiency and quality. J Med Syst 2008;32(3):193–9.
[40] Ferrier GD, Trivitt JS. Incorporating quality into the measurement of hospital efficiency: a double DEA approach. J Prod Anal 2013;40(3):337–55.
[41] Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? Milbank Q 2005;83(4):843–95.
[42] Bilsel M, Davutyan N. Hospital efficiency with risk adjusted mortality as undesirable output: the Turkish case. Ann Oper Res 2014;221(1):73–88.
[43] Chang SJ, Hsiao HC, Huang LH, Chang H. Taiwan quality indicator project and hospital productivity growth. Omega: Int J Manag Sci 2011;39(1):14–22.
[44] Clement JP, Valdmanis VG, Bazzoli GJ, Zhao M, Chukmaitov A. Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion. Health Care Manag Sci 2008;11(1):67–77.
[45] Du J, Wang J, Chen Y, Chou SY, Zhu J. Incorporating health outcomes in Pennsylvania hospital efficiency: an additive super-efficiency DEA approach. Ann Oper Res 2014;221(1):161–72.
[46] Ferrier GD, Valdmanis V. Rural hospital performance and its correlates. J Prod Anal 1996;7(1):63–80.
[47] Gok MS, Sezen B. Analyzing the ambiguous relationship between efficiency, quality and patient satisfaction in healthcare services: the case of public hospitals in Turkey. Health Pol 2013;111(3):290–300.
[48] Matranga D, Sapienza F. Congestion analysis to evaluate the efficiency and appropriateness of hospitals in Sicily. Health Pol 2015;119(3):324–32.
[49] Mutter R, Valdmanis V, Rosko M. High versus lower quality hospitals: a comparison of environmental characteristics and technical efficiency. Health Serv Outcome Res Methodol 2010;10(3–4):134–53.
[50] Prior D. Efficiency and total quality management in health care organizations: a dynamic frontier approach. Ann Oper Res 2006;145(1):281–99.
[51] Tiemann O, Schreyögg J. Effects of ownership on hospital efficiency in Germany. Bus Res 2009;2(2):115–45.
[52] Valdmanis VG, Rosko MD, Mutter RL. Hospital quality, efficiency, and input slack differentials. Health Serv Res 2008;43(5):1830–48.
[53] Varabyova Y, Blankart CR, Schreyögg J. Using nonparametric conditional approach to integrate quality into efficiency analysis: empirical evidence from cardiology departments. Health Care Manag Sci 2016. https://doi.org/10.1007/s10729-016-9372-4.
[54] Wu CH, Chang CC, Chen PC, Kuo KN. Efficiency and productivity change in Taiwan’s hospitals: a non-radial quality-adjusted measurement. Cent Eur J Oper Res 2013;21(2):115–45.
[55] Ellis L, Woods LM, Estéve J, Eloranta S, Coleman MP, Rachet B. Cancer incidence, survival and mortality: explaining the concepts. Int J Canc 2014;135(8):1774–82.
[56] Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, Hudson M. What does ’access to health care’ mean? J Health Serv Res Pol 2002;7 (3):186–8.
[57] Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH. Poverty and access to health care in developing countries. Ann N Y Acad Sci 2008;1136:161–71.
[58] Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. JAMA. Internal. Med. 2000;160:1074–81.
[59] Ferreira DC, Marques RC. Identifying congestion levels, sources and determinants on intensive care units: the Portuguese case. Health Care Manag Sci 2016:1–28.
[60] Lyder CH, Wang Y, Mtersky M, Curry M, Kliman R, Verzier N, Hunt DR. Hospital-acquired pressure ulcers: results from the national medicare patient safety monitoring system study. J Am Geriatr Soc 2012. https://doi.org/10.1111/j.1532-5415.2012.04106.x.
[61] Sullivan N, Schoelles KM. Preventing in-facility pressure ulcers as a patient safety strategy: a systematic review. Ann Intern Med 2013;158:410–6.
[62] Loiselle A, Parish JM, Wilkens JA, Jaroszewski DE. Managing iatrogenic pneumothorax and chest tubes. J Hosp Med 2013;8(7):402–8.
[63] Schang L, Hynninen Y, Morton A, Salo A. Developing robust composite measures of healthcare quality: ranking intervals and dominance relations for scottish health boards. Soc Sci Med 2016. https://doi.org/10.1016/j.socscimed.2016.06.026.
[64] Shieh L, Go M, Gessner D, Chen JH, Hopkins J, Maggio P. Improving and sustaining a reduction in iatrogenic pneumothorax through a multifaceted quality-improvement approach. J Hosp Med 2015;10(9):599–607.
[65] Lee DJ, Elfar JC. Timing of hip fracture surgery in the elderly. Geriatr Orthop Surg Rehabil 2014;5(3):138–40.
[66] Zuckerman JD. Hip fracture. N Engl J Med 1996;334(23):1519–25.
[67] Ferreira DC, Marques RC. Should inpatients be adjusted by their complexity and severity for efficiency assessment? Evidence from Portugal. Health Care Manag Sci 2016;19(1):43–57.
[68] Grossman M. On the concept of health capital and the demand for health. J Polit Econ 1972;80(2):223–55.
[69] Portela MCAS, Thanassoulis E, Simpson G. Negative data in DEA: a directional distance approach applied to bank branches. J Oper Res Soc 2004;55(10):1111–21.
[70] Cherchye L, Moesen W, Rogge N, Puyenbroeck TV. Constructing composite indicators with imprecise data: a proposal. Expert Syst Appl 2011;38(9):10940–9.
[71] Zanella A, Camanho AS, Dias TG. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. Eur J Oper Res 2015;245 (2):517–30.
[72] Zha Y, Song A, Xu C, Yang H. Dealing with missing data based on data envelopment analysis and halo effect. Appl Math Model 2013;37(9):6135–45.
[73] Smirlis YG, Maragos EK, Despotis DK. Data envelopment analysis with missing values: an interval dea approach. Appl Math Comput 2006;177(1):1–10.
[74] Kuosmanen T. Modelling blank data entries in data envelopment analysis. URL, http://econwpa.repec.org/eps/em/papers/0210/0210001.pdf; 2001.

[75] Despotis DK, Smirlis YG. Data envelopment analysis with imprecise data. Eur J Oper Res 2002;140(1):24–36.
[76] Barros PP. Competition policy for health care provision in Portugal. Health Pol 2017;121:141–8.
[77] Ferreira DC, Marques RC, Nunes AM, Figueira JR. Patients' satisfaction: the medical appointments valence in Portuguese public hospitals. Omega: Int J Manag Sci 2017. https://doi.org/10.1016/j.omega.2017.08.009.