Portuguese Economic Journal (2022) 21:351–381
https://doi.org/10.1007/s10258-022-00223-0

ORIGINAL ARTICLE

Quality decreases from introducing patient choice in a National Health Service

Pedro Pita Barros[1]

Received: 6 January 2022 / Accepted: 11 October 2022 / Published online: 24 October 2022

Abstract
A view often expressed about patient choice of health care providers is that it will increase competition between providers, which benefits the efficiency of the health system. We address here a patient choice initiative, regarding selection of hospital for specialty consultations, in the Portuguese National Health Service (NHS) that has two specific features. The first feature involves shared decision making between patients and GPs, in the choice of hospital for referral, which should be based on publicly available information on "quality". The second specific feature is that the patient choice initiative did not involve payment changes to NHS hospitals associated with patients' movements. We show that explaining initial asymmetries in qualities (waiting times) with systematic differences in hospital characteristics (cost advantages and managerial talent) leads to potential asymmetric responses to the introduction of patient choice in the NHS. This implies that the empirical analysis has to accommodate such asymmetries. Explicitly allowing for asymmetries in responses to the policy measure reveals that reactions were indeed different, with top-performance hospitals reducing their qualities (increasing waiting times) after the patient choice initiative was introduced.

Keywords Hospital quality · Patient choice

1 Introduction

A view often expressed about patient choice of health care providers is that it will increase competition between providers, which benefits the efficiency of the health system. This view on the benefits of competition, common in many, if not most, markets outside the health sector, requires some careful analysis when applied to health care markets. These are markets with several imperfections and often extensive regulation is present. Thus, policy makers need to be more cautious about the effects of

✉ Pedro Pita Barros
ppbarros@novasbe.pt

[1] Nova School of Business and Economics, Universidade Nova de Lisboa, Lisbon, Portugal

policies that aim at increasing competition in the healthcare marketplace. Competition should be seen as a tool, as a policy instrument, and not as an end in itself.[1]

We address here a patient choice initiative in the Portuguese National Health Service (NHS) that has two particular features. It involves shared decision making between patients and GPs, in the choice of hospital for referral, which should be based on publicly available information on “quality”. This patient choice initiative (described below), unlike what occurred in other countries, did not involve payment changes to NHS hospitals, which are funded by global budgets and not paid directly by episode of treatment. Thus, endogenous “quality” choices by hospitals can be considerably different from the cases in which more demand is associated with more net revenues (while in the patient choice initiative addresses here, more demand results in more costs but not more revenues). And the way different providers react is related to initial asymmetries in “quality” across hospitals, requiring an adequate empirical strategy. The empirical findings corroborate the importance to distinguish different reactions to the policy choice initiative by providers, according to their initial asymmetries in quality choices.

There is a growing health economics literature on the introduction of competition in health care markets. The work of Siciliani and Straume (2019), building on the model introduced by Brekke et al. (2011), provides a useful framework to discuss quality choices under regulated prices (although the existence of regulation is not explained). It sets out that ambiguous effects emerge, with particular outcomes being dependent on technical aspects of benefit functions, among other aspects. This leaves for empirical findings to sort out the actual effect of patient choice reforms.

The main literature on the role of competition for hospital quality decisions is summarised in Brekke et al. (2021). Previous literature has found, in a majority of cases, positive effects associated with competition (free choice by patients), although some negative impacts have also been documented. Such previous literature has not considered, up to our knowledge, a NHS context with hospital funding not responding to patient flows (by design) and explicit consideration of asymmetric responses by providers to the policy reform.

The work of Brekke et al. (2021) considers a patient choice reform in Norway, as hospitals in a NHS were exposed to non-price competition similar to the one explored here. Some crucial differences exist between the two reforms, in Portugal and in Norway, while similarities can also be found. The reform in Norway is the closest to the one that took place in the Portuguese NHS, with a crucial difference being the absence of direct revenue effects in the Portuguese case. Revenue streams are guided by patient flows in the Norwegian case. Bjorvatn and Ma (2011), looking at the 2001 choice reform in Norway, show that differences in waiting times across hospitals determined mobility flows (mainly in younger patients). Brekke et al. (2021) document a (larger) decline in waiting times in hospitals facing stronger competition. They take their testable hypothesis from a model assuming a

[1] Barros et al. (2016) discuss the role of competition in health care markets.

symmetric equilibrium, and as such all hospitals are expected to behave in the same way in response to the policy reform introducing free choice of hospital by patients.

However, there is no reason to use such assumption, at least in the Portuguese context, as it will be detailed below. Differences in qualities (waiting times, in the empirical discussion) across hospitals before the policy reform arise due to systematic hospital differences and are not the outcome of random elements. Thus, the source of the initial asymmetric qualities of hospitals is also a likely source for asymmetric reactions to the introduction of a free choice policy of hospital by patient (and GPs).

Moscelli et al. (2019) find that relaxing patient constraints on hospital choice in England resulted in lower quality in some procedures (hip and knee replacement) and no effect on others (coronnary bypass). The authors mention the role of (lack of) profitability in those procedures is behind their findings. The view that increasing competition amongst hospitals leads to higher quality is common, but does not seem to hold in several empirical settings. Gaynor (2006) provides a simple argument for this to be the case when hospitals face regulated prices and the price received by treating an extra patient is above the corresponding cost. Other models question this rationale by introducing other features from health systems (Katz 2013; Brekke et al. 2014).

The model laid down by Moscelli et al. (2019) illustrates how the effect of competition depends on the properties of the cost function and the degree of altruism (by providers). We differ from their model in one important respect: we consider heterogeneous providers, giving origin to different starting points and to different responses to the policy change. We share with Moscelli et al. (2019) the explicit inclusion of the role of altruism, as it may lead to treatment of patients and choice of quality levels that bring a loss to the provider.

In a different direction, Santos et al. (2015) document the impact of a patient choice initiative in England upon choices in primary care use. Their starting point is that “a necessary condition for greater patient choice to improve quality is that a provider will face higher demand if they improve their quality” (Santos et al. 2015, p. 445). Implicit is the assumption that higher demand will bring more revenues. Their main finding is that patients do indeed react to differences in quality across providers.[2] Thus, patient choice reforms across different health systems and contexts yield context-dependent results.

We explore below a patient choice initiative in the Portuguese NHS. The first element of analysis, before proceeding to the data, is to establish in a clear way the expected effects of the policy, given the institutional features of Portuguese National Health Service. We show that explaining initial asymmetries in qualities (waiting times) with systematic differences in hospital characteristics (cost advantages, managerial talent) leads to potential asymmetric responses to the introduction of patient choice in the NHS. This implies that the empirical analysis has to accommodate such asymmetries. Explicitly allowing for it reveals that reactions are indeed different,

[2] See Santos et al. (2015) for a brief review of studies on patient choice and its relation to information on “quality”.

with top quality hospitals reducing their qualities (increasing waiting times) after the patient choice initiative is introduced.

The paper is organised as follow. Section 2 briefly details the patient choice initiative. Next, Sect. 3 presents a simple model illustrating the main economic forces at play. Section 4 reports on the data and data sources for the empirical analysis. Then, Sect. 5 presents the main empirical results. Section 6 reports on robustness checks. Finally, Sect. 7 has concluding remarks.

2 The policy

In 2016, the Portuguese Ministry of Health enacted a normative document (Despacho 6170-A/2016, of May 9), establishing a patient choice initiative (PCI). In this initiative, the patient, jointly with the GP, decides which hospital in the National Health Service (wherever located in the country) to get an appointment for a hospital specialty consultation. In the Portuguese National Health Service, patients register with a GP in an NHS primary care unit, usually located in their residential area. Prior to the patient choice initiative, if a hospital referral was needed, the patient would be automatically directed to the NHS hospital to which network the primary care unit belongs to.[3] Previously to this initiative, the match of patient to hospital was based on regulated catchment areas defined by geography, and which may differ according to the relevant hospital specialty (as not all hospitals offer the full spectrum of medical specialties). The choices presented to the patients, under the PCI, are based on geographic proximity and on available information of average waiting times per specialty in each hospital. When surgery is relevant, the information on waiting time for surgical intervention is provided to the decision process. The information on waiting times for first time consultations in each hospital is available online, at the NHS web portal, including the information on how many people are in the waiting list according to priority level.

This patient choice initiative was expected to create several types of benefits, according to the NHS documents supporting its adoption. To patients, the benefit of choice of NHS hospital and increased trust in the NHS. To the GP, an increased role in the health system, an improved relationship with patients and hospitals and the gains from involving the patient in the management of his/her health. To the hospital, the documents noted as benefits an increase in efficiency and competitiveness and an increase in loyalty of patients. To the National Health Service, the expected benefits pointed out were an increase of overall efficiency, better use of installed capacity in the NHS, ability to respect target waiting times, to value patients and increased transparency and information sharing.

These expected benefits resulting from official discourse of the Ministry of Health may, or may not, be reasonable ones, as it will be discussed below. Before moving to the analysis of the policy, it is worth mentioning an early view from one hospital on the merits and potential impact of the policy. According to an internal document of a major

[3] For further details on the Portuguese health system, see Simões et al. (2017b).

NHS hospital, several concerns were raised. Existing referrals, at that very early stage, seemed to be more linked to geographical proximity than to waiting times. The relative importance of the two elements may be different from the one required for the policy to have any visible effect. This is, of course, an aspect prone to empirical measurement.

The second concern was about the internal organisation of the hospital. In particular, since that NHS hospital expected an inflow of patients under this patient choice initiative, the specific concern was the difficulty to increase their supply of services in hospital specialty consultations to patients without a reduction in other areas of activity of the hospital (say, programmed surgeries in the more affected specialties or support to the emergency room services). A third managerial concern was that the additional number of appointments would not carry out an increased stream of revenues. The hospital is funded by an annual global budget, which is insensitive to the number of patients that use the patient choice initiative (both inbound and outbound flows). The envisaged internal adjustment considered several elements: reducing time duration of appointments, reducing the turnover time between patients, reducing the number of follow-up appointments (in order to increase physicians' time for first-time consultations), propose a change in referrals by GPs to make mandatory to account for waiting times, and propose a financial reward for taking in patients, paid by the natural "catchment area" hospital.

Thus, according to this view, we should expect that hospitals change their "quality" of service, including effects in patients' waiting times, and/or request a different payment system. Thus, the hospital adjustment to the new policy needs to be adequately considered in the discussion of the policy.

The initial stage of the PCI was described by Simões et al. (2017a), with a focus on patient mobility. According to the information collected and reported in Simões et al. (2017b), about 10% of outpatient referrals from NHS primary care units were made to an NHS hospital outside the previous catchment area, in the period from June 2016 to May 2017. The authors also report an increase in median waiting time for first outpatient visits. According to the national report on access to care in the NHS, up to the end of 2020, 11.6% of patients opted to have a consultation in an hospital outside the traditional catchment area.[4]

3 The model

Our starting point is the model of Siciliani and Straume (2019), which builds on the original contribution of Brekke et al. (2011). We introduce of a crucial element of the Portuguese NHS 2016 patient choice initiative - demand shifts across hospitals do not carry payments (or revenues) with them. This changes the incentives faced by hospitals. Intuitively, if more demand does not bring more revenues, why should a provider care about quality to attract patients is an obvious question. But another question needs to be answered, why is not quality set at the lowest possible level if quality does not play a role in revenues' attraction. Thus, a model of analysis of the PCI has to include

[4] Ministério da Saúde, Relatório Anual do Acesso, p. 120.

a motive for valuation of quality by the hospital management, even in the absence of financial flows that are sensitive to quality by means of patient choice.

Taking waiting times as a relevant quality dimension in the service provided by a hospital, as waiting times often are present as key performance indicators in management assessment and reputation, this provides a motive for quality considerations to enter the objective function of hospital managers. In addition, hospitals are often seen as semi-altruistic providers, as they care about patient benefits, resulting from their case activities. These aspects are included in the model of analysis.

As to the decision process by the patient, the patient choice initiative introduces competition (choice across hospitals) by removing the restriction of strict geographic catchment areas. We model the choice of hospital as being determined by distance and by quality (measured by an indicator related to waiting times, in the empirical analysis, as described below). We assume that no conflict exists between patient and GP in the shared decision making process.[5] The type of competition introduced by the PCI is different from increasing competition in a context where there is already some level of competitive pressure across providers.[6]

We use a standard Hotelling model, following the tradition of previous literature, with two hospitals and patients sensitive to quality (as mentioned above, an observable element of quality of the service provided by a hospital is the waiting time, which is the dimension that will be addressed in the empirical analysis below).

We first consider the utility of hospitals' decision makers. Hospital managers derive utility from three distinct elements. First, the net revenue generated. Second, hospital managers value own patients' benefits. Third, hospital managers consider the costs of quality provision.

Formally, we have: (i) utility from revenues (financial situation of the hospital), given by: $U_{1i} = R_i - c_i x_i$, where R_i is hospital i's revenues, exogenously determined (global budget, in the case of the hospitals of the Portuguese NHS), x_i is the number of patients that choose to be treated in hospital i (in the context of the Hotelling model representation, it is the indifferent patient choosing from two possible hospitals, each located at the endpoints of a line segment of length 1 and patients uniformly distributed with density 1) and $c_i\,(> 0)$ is the (constant) marginal cost of treating a patient; (ii) utility from patients' benefits: $U_{2i} = \alpha q_i x_i$, in which higher quality q means a higher perceived (non-monetary) benefit to the hospital managers for any given number of patients treated, valued by α per quality unit. We can also interpret $\alpha\,(> 0)$ as a reputation effect to the hospital management team. For example, high waiting times/low quality implies a lower managerial reputation (or lower performance assessment) to the hospital management team; and, (iii) costs of providing quality, $\beta_i q_i^2/2$, with increasing marginal costs of quality.[7] Utility of hospital's decision makers is normalised in monetary units without loss of generality.

[5] Exploring the nature of the patient - GP relationship and how shared decision making translates into a final decision is beyond the scope of this paper. It is left for future research.

[6] See Brekke et al. (2011) on this aspect.

[7] Note that having higher costs of treating patients with higher quality, cqx, can be seen as included in U_2, under a simple reinterpretation of α as the degree of altruism/reputation minus the marginal costs of treating patients with quality α, while retaining $\alpha > 0$.

The objective function of each hospital i management team is

$$V_i = R_i - c_i x_i + \alpha x_i q_i - \frac{1}{2}\beta_i q_i^2, \qquad i = 1, 2 \tag{1}$$

where α includes altruism net of costs of providing quality q to treated patients.

Patients have preferences over the two hospitals following an horizontal differentiation framework. Quality differences guide relative demand. This allows for an interpretation of geographic proximity to a hospital, but also for other aspects that matter for patients' choices. Patients are fully insured by the NHS, and face no monetary prices.

Patients can have preferences for hospitals by other factors than waiting time and distance (geographical proximity). These preferences can be accommodated in the model in two different ways (at least). First, if all patients have a higher preference for one hospital, that amounts to include a patient valuation difference in the expression that defins the indifferent patient (between going to hospital 1 or hospital 2), which would lead to a higher demand for that hospital for every difference in qualities. The adjustment in quality after the policy is adopted would be essentially like the one described in the model (although with a slightly modified analytical expression). The second way to model this preference is to take the location of patients as also reflecting the preference for one or the other hospital. The location of the patient at a certain point of line describing the "distance" between the two hospitals can be interpreted as including that element and not just physical distance, with "transport costs" being inclusive, in this interpretation, of the utility cost of not choosing the most preferred hospital. In this case, the results of the model hold as reported below. The main difference between the two ways of addressing the possibility of other motives for preference of patients across hospitals is that the first option is more in line with the notion of "vertical differentiation" (everyone agrees that one hospital is better) while the second one is in line with the notion of "horizontal differentiation" (patients differ in their views about which is the preferred hospital).

For simplicity, we consider the simplest setting of patients having a valuation $bq_i - tx_i$ when going to hospital i, where b is the benefit associated with receiving quality q_i of care, t is "transport cost" of covering one unit of "distance" and x_i is how much they have "to travel" to hospital i.

3.1 Before the patient choice initiative

In the absence of choice by the patients, we have, for each hospital i, $x_i = \bar{x}_i$, $i = 1, 2$, with $\bar{x}_1 + \bar{x}_2 = 1$,[8] with $\bar{x}_i$ fixed administratively by the NS as the relevant catchment area.

Initial asymmetric catchment areas can be present, which will induce asymmetric quality choices before the introduction of the patient choice initiative. The implications of demand asymmetries may include asymmetric responses, even with otherwise symmetric hospitals as discussed in Sá et al. (2019). Sá et al. (2019) present

[8] To keep the line of length one, the sum of demands must be one. Nothing essential depends on this assumption.

a model of hospital competition and waiting times in which the initial waiting time differs across hospitals, which raises the question of why does it differ. In Sá et al. (2019) the differences are justified by assuming different demand levels and equally efficient hospitals. Endogeneising hospital efficiency by realisation of management effort means that different demand levels do lead to distinct effort levels and therefore different production efficiency levels (including in this notion both the level of production - patients treated - and quality levels). Our paper also adds another source of asymmetry: the managerial quality of hospitals can differ, leading to further effects.

Another source of asymmetry would be distinct altruistic valuations across hospitals (distinct intrinsic value of treating patients relative to the financial results of the hospital). As the policy change of introduction of choice by patients (and their GPs) takes place only within the public National Health Service, and the sample of hospitals in the empirical illustration includes only NHS hospitals, the working assumption is that the role of this potential asymmetry is negligible.

The optimal choice of quality by hospital i is characterised by the following first-order condition, resulting from maximising V_i, defined by Eq. (1), subject to $\bar{x}_i = 1/2$ (we assume symmetric catchment areas to focus attention in the other sources of asymmetry).

$$\frac{\partial V_i}{\partial q_i} = 0 \tag{2}$$

from which, under the assumed functional form,

$$q_i^* = \frac{\alpha \bar{x}_i}{\beta} = \frac{\alpha}{2\beta}\bigg|_{\bar{x}_i = \frac{1}{2}} \tag{3}$$

Note that equilibrium quality has values above its minimum level (zero) due to the role of altruism ($\alpha > 0$).

Under administratively fixed catchment areas, quality choices by each hospital are driven by two forces. First, hospital valuations of health outcomes to patients (higher quality leading to better outcomes) provides an incentive to increase quality. Second, increasing quality has a cost to the hospital, thus setting a downward force. The first effect is captured by α while the second is captured by β_i. Finally, having more patients, higher $\bar{x}_i$, makes relatively more important the first effect and, as such, higher quality should be selected for larger catchment areas, all other elements constant.

3.2 Under the patient choice initiative

The next step in the analysis is to define the demand faced by each hospital under patients' freedom of choice, which is determined by the indifferent patient between hospital 1 or choosing hospital 2. Under the patient choice initiative, the indifferent patient location is given by:[9]

[9] A possible interpretation, relating to waiting times, is to take $q_i = q_0 - w_i$, where w_i is waiting time at provider i, in which case q_0 is the highest level of quality achievable, under zero waiting time.

$$x = \frac{1}{2} + \frac{1}{2t}\left(q_i - q_j\right), \qquad i, j = 1, 2; i \neq j \tag{4}$$

The optimal choice of quality is given by the solution to the following first-order condition, for each provider i,[10]

$$\left(-c_i + \alpha q_i\right)\frac{\partial x_i}{\partial q_i} + \alpha x_i - \beta q_i = 0 \tag{5}$$

The first term reflects the impact of the policy in the incentives for quality, while the second and third terms are the direct effects of quality on the objective function of the management team (and were the ones already present under fixed catchment areas).

If $(-c_i + \alpha q_i) < 0$ then, because $\partial x_i / \partial q_i > 0$, the marginal valuation of increasing quality is smaller than before the policy. A quality increase attracts now more patients, which bring a financial loss to providers of care. The cost of treatment c_i can be larger than the valuation, to the hospital, of the benefits of treating a patient with quality q_i. Note that under the policy rules we will address empirically below, there is no extra payment associated with treating more patients. If that was the case, and the price paid for treating one additional patient covered treatment costs, then the incentive could easily operate in the opposite direction. In such case $p_i - c_i + \alpha q_i > 0$, where p_i would be the price paid per patient treated, and $p_i - c_i > 0$ would mean that attracting patients is financially profitable to the hospital.

Under the assumption of symmetric providers ($c_i = c$, $\beta_i = \beta$, $i = 1, 2$), it results the following expression for optimal quality:

$$q^s = \frac{t\alpha - c}{2t\beta - \alpha} \tag{6}$$

It follows that

$$\lim_{t \to \infty} q^s = q^* = \frac{\alpha}{2\beta} \tag{7}$$

recovering the regulated catchment areas result (with symmetric providers).

Given the second-order condition for a maximum in each provider's problem, for $q^s > 0$, it needs to hold $\alpha t - c > 0$. Otherwise, $q^s = 0$. That is, if the costs of treatment c are too large, the new policy will lead to the lowest quality level, as providers want to avoid the cost of treating patients without any additional revenue. We assume this is not the case. We add the requirement of a locally stable Nash equilibrium, which amounts to impose:[11]

$$\frac{\partial^2 V_1}{\partial q_1^2} \times \frac{\partial^2 V_2}{\partial q_2^2} - \frac{\partial^2 V_1}{\partial q_1 \partial q_2}\frac{\partial^2 V_2}{\partial q_1 \partial q_2} > 0. \tag{8}$$

[10] The second-order condition for a maximum is given by $\alpha < \beta_i t, i = 1, 2$.

[11] See Dragone and Lambertini (2020) and Hoernig (2003) for details.The relevant expression for the case of asymmetric providers will be made explicit below.

Equilibrium qualities under freedom of choice can be smaller or larger than under mandatory catchment areas. For low (high) costs of treatment, c, equilibrium qualities will increase (decrease). In the model parameters, $\bar{c} = \alpha^2/2\beta$ is the critical threshold for this characterisation. For c above this threshold, equilibrium qualities of (symmetric) providers will decrease after freedom of choice is introduced. Thus, the model does not support unambiguously that patient choice promotes higher quality by providers. This result is in line, without surprise, with the previous literature. We now move to the richer context of hospital/provider asymmetries. To obtain the implications and predictions of the model associated with the policy of freedom of choice, we proceed by steps as way to highlight the key intuitions. The total effect can be stated as the equilibrium quality difference between the two polar situations: legally defined catchment areas, on the one hand, and free choice of hospitals, on the other hand. This difference is given by

$$\Delta q_i = q_i^a - q_i^*, \qquad i = 1, 2 \tag{9}$$

Before reporting the expressions for Δq_i under asymmetries in cost structures and managerial costs, it is important to establish a set of remarks, to help understand the several effects present. The first remark is that under binding and pre-defined catchment areas, increases in demand induce more quality, due to the semi-altruistic valuation of patients treated and the quality of treatment they receive. Thus, even without any strategic interaction, changes in demand lead to changes in quality choices by hospitals. The second remark is that with symmetric hospitals (same costs of treating patients and same managerial costs of quality), the equilibrium quality choices under freedom of choice by patients result in lower quality than under the fixed (and equal) catchment areas if costs of treament are large relative to the altruism effect. Under symmetry of hospitals, the split of patients after the freedom of choice policy is 50% of total demand for each hospital, which is equal to the symmetric split under fixed catchment areas (by definition). Thus, if the catchment areas were not initially symmetric, the freedom of choice policy will lead to distinct adjustments in quality as response to the policy measure if hospitals are similar in all cost elements. The lower quality under the new policy emerges as a direct consequence from patients' choices and treatment costs. Lower quality drives away patients, which lowers total production costs (it also lowers total valuation from treating patients, which is of smaller magnitude relative to the treatment costs impact, by assumption). It is straightforward to check that having revenues sensitive to patients' movements leads to increases in quality if that revenue effect is large enough. Thus, movements in quality that differ across hospitals must result from asymmetries in costs (either treatment costs or managerial costs or both) and can be easily illustrated by comparison of the asymmetric equilibrium (under several parameter choices) and the symmetric equilibrium.

The total effect of the new policy in terms of in quality choices is the sum of several different effects. There is mechanic effect resulting from the immediate impact of allowing choice of hospital, without any change in quality decisions, as long as pre-policy qualities in each hospital differ. The second effect is the change of decisions regarding quality by each hospital, holding constant the quality of the other hospital.

The change demand arriving at the hospital will drive a different decision, even if there is no incentive resulting from strategic interaction. This can be easily seen from the impact of a demand change under the fixed catchment area policy. The third effect is associated with the strategic interaction between hospitals, resulting from the freedom of choice. By changing its quality, each hospital recognises the impact on the flow of patients.

3.3 The asymmetric case

The previous section highlighted elements of hospital interaction in the choice of quality when patients can choose which hospital to go, "money does not follow the patient" and hospitals are symmetric. The assumption of symmetry, however, does not reveal the richer set of possibilities in how hospitals may interact in reaction to the patient choice initiative. This requires an explicit treatment of the implications of hospitals' asymmetries to the way they will adjust qualities after the new policy is in place. On the role of GPs choosing hospitals for their patients, Xavier (2003) considers a quadratic cost function for a hospital providing a certain waiting time. Then proceeds to analyse demand for care by a GP on behalf of patients. GPs have to manage a budget. Our focus on the policy change and hospital adjustment, and not on the budget constrained GP decision choice, makes our analysis distinct from Xavier (2003).

We consider two basic elements of asymmetry. The first one is the cost of treatment, c_i. The second source of asymmetry is the managerial cost of quality β_i. We normalize $\beta_1 = 1$ for hospital 1 and take $c_1 = c + \Delta, c_2 = c$. The parameter Δ is the cost disadvantage of hospital 1 (for negative values of Δ, it will have a cost advantage). The parameter $\beta_2 = \beta < 1$ is the managerial cost advantage of hospital 2. Parameter β is the element giving a quality difference, as quality is selected endogenously by the hospital, under the mandatory catchment areas rule. The quality of hospital 2 prior to the new policy will be higher than the quality of hospital 1 for $\beta < 1$.

The set of first-order conditions for maximisation of the objective function of each hospitals can be (re)written as:

$$-(c + \Delta) + \alpha q_1 + \alpha t + \alpha(q_1 - q_2) - 2tq_1 = 0 \quad \text{(Hosp. 1)} \tag{10}$$

$$-c + \alpha q_2 + \alpha t + \alpha(q_2 - q_1) - 2t\beta q_2 = 0 \quad \text{(Hosp. 2)} \tag{11}$$

It can be readily checked that second-order conditions for a maximum require that β is not too small. We will consider below only the range of parameters that ensure second-order conditions to hold. The condition for a locally stable Nash equilibrium is given by:

$$\beta > \frac{\alpha}{t}\frac{4t - 3\alpha}{4t - 4\alpha} \tag{12}$$

and it is assumed to hold. Given that second-order conditions for value maximisation by each hospital are assumed to hold, this condition amounts to $\beta > 3\alpha/(2t)$.

Solving for the optimum qualities under the freedom of choice rule yields:

$$q_1^a = \frac{(\alpha t - c)(2t\beta - 3\alpha) - 2(t\beta - \alpha)\Delta}{4(t-\alpha)(\beta t - \alpha) - \alpha^2} \tag{13}$$

$$q_2^a = \frac{(2t - 3\alpha)(t\alpha - c) + \alpha\Delta}{4(t-\alpha)(\beta t - \alpha) - \alpha^2} \tag{14}$$

A quick inspection of these expressions reveals that for $(\beta = 1, \Delta > 0)$ we have $q_1^a < q_1^s$ and $q_2^a > q_2^s$. For $(\beta = 1, \Delta = 0)$ one retrieves, obviously, the symmetric situation. Also, for $(\beta < 1, \Delta = 0)$ it results $q_1^a < q_2^a$.

A related question that can be easily addressed is whether, or not, hospitals exposed to greater competition respond more strongly to the policy change. This amounts to the impact of increasing t (an indirect measure of competitive pressure, as it is related to hospital substitutability at the eyes of patients). Intuitively, a lower t implies more movement of patients across hospitals in response to quality differences, which will drive down quality levels in the higher quality hospitals prior to the policy change. Thus, higher competitive pressure, in the absence of revenue effects associated with patients' mobility across hospitals, leads to increased costs, that are detrimental to quality choices by hospitals. On the other hand, receiving more patients at the hospital contributes to the altruistic component of the valuation function, which operates in the opposite direction of the former effect.

The impact of competitive pressure on the policy effect strength is, thus, ambiguous. This can be easily seen in the symmetric situation $(\beta = 1, \Delta = 0)$. Under the assumption of symmetry and $\beta = 1$, post-policy choices are given by expression (6) and straightforward derivation yields:

$$\frac{\partial q^s}{\partial t} = \frac{2c - \alpha^2}{(2t - \alpha)^2} \tag{15}$$

which is negative when α is large enough relative to the cost effect c (which captures the negative financial impact of treating more patients in the absence of positive revenue effects) and positive otherwise. In the presence of asymmetries, this ambiguity remains. Also note that, in the context of the model, under symmetry, there is an upper limit to the level of competitive pressure (lower limit to t) to ensure that qualities are not set the lowest possible value (technically, a corner solution).

We are now interested in what happens to qualities selected in equilibrium when the freedom of choice policy replaces mandatory catchment areas, and hospitals are not equal. That is, what can be said about $\Delta q_i = q_i^a - q_i^*, i = 1, 2$.

Recall the first-order condition of the maximization problem of hospital i given by:

$$(-c_i + \alpha q_i)\frac{\partial x_i}{\partial q_i} + \alpha x_i - \beta_i q_i = 0 \tag{16}$$

The first term on the left-hand side does not appear under the pre-freedom of choice policy period. Since $\partial x_i/\partial q_i > 0$, for $-c_i + \alpha q_i < 0$, the hospital chooses to lower its quality after the policy is introduced.

Evaluating at the pre-policy equilibrium values,

$$q_i^* = \alpha x_i/\beta_i$$

and under the condition $-c_i + \alpha^2 x_i/\beta_i < 0$, one obtains a lower equilibrium quality choice by hospital i, relative to the pre-policy decision. Thus, for low α and/or high c (or high β), the PCI will lead to lower equilibrium qualities.[12] The freedom of choice policy introduces the marginal gain (or loss) of receiving more patients at hospital i that is created by quality changes. That marginal return to quality, in the absence of additional payments, will be positive only if altruism (valuation of patient benefits) is large relative the marginal cost of treatment. Hospitals with higher treatment costs will more likely reduce quality in equilibrium, relative to pre-policy decisions. Hospitals with a lower β (better management in our interpretation) will have a higher pre-policy quality and also gain less, in terms of costs, from reducing quality. They are more likely to have a positive marginal return from inducing patients' movements and, therefore, more likely to have, in equilibrium, a quality increase after the policy being adopted. Depending on the magnitude of these effects in each hospital and the magnitude of asymmetries across hospitals, equilibrium quality changes associated with the policy can produce any empirical combination of increase/decrease of quality in each hospital.

Thus, depending on parameter values, several situations may occur. From the symmetry case, we know that the possibility that both qualities increase or both qualities decrease cannot be ruled out. We add now the situations where qualities of each provider move in opposite directions of each other.

The strategic interaction between asymmetric hospitals can lead to any evolution of equilibrium qualities. Treatment cost advantages and managerial talent advantages provide incentives for higher quality. Depending on the strength of these advantages for each hospital and of the strategic interaction (which provides an incentive for lower quality, as described earlier) any combination of quality changes may emerge. Table 1 reports numerical examples of all four possibilities[13], illustrating the main point that empirical analysis needs to explicitly account for asymmetries lying behind initial different waiting times (before patients' choice being introduced). These asymmetries lead to distinct responses to the new environment with patients' choice.

The important point to retain is that starting from an initial asymmetric position, hospitals with high and with low quality choices may end up moving in different directions. The empirical implication is the need to treat them separately in estimation.

[12] If revenues are sensitive to patients' movement across hospitals, $\partial R_i/\partial x_i > 0$ and for $\partial R_i/\partial x_i - c_i > 0$, qualities will increase in equilibrium.

[13] As discussed below, on the particular proxy of quality used in the present work, the fraction of people waiting more than a pre-defined target value, the cost interpretation we assume to be more adequate is the organisational effort to achieve lower waiting times, which in turn leads to a smaller proportion of patients with waiting times above the target.

Table 1 Numerical examples of all possible equilibrium configurations

	$\Delta q_1 > 0$ $\Delta q_2 > 0$	$\Delta q_1 > 0$ $\Delta q_2 < 0$	$\Delta q_1 < 0$ $\Delta q_2 > 0$	$\Delta q_1 < 0$ $\Delta q_2 < 0$
α	3	3	3	3
t	5	5	5	5
β	0.95	0.98	0.92	0.95
c	4.5	4.5	4.5	10
Δ	–0.7	–0.7	–0.7	0
q_1^*	1.50	1.50	1.50	1.50
q_1^a	1.54	1.78	1.14	0.5
q_2^*	1.58	1.53	1.63	1.58
q_2^a	1.68	1.35	2.21	1.00

It is straightforward to verify that all second-order conditions for profit maximization and local stability of Nash equilibrium are satisfied in the cases above (The effects hold for an open set of values in the neighborhood of the parameters reported in the table)

In the next sections we provide evidence on the effects produced by the patient choice initiative, using the conceptual framework defined above to guide the analysis.

4 Empirical methodology

4.1 The testable implications from the model

The general model was presented in terms of quality decisions. To move to the empirical application, a measure for quality is used, the proportion of patients in waiting list for a first appointment at hospital specialty consultation within the target waiting time. Of course, hospital quality has more dimensions than waiting time. The use of waiting time here, with no reference to other quality dimensions, is motivated by the policy change under discussion, as the official documents supporting the policy make explicit the role of waiting times in the set of information to be used in the joint decision process of patients and GPs. Other global measures of quality widely used, such as readmission rates or hospital-related mortality rates, do not seem adequate to address the impact of the policy. In addition, waiting time is a concern in several health systems and satisfy the notion of quality we want to address: it is certainly costly to reduce waiting times to hospitals. It is also natural to assume that patients value lower waiting times and that hospital managers may take into account in their decisions the benefit to patients from lower waiting times (thus, being consistent with our modelling of hospitals as semi-altruistic providers of care). In addition, under the policy change addressed here, information on waiting times is available to patients when choosing the hospital. These two other quality measures, moreover, are unlikely to the major drivers of decisions related to booking an appointment with a hospital specialist by a patient.

The use of waiting times as a measure of quality has also a more subtle point. Waiting times have a time dimension, an underlying dynamic nature, that is not made explicit in the model presented above. The link between a dynamic model of demand and supply of health services with waiting times acting as a balancing factor and modeling it as a static phenomenon was explored in the literature (Iversen 1997; Van Ackere and Smith 1999; Gravelle et al. 2003; Xavier 2003; Siciliani 2005, 2006, 2008, among others).

On the treatment of waiting time as "negative quality", the best way to interpret the analysis is to consider that hospitals can take managerial effort that improves the ability (time) to treat, thus setting indirectly waiting times when facing a stable structure of random demand. The costs associated with providing quality can be generally interpreted as managerial effort produced to achieve that quality level. Our product differentiation structure does not explicitly consider time or the moment of need. In terms of the dynamics of queues, higher managerial effort leads to lower server time for each treated patient (on average), which decreases average waiting time and increases the number of patients that wait less than a certain target time.

Associating lower waiting times with higher quality, the model predicts that after the policy change, initially asymmetric hospitals will react differently to the policy. Also from the model, the advantage/disadvantage leading to initially different waiting times cannot be attributed to treatment costs of patients alone. The differences have to be based on distinct managerial efficiencies and managerial efforts. From the theoretical model, depending on the particular starting point of each hospital, quality may increase or decrease after the policy change. The symmetric version of the model provides one component of the effects at play. Conditional on value of waiting times reduction being sufficiently important to hospitals, equilibrium qualities will increase under freedom of choice of provider. The "sufficiently important" term has a precise meaning: the valuation (altruism) that hospitals have of benefits to patients from reduced waiting times must be larger than the cost of providing care to more patients, which the hospital will attempt to attract through higher managerial effort to reducing waiting times. Otherwise, if this condition is not satisfied, quality will move in the opposite direction. High financial pressure over hospital management can result in more weight being given to financial results in decisions of hospital management. That corresponds to a lower value of α in the model. Hospitals, in this latter case, prefer that patients move to another location to be treated. In the absence of altruism and with no payment associated with treating more patients, either reputation effects are strong enough to have the same role as altruism to keep hospitals with a focus on reducing waiting times or the rational decision would be to allow quality deterioration (waiting times increase) as a way to save on costs of patients (that will move to another hospital).

A potential issue with the use of waiting times, more exactly, the proportion of patients waiting more than a pre-specified threshold set by the Government, is that high waiting times may be also a signal to high-quality care provided by the hospital (leading to higher demand and to higher waiting times). In the setting of the Portuguese NHS this is less of a concern. Prior to the policy being adopted, there was no choice, so higher waiting times were the result of demand (determined by patients' conditions in the catchment area) not being met by hospital operations (production falling behind demand) and could result from either demand differences or lower hospital efficiency (lower managerial effort, due to higher costs of exerting effort), even in the presence of equal clinical quality in treatment of patients. After the policy is adopted, the shared decision-making process implies that GPs are actively involved in discussion with the patient and clinical quality needs not to be inferred by patients through observation of waiting times as GPs will presumably have knowledge and experience to assess it (from interaction with the hospital, with colleagues and experience from other patients). This leaves waiting times to be a dimension of quality that is valued by patients and a potential proxy for hospital quality. Essentially, the signalling effect of high waiting times is not relevant prior to the policy due to the legal catchment areas approach and it is not relevant after the policy as shared decision making will involve the GP, an informed agent.

Benefits from competition among hospitals based on analogies to other markets of products and services cannot be taken for granted. Introducing asymmetries brings in further effects, associated with the reasons for that asymmetry. Hospitals that are initially more efficient in reducing waiting times will now face an additional demand, even if they do not change their decisions. This flow will either be desirable (if the value of treating more patients exceeds the extra cost of waiting times) or not. Therefore, the new equilibrium choices may have lower or higher waiting times, leaving room for empirical analysis to step in and inform about it.

The implications from the model can be summarised in the following way: (a) the changes in waiting times (quality) are likely to be different in low waiting time and high waiting time hospitals (prior to the introduction of the patient choice initiative); (b) the changes in quantity - number of appointments - will reflect the movement of patients to lower waiting time units, and will be guided by (lagged) information on performance; and, (c) increased geographic dispersion of patients going to low-waiting time hospitals is expected, with no discernible effect on the other hospitals (they lose patients but not necessarily with a clear geographic pattern inside the initial catchment area).

A non-significant effect in statistical terms can be due either to absence of effect (in patients' decisions) or to opposing effects at the hospital level, with strategic decisions upon quality compensating mechanic effects of the specific indicator used here (with different underlying incentives for top and non-top performance hospitals). Below, we explore implications (a) and (b), based on publicly available data.

4.2 The data: qualities and quantities with asymmetric providers

The data used comes from publicly available data on Portuguese hospitals, retrieved in the NHS portal. It provides data per hospital on specialist' appointments in hospitals and percentage of those appointments that is done within the regulated waiting time target (legally defined maximum response time).[14] This data set is used to address both changes in quality and changes in quantity (number of patients).

[14] Ideally, the analysis would be performed with data per patient, but such data is not available.

The empirical measurement available is not the average (or median) waiting time. It is the proportion of patients waiting longer than a pre-specified length of time. Thus, gains in median waiting times may not be captured if they take place mainly at the lower end of the distribution of waiting times (those that wait less benefit more from the decisions of the hospital management). Similarly, worsening median waiting times can be compatible with a decrease in the proportion of patients waiting more than a pre-specified threshold. Thus, the information on the proportion of patients waiting longer than a certain number of days (or months) is not a perfect informant on the more standard measure of median waiting time. On the other hand, this measure is less dependent on short-term fluctuations in demand, which may leave unaffected the long waiting times cases that are captured by this variable. Thus, the empirical analysis will be carried out under the assumption that hospital's management decisions on "quality" of the services they provide will be reflected in a reduction of the proportion of people waiting for surgery more than a target length of time (specified exogenously to all hospitals).

The first step of the empirical analysis is the definition of performance groups based on data prior to the patient choice initiative. The second step will be a regression analysis (a before/after comparison). A difference-in-differences analysis is not possible as there is no control group. All hospitals of the National Health Service were subject, at the same time, to the patient choice initiative. The data on waiting times and patients' appointments runs from January 2013 to December 2019.[15] An empirical strategy of continuous difference-in-differences used in previous works is due to Cooper et al. (2011). The analysis introduced in Cooper et al. (2011) is done at the patient level, with a difference-in-differences structure that explores different (presumed) competitive pressure levels across regional areas. The nature of the policy change and the level of aggregation of available data does not allow for a meaningful analysis of competitive pressure using that approach in our context. One could use geographical proximity to define more narrow, delimited, geographic markets, although the pooling of all hospital specialties waiting times into a single aggregate measure tenders such exercise difficult to interpret. That is, the 'continuous difference-in-differences" approach relies on the definition of regional markets that are subject to different levels of competition pressure. The definition of such regional markets for programmed activities (such as the booking of a specialist appointment at the hospital) is more difficult to establish than for acute needs of patients. In principle, under the patient choice initiative in the Portuguese NHS, any hospital offering the necessary type of consultations can be selected, with no geographic restrictions. The aggregated data available does not allow to separate waiting time by procedure (or even specialty). As some hospital specialty appointments will only be available at a small number of hospitals, the set of choice of hospitals offering such consultations is national. In a more precise way, in our case, the aggregate nature of the dependent variable means that it includes many different conditions in the same indicator. As one example, advanced neurosurgery interventions would take a national hospital market, with few hospitals providing it. It will also

[15] The year 2020 was excluded from the analysis as the COVID-19 pandemic hit Portugal in March 2020.

have a small number of cases. At the opposite end of the spectrum, we can highlight cataract removal, for which "local competition" of nearby hospitals may be seen as closer to the idea of "continuous effects" differing across local markets. These will have a high volume of cases. In the middle, we may consider cardio-thoracic surgeries, for which regional hospital markets can be a good approximation. Oncology is another area in which the "geographic market" can be either regional or national but certainly not local. In addition, in some places, only one hospital exists for a low travel time of patients but it may not carry all types of medical and surgical interventions. For a patient located in the interior of the country, the local hospital will have a strong location advantage relative to other hospitals in the same region and in the country, although for the procedures not carried out in that hospital, "competition" can be national. In the main cities, within roughly the same distance, patients may opt for different hospitals of similar profile. A final note related to geographic market scope. It is defined, in the data used, by Continental Portugal. The islands of Azores and of Madeira have their own health systems and their data is not included in the available NHS data.

To define the "top" and "non-top" groups, the following procedure was followed. The "top" units were defined as those having, prior to the patient choice initiative, a percentage of patients within target waiting times above the median value (over the full period) in all months in the sample (achieving a balance of units vs. being a "good performer"). Variations in the definition of top units according to criteria associated with waiting times performance prior to the patient choice initiative produced essentially the same qualitative results.[16]

As explanatory variables we use a monthly time trend variable, a dummy variable for the period of the patient choice initiative (after June 2016), the time trend is also allowed to have a shift at this date, the percentage of patients within the target waiting time for surgery (1 month lag) and the percentage of patients within target time for appointments (1 month lag). Finally, fixed effects at the hospital level and at the month level (to capture seasonality) were included.

A technical point in the design of the empirical analysis is the appropriate bandwidth (of months) to be included. Given the large cross-section dimension (number of hospitals) relative to the time dimension, the option was to use all the available information in the time dimension, which also ensures a balance between time periods prior and after the policy was adopted. Also relevant is the time unit to be used, with the available choices ranging from monthly to annual data. One may reasonable argue that "quality" responses by hospital management to the policy will take time to unfold and will not be adequately captured by monthly variations. Using annual data (or semester data) leads to an implicit higher weight placed in the cross-section dimension (differences across hospitals) rather than reaction to the policy. Quarterly data is another possibility that one may consider. Taking all into consideration, including the role of a time trend variable, monthly data makes use of all available information.

At the time of the PCI, there were no other public policies aimed at waiting times (or hospital quality, in general). A different issue is about lead and lag times associated

[16] Details available upon request.

with the policy. In case hospitals anticipate the policy, effects could start earlier than the formal introduction of the policy. On another direction, if hospitals are mainly reactive to the enactment of the policy, then a lag in changes if quality may occur. Given the uncertainties surrounding the political process of policy approval in Portugal, only after official publication of policy measures the exact details become known. It is more natural to have lags in reaction than leads. This will be taken into account in robustness checks on the results.

To relate the testable implications from the model with the available data, several assumptions are made. In particular, we assume that a higher percentage of patients within waiting-time target has a positive association with lower waiting times in the hospital, and that hospital admissions pattern reflects specialists' appointments geographic pattern (to some extent, as not all hospitals treat all conditions). Only elective, programmed, activity is considered, with emergency department cases being removed.

The estimated equations have the following form:

$$y_{it} = \theta_0 + \theta_1 T_t + \theta_2 Dp_{it} \times T_t + \theta_3 Dp_{it} + \sum_i \gamma_i X_{it} + s_m + \lambda_i + \varepsilon_{it} \tag{17}$$

where T denotes a trend variable, Dp represents the period after introduction of the policy (taking value 1 after and value 0 before the policy), X_{it} are other explanatory elements (information-related variables in the regression with the dependent variable "number of first consultations", the change in target waiting time in the case of dependent variable "proportion of patients within the target waiting time"). The term λ_i captures hospital fixed effects and s_m are a set of monthly seasonal dummies. Finally, ε_{it} denotes a stochastic error. The relevant dependent variable, y_{it}, is defined, in one set of estimates, as the proportion of patients waiting for a first appointment within the target waiting time, and in the other set of estimates, as the total number of first appointments. Subscripit t indexes months. A separate equation, for each dependent variable of interest, is estimated for "top" and "non-top" performance hospitals.

Table 3 in Appendix describes the main variables used in the analysis. Tables 4 and 5 in Appendix reports basic descriptive statistics on the main variables.

4.3 Clustering standard errors

An essential decision for statistical inference is how to compute the correct variance-covariance matrix of estimators. The concern with a common unobserved shock at the group level, creating correlation between all observations of the group, leads to the suggestion of computing clustered standard errors for statistical inference (Bertrand et al. 2004; Hansen 2007). Abadie et al. (2020) address the issue of clustering standard errors when the entire population is observed, which is the case here. According to their analysis, if sampling is completely random, then adjusting for clustering in standard errors is not appropriate. Abadie et al. (2021) suggest that the clustering adjustment can be unnecessarily conservative under some conditions.

The residual is likely to be correlated by hospital, leading to the suggestion of clustering by hospital. But common shocks per period may also be relevant. In this case, clustering by month-year would be suggestion. Whether one should cluster, or not, should not be decided by the observation that changes something in the results (Abadie et al. 2020). Cameron and Miller (2015) suggest that clustering should take place whenever there is the expectation of correlation within groups. They also suggest to compute cluster robust standard errors and compare to default standard errors. Clustered standard errors should be used when there is an important difference. Wrongly assuming that the clusters are a sample of more clusters in the main population may lead to incorrect clustering. In our data, the full population and clusters (hospitals) are used. In addition, all hospitals were assigned to treatment (subject to the new policy). Thus, the current literature suggests that either clustering at hospital level or at month level may be appropriate (there are hospital unobservable factors, there are time unobservable factors common to all hospitals in the sample, as they all belong to the Portuguese NHS) (MacKinnon et al. 2022). For completeness, both approaches to clustering standard errors are implemented and reported.

5 The effects of patient choice

Figure 1 reports the average value, over hospitals, per month of observation of the percentage of patients within the target waiting time for first specialty appointment at the hospital, divided by group, top and non-top hospitals. The vertical dashed line points to the starting month of the patient choice initiative. From visual inspection,

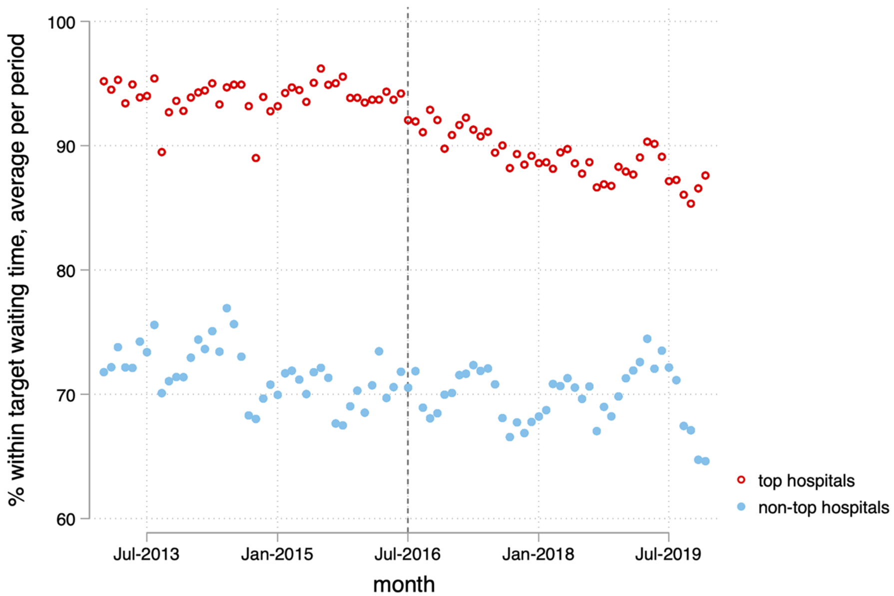

Fig. 1 Proportion of patients within target waiting time (average per month, by group of hospitals)

it results visible a decrease over time of the proportion of patients having a first specialty consultation within the target waiting time in the top performance hospitals after entering in place of the new freedom of choice policy. No clear evolution or change takes place in non-top performance hospitals. Thus, lower quality, meaning higher waiting times on average, seems to have resulted from the PCI.

Figure 2 describes the average per month of the total number of first appointments, divided according to top performance and non-top performance hospitals. Top performance hospitals are, on average, smaller in size (receive smaller numbers of patients). There seems to be no systematic change associated with the PCI.

Regression analysis, reported in Table 2, provides a quantitative view of the changes produced by the new policy, using as unit of observation each hospital in each month over the years 2013-2019.

According to the results in Table 2, the changes in the trend of the variable "percent of patients within the target waiting time", with non-clustered standard errors, indicate asymmetric responses. The hospitals with a better starting point (higher number of patients within target time) see a negative trend after the policy change. The introduction of freedom of choice by patients resulted in an increase of waiting times (lower quality), as measured by the proportion of patients within the target waiting time. On the other hand, the opposite holds for the hospitals that were not in the top performance group: these hospitals on average increased the proportion of patients within the target waiting time, meaning that waiting times have been reduced (better quality after the policy was introduced).

Regarding the intensity of patient flows and the role of information decision variables (relevant for the shared decision making introduced by the freedom of choice

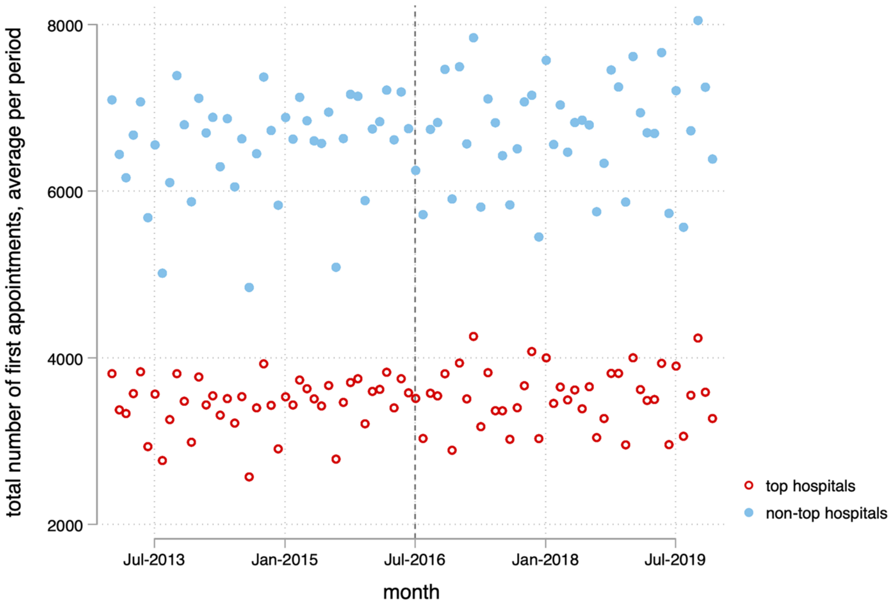

Fig. 2 Number of first appointments (average per month, by group of hospitals

Table 2 Changes associated with the patient choice initiative

	% within target waiting time		number of first appointments	
	top	non-top	top	non-top
Trend	0.006	−0.081	3.38	6.81
	(0.25)	(−3.20)***	(0.59)	(1.96)**
	[0.28]	[−0.99]	[0.55]	[1.85]
	{0.46}	{−5.08}***	{0.86}	{1.18}
Policy	1.766	−4.370	−1934.00	1200.68
	(0.85)	(−2.20)***	(−1.37)	(2.68)***
	[0.42]	[−0.71]	[−2.17]	[1.54]
	{2.09}**	{−2.56}**	{−1.78}**	{2.73}***
Trend × Policy	−0.089	0.117	0.238	−0.955
	(−1.86)	(2.58)**	(0.04)	(−0.26)
	[−1.33]	[0.72]	[0.03]	[−0.22]
	{−4.32}***	{3.08}***	{0.05}	{−0.13}
Surgery WT_{t-1}			2.397	15.12
			(0.26)	(3.17)***
			[0.53]	[1.79]*
			{0.34}	{3.56}***
First appointments$_{t-1}$			−1.92	−2.45
			(−0.15)	(−1.15)
			[−0.19]	[−0.84]
			{−0.20}	{−1.93}*
Surgery WT_{t-1} × Policy			12.641	−13.61
			(1.97)	(−3.11)
			[3.05]***	[−1.89]*
			{2.54}**	{−3.98}***
First appointments$_{t-1}$ × Policy			9.15	−0,35
			(0.76)	(−0.13)
			[1.24]	[−0.08]
			{1.03}	{−0.21}
Δ definition of waiting time	−1.593	−1.646		
	(−1.45)	(−1.56)		
	[−0.86]	[−0.48]		
	{−4.13}***	{−2.60}**		
Observ.	1003	3104	567	2331
Fixed effects	yes	yes	yes	yes
Nr Hospitals	12	37	9	37

t-statistics based on standard errors clustered at the hospital level in square brackets, based on clustered standard errors at the period level in curly brackets

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

policy), we see that differences across the two groups are substantial. The information variables do not seem to guide demand of top performance hospitals but for surgery indications within target waiting times in the previous period. In contrast, information on past performance is relevant for the demand directed to the non-top performance hospitals. However, the change associated with the new policy of freedom of choice is in the direction of becoming less, not more, sensitive to information regarding the waiting times for surgery in the hospital (actually, testing for a zero role of information after the introduction of the new policy does not reject the null hypothesis).[17]

The more relevant element of information is, according to the results, the waiting for surgery, with patients choosing to go to hospitals with lower waiting times (higher proportion of patients within the target waiting time for surgery). Since the movement of patients is the booking of first appointments for specialty consultations, it suggests that concerns with eventual future surgery needs are present in patients' decisions.

From the several groups of results, two features are worth highlighting even though statistical significance is not very high. The impact of the policy on the top performing hospitals is in line with an increase in waiting times (lower quality), as proxied by the proportion of patients that are in the waiting list below the target waiting time. The opposite effect seems to be present in the non-top performing hospitals, with an increase in the proportion of patients within the target time.

Taking as quality measure the proportion of patients waiting longer than a target waiting time, for the same managerial effort, patients' reaction to the PCI would lead to an automatic change in the indicator used, even if quality set by the hospital is unchanged. This mechanic effect cannot be independently observed in the data and separated from the impact of hospitals' management decisions. If the hospital management has a pre-policy target for waiting times (quality), then the PCI would lead to adjustment in the hospital to keep the target quality level. The mechanical effect on waiting times resulting from flows of demand would be accommodated to keep quality constant. However, the quality objective of a hospital does change with intensity of demand, even if the hospital does not internalize the strategic interaction with other hospitals (resulting from quality choices driving patients' decisions). Ignoring that hospitals' qualities influence the demand faced, the observation of an increase in demand should lead to a quality (decrease in waiting time). Finally, when hospitals recognize their interdependence, the ability to influence flows of patients with quality choices (waiting times) creates an additional incentive that, under sufficiently high costs of treating patients and/or sufficiently low altruism) leads hospitals to decrease quality (increase waiting times) as an indirect way to avoid increases in costs of operation.

Overall, there is no clear systematic regularity in terms of statistical significance. Most coefficients do not pass the standard t-test of individual statistical significance of being different from zero. For the few cases where the results suggest different

[17] The test corresponds to the null hypothesis of the sum of the coefficient of the variable plus the coefficient of the same variable interacted with the past policy period being equal to zero.

effects prevailing at the top and at the bottom (performance) of hospitals in accordance to the predictions of the model. Moreover, the effect for the best performing hospitals (measured before introduction of the new policy introduced in 2016) points to a negative effect. Those hospitals have fewer proportion of patients within the target waiting time for appointments, our measure of quality, after the policy. Hospitals that were not classified in the top performing group (prior to the new policy) appear to have a larger proportion of patients within the target waiting time (lower waiting times).

In the future, the absence of payments associated with patient flows creates a burden on the financial position of hospitals that attract more patients, while alleviating the financial pressure on the others. At some point, by the way hospital global budgets are set, financial flows will be affected with a two-years lag, or more. The anticipation of revenues two or more years into the future, together with many other elements of hospital activity that will change and that also shape the global budget awarded (yearly) to the hospital is unlikely to produce a financial incentive that induces an interest in attracting patients by offering higher quality to obtain a higher budget. Thus, over time, the global budget may accommodate some of the financial impact without creating an explicit incentive (operating more as a future lump-sum change rather than a direct consequence of increased patient flows).

6 Robustness checks

Several robustness checks were performed. The monthly time trend in Tables 6 and 7 was assumed to have the same linear functional form prior and after the policy measure entered place. To test for the robustness of results to this assumption, a more flexible quadratic functional form for the trend prior and after the policy adoption. Table 6 in Appendix presents the results. This more flexible structure has the same features of the linear trend version, pointing to an increase in waiting times of top performance hospitals and a decrease in waiting times of non-top performance hospitals.

Using a policy dummy assumes that policy is in full effect immediately. But patients and GPs may take some time to learn about the policy and how to use it. And after patients start to eventually use the policy, hospitals may try to act upon it. This suggests that the impact of the policy will not take place at a moment in time, and it eventually unfolds only after some time. Thus, another robustness check involved the exclusion from the regression analysis of the 6 months after enactment of the policy, to account for an adjustment period. Identification of treatment effects from regression discontinuity donuts is subject to discussion (Dowd 2021), thus we retain as main results the ones without data omission.

The essential features of the results (including the lack of precision in estimating effects) are common to all variants.[18]

[18] Details available upon request.

7 Final remarks

Health care markets are widely recognised to be subject to specific structural conditions, including price regulation in the payment to providers of care, definition of mandatory catchment areas (in some healths, in particular those where a public National Health Service is present, etc.). We explored here a policy change that introduced more freedom of choice in a hospital system with regulated prices (and previously mandatory catchment areas). The main concern addressed was the choice of hospitals, and the likely effects of the new policy into quality decisions of hospitals.

Policies that introduce freedom of choice to patients are often regarded as pro-competitive. In a context of regulated prices, the impact on quality of care provided will be influenced by several elements. We highlight here the role of initial asymmetries across providers, in their quality levels, and how such asymmetries will influence their response to the policy change. A stylised model of the market, under symmetry of providers of care (hospitals), shows that, under conditions that are reasonable, an increase in quality should be anticipated as optimal response. However, allowing for asymmetries across providers pre-policy change leads to a more diverse picture. In particular, allowing for two basic sources of asymmetry – different costs of treating patients and different costs of improving quality – leads to the possibility that each provider of care may increase or decrease quality in response to a policy change introducing freedom of choice to patients.

An empirical testing of the model with Portuguese hospital data is broadly in line with predictions from the model on asymmetric responses in reaction to the introduction of patient choice. The overall statistical significance of results is weak. It suggests, still, that reactions to the policy change are richer and more uncertain than the simple expectation that more freedom of choice triggers quality competition and quality improvements uniformly across providers of care.

In terms of demand directed to each hospital, there is no clear picture. Patients that choose to go to the non-top performing hospitals seem to react to information variables, with hospitals with a lower proportion of patients within the target within time experiencing a drop in the demand, although the effect becomes smaller after the policy. This demand effect is consistent with the change in trend observed in this group of hospitals, after the policy adoption with an improvement in the proportion of patients within target waiting time. These broad observations emerge, with variable levels of statistical significance, under the several variants explored here.

Appendix

Table 3 Definition of variables

Variable	Definition
Trend	Takes value 1 in January 2013. Increments 1 each month.
Policy	Takes value 1 after June 2016. Takes value 0 otherwise.
Surgery WT_{t-1}	Percent of patients waiting for surgery within the target waiting time, in the previous month
First $appointments_{t-1}$	Percent of patients with first appointments within the target waiting time, in the previous month
Δ definition of waiting time	Takes value 1 after May 2017. Takes value 0 otherwise. Captures change in the target waiting time definition.

Table 4 Descriptive statistics I

	Obs	Mean	Std. Dev.	Min	Max
Proportion of patients within waiting target time					
Top performance hospitals					
Total	1003	91.57	10.2	40.3	100
Policy = 0	487	93,99	6,32	49,61	100
Policy = 1	516	89.29	12.49	40.34	100
Non-top performance hospitals					
Total	3104	70.76	14.16	0	99.89
Policy = 0	1513	71.67	14.71	0	99.89
Policy = 1	1591	69.89	13.57	6.91	99.44
Number of first appointments					
Top performance hospitals					
Total	567	4623.8	4382,76	762	19243
Policy = 0	180	4640.8	4505.5	776	18264
Policy = 1	387	4615.9	4330.4	702	19243
Non-top performance hospitals					
Total	2331	6727.3	4388.8	0	23558
Policy = 0	740	6726.5	4284.6	0	21601
Policy = 1	1591	6727.6	4437.6	487	23558

Table 5 Descriptive statistics II

Proportion of patients in list for surgery					
previous month, within target waiting time					
Top performance hospitals					
	Obs	Mean	Std. Dev.	Min	Max
Total	567	85.93	12.00	48.90	100
Policy = 0	180	87.93	9.10	71.8	100
Policy = 1	387	84.99	13.05	48.9	100
Non-top performance hospitals					
	Obs	Mean	Std. Dev.	Min	Max
Total	2331	87.44	8.72	3.3	100
Policy = 0	740	89.50	7.51	67.6	100
Policy = 1	1591	86.48	9.08	3.3	100
Proportion of patients in list for appointment					
previous month, within target waiting time					
Top performance hospitals					
	Obs	Mean	Std. Dev.	Min	Max
Total	567	89.55	10.46	49.2	100
Policy = 0	180	93.28	4.38	81.6	99.9
Policy = 1	387	87.57	11.80	49.2	100
Non-top performance hospitals					
	Obs	Mean	Std. Dev.	Min	Max
Total	2331	70.86	13.12	0	99.4
Policy = 0	740	71.22	14.53	0	99.4
Policy = 1	1591	70.70	12.42	10.1	99.4

Table 6 Change in quantities and qualities associated with the PCI

	% within target waiting time		number of first appointments	
	top	non-top	top	non-top
trend (prior)	–0.021	0.048	–9.88	–15.51
	(–0.19)	(0.45)	(–0,38)	(–0.99)
	[–0.14]	[0.14]	[–0.36]	[–0.95]
	{–0.37}	{0.83}	{–0.61}	{–0.66}
trend (prior) squared	0.66	–3.07	0.26	0.44
	(0.25)	(–1.24)	(0.53)	(1.45)
	[0.19]	[–0,42]	[0.60]	[1.42]
	{0.80}	{–2.16}**	{0.77}	{0.89}
trend (after)	–0.43	0.74	–46.78	–63.90
	(–0.90)	(1.63)	(–1.57)	(–3.54)***
	[–0.43]	[1.05]	[–1.71]	[–2.36]**
	{2.06}**	{–2.05}**	{–1.47}	{–1.57}
trend (after) squared	0.003	–0.005	0.398	0.553
	(0.73)	(–1.56)	(1.70)*	(3.88)***
	[–0.35[]	[–0.91]	[2.01]*	[2.57]**
	{2.06}**	{–2.05}**	{1.55}	{1.71}
Policy	12.23	–25.20	–510.24	3108.88
	(0.82)	(0.07)	(–0.30)	(4.38)**
	[0.62]	[–1.25]	[0.35]	[2.98]**
	{2.34}**	{–2.52}**	{–0.37}	{2.20}**

t-statistics based on standard errors clustered at the hospital level in square brackets, based on clustered standard errors at the period level in curly brackets

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table 7 Change in quantities and qualities associated with the PCI (continuation)

	% within target waiting time		number of first appointments	
	top	non-top	top	non-top
Δ definition of waiting time target	–0.678	–3.50		
	(–0.40)	(–2.18)**		
	[–0.51]	[–1.93]*		
	{–1.31}	{4.08}***		
Surgery WT_{t-1}			5.25	18.11
			(0.56)	(3.75)***
			[1.07]	[2.12]**
			{0.81}	{4.43}***
Surgery $WT_{t-1} \times$ Policy			11.77	–13.54
			(1.83)*	(–3.09)***
			[3.54]***	[–1.86]*
			{2.38}**	{–4.05}***
First $appointments_{t-1}$			–3.50	–2.53
			(–0.27)	(–1.18)
			[–0.35]	[–0.87]
			{–0,36}	{–1.97}*
First $appointments_{t-1} \times$ Policy			9.61	–0.789
			(0.80)	(–0.31)
			[1.39]	[–0.19]
			{1.09}	{–0.96}
Observ.	1003	3104	567	2331
Fixed effects (hosp. & month)	yes	yes	yes	yes
Nr Hospitals	12	37	9	37

t-statistics based on standard errors clustered at the hospital level in square brackets, based on clustered standard errors at the period level in curly brackets

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Acknowledgements I am grateful to the editors Luis Costa and Miguel Portela for their support and patience. I thank the thoughtful comments of the referees and of Francisca V Lopes, which gratefully improved the paper. All errors are my own.

References

Abadie A, Athey S, Imbens G, Wooldridge J (2020) Sampling-based versus design-based uncertainty in regression analysis. Econometrica 88(1):265–296

Abadie A, Athey S, Imbens G, Wooldridge J (2021) When should you adjust standard errors for clustering?. Manuscript

Barros PP, Brouwer WB, Thomson S, Varkevisser M (2016) Competition among health care providers: Helpful or harmful? Eur J Health Econ 17(3):229–233. https://doi.org/10.1007/s10198-015-0736-3

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? Q J Econ 119(1):249–275

Bjorvatn A, Ma A (2011) Patient mobility and public hospital waiting time. In Four Essays on Health Care Reforms in Norway, PhD Dissertation. University of Bergen

Brekke K, Canta C, Siciliani L, Straume O-R (2021) Hospital competition in a national health service: Evidence from a patient choice reform. J Health Econ 79. https://www.sciencedirect.com/science/article/pii/S0167629621000941

Brekke K, Siciliani L, Straume O-R (2011) Hospital competition and quality with regulated prices. Scand J Econ 113(2):444–469

Brekke KR, Gravelle H, Siciliani L, Straume OR (2014) Patient choice, mobility and competition among health care providers. In: Levaggi, R., Montefiori, M. (eds) health care provision and patient mobility. Dev Health Econ Public Policy 12. Springer, Milano. https://doi.org/10.1007/978-88-470-5480-6_1

Cameron AC, Miller D (2015) A practitioner's guide to cluster-robust inference. J Human Resour 50(2):317–372

Cooper Z, Gibbons S, Jones S, McGuire A (2011) Does hospital competition save lives? Evidence from the English NHS patient choice reforms. Econ J 121(554):F228–F260

Dowd C (2021) Donuts and distant CATEs: Derivative bounds for R&D extrapolation. Available at SSRN

Dragone D, Lambertini L (2020) Equilibrium existence in the Hotelling model with convex production costs. Reg Sci Urban Econ 84

Gaynor M (2006) What do we know about competition and quality in health care markets?. NBER Working Paper 12301. https://doi.org/10.3386/w12301

Gravelle H, Smith PC, Xavier A (2003) Performance signals in the public sector: the case of health care. Oxf Econ Pap 55:81–103

Hansen C (2007) Generalised least squares inference in panel and multilevel models with serial correlation and fixed effects. J Econom 140(2):670–694

Hoernig S (2003) Existence of equilibrium and comparative statics in differentiated goods Cournot oligopolies. Int J Ind Organ 21(7):989–1019

Iversen T (1997) The effect of private sector on waiting time in a National Health Service. J Health Econ 16:381–396

Katz M (2013) Provider competition and healthcare quality: more bang for the buck? Int J Ind Organ 31(5):612–625

Mackinnon J, Nielsen M, Webb M (2022) Cluster-robust inference: a guide to empirical practice. Working Paper

Moscelli G, Gravelle H, Siciliani L (2019) Effects of market structure and patient choice on hospital quality for planned patients. CHE Research Paper 162

Sá L, Siciliano L, Straume OR (2019) Dynamic hospital competition under rationing by waiting times. J Health Econ 66(260):282

Santos R, Gravelle H, Propper C (2015) Does quality affect patients' choice of doctor? Evidence from England. Econ J 445–494

Siciliani L (2005) Does more choice reduce waiting times? Health Econ 14(1):17–23

Siciliani L (2006) A dynamic model of supply in the presence of waiting times and waiting lists. J Health Econ 25(6):891–907

Siciliani L (2008) A note on the dynamic interaction between waiting lists and waiting times. Health Econ 17:639–647
Siciliani L, Straume O-R (2019) Competition and equity in health care markets. J Health Econ 64(March):1–14. https://doi.org/10.1016/j.jhealeco.2018.12.002
Simões J, Augusto GF, Fronteira I (2017a) Introduction of freedom of choice for hospital outpatient care in Portugal: Implications and results of the 2016 reform. Health Policy 121:1203–1207
Simões J, Augusto GF, Fronteira I, Hernández-Quevedo C (2017b) Portugal: Health system review. Health Syst Transit 19(2):1–184
Van Ackere A, Smith PC (1999) Towards a macro model of National Health Service waiting lists. Syst Dyn Rev 15(3):225–252
Xavier A (2003) Hospital competition, GP fundholders and waiting times in the UK internal market: the case of elective surgery. Int J Health Care Finance Econ 3:25–51