Efficiency spillovers from a waiting list management program

Pedro Pita Barros[1], Miguel Ferreira[1,2], Julian Perelman[2]

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

[2]Escola Nacional de Saúde Pública, Universidade Nova de Lisboa

INTRODUCTION

In Portugal as elsewhere, there has been a remarkable investment aimed at the reduction of waiting times and waiting lists for elective surgery. The Portuguese health system is indeed mostly characterized by public financing and provision, through a National Health Service (NHS); as it the case in most NHS-type health systems, co-payments are very low, and rationing is mostly achieved through waiting times. An ambitious program was launched in 2004, called SIGIC (Integrated System for the Management of Patients Listed for Surgery). This program included, among other features, an extra pay to hospitals for performing specific elective surgeries in waiting lists, the so-called "SIGIC interventions", and a maximum waiting time guarantee (MWTG) entitling the patients to be transferred to another public or private facility in case the MWTG was exceeded. By the end of 2005, the median waiting time for surgery was 8.6 months; in 2011, it had decreased to 3.2 months (for more data and information on waiting time policies in Portugal, see Barros, Cristovão, and Gomes (2013)).

In this paper, we examined the consequences of supply-side wait-reduction policies on untargeted hospital activities, an issue that has not been analysed so far in the literature. To do so, we studied the Portuguese experience of waiting list management, which provided generous financial incentives to increase the volume of interventions under the waiting list management program, based on extra payments for extra activity. Our hypothesis is relatively straightforward: under fixed capacity, an increased specialization in extra-paid surgeries may have a crowding-out effect, causing non-SIGIC interventions to decrease. However, if incentives also prompt hospitals to become more efficient, this boils down to a capacity increase with possible spillovers to non-SIGIC interventions, which therefore increase too. Under flexible capacity the situation is completely different, as extra capacity will have no impact on the normal activity. We tested these competing hypotheses (crowding-out *versus* spillovers), showing that the program, in addition to improving the efficiency of targeted interventions, also produced spillovers to untargeted interventions, whose efficiency and volume improved.

A large literature has examined the impact of relatively similar policies targeting the supply side implemented in various countries. We briefly summarize here-below some of these policy experiences and their impact (examples were mostly borrowed from Cullis, Jones, and Propper (2000); Schut and Varkevisser (2013); Siciliani, Borowitz, and Moran (2013); Siciliani and Hurst (2005)).

A first group of policies consisted in activity-based payments, often coupled with additional financial incentives. In Spain, in 1996, a package of measures was introduced including extra funding for extra activity, maximum waiting times, and financial rewards for hospitals reducing waiting lists. Between 1996 and 2000 activity increased by 28% and average waiting times were reduced from 210 days to 67 days. In Norway, the government implemented in 1997 an activity-based payment scheme covering 30% of the average Diagnosis Related Group (DRG) costs per inpatient treated, increasing coverage up to 55% in 2002. This policy led to an increase in the annual growth rate of hospital activity. In 1991, the Australian State of Victoria also started funding hospitals directly on activity for patients in waiting lists, with successful outcomes in terms of reduced waiting lists. A policy focusing hospital activity was equally implemented in 2001 in the Netherlands, with extra funding attributed to hospitals exceeding their pre-determined yearly base production. As a result, productivity increased by around 15%. Though, this policy was not that successful at reducing waiting list since higher productivity gains were not obtained for diagnoses with the longest waiting times.

A second group of policies consisted in increasing hospitals capacity. Among recent experiences, the Australian government made USD 650 million available in 2008 to fight waiting lists and waiting time. The objective was to increase capacity either in the public health system or by contracting activity in private hospitals. The plan increased capacity in 120 hospitals with more than 62,000 additional surgeries performed. However, waiting times did not decline, as the increase in the number of surgeries caused a raise in demand, which offset the higher hospitals' capacity.

A third type of supply-side policy was implemented in England, known as "targets and terror". In 2002-2003, a maximum waiting time of 12 months was established in England and Scotland, progressively decreasing until 18 weeks in 2006 (the "target"). In England, if the maximum waiting time was not respected, the hospital would suffer a reduction of up to 5% of its revenues (the "terror"); conversely, these penalties were not implemented in Scotland. Subsequently, the percentage of patients waiting more than 6 months became 6 to 9 percentage points lower in England as compared to Scotland. A relatively similar policy was implemented in Sweden in 2009, which also guaranteed patients a maximum waiting time ("target"); it also provided untreated patients the option to be treated in a different hospital after the maximum time elapsed, fully paid by the origin hospital ("terror"). Also extra funding was made available for counties that reduced waiting times. The results were promising, with reduction in waiting times and more patients receiving treatment within the maximum time.

To conclude, supply-side policies have been relatively successful, at least in the short term, to reduce waiting times and list and promote extra activity. However, these policies may have unintended consequences that deserve close scrutiny. We have already mentioned the case of lower waiting lists increasing demand, hence reducing expected benefits. Iversen (1993) also alerted to the reverse incentives created by giving greater budgets to hospitals with longer waiting lists, leading hospitals to manipulate their lists. In this paper, we address the issue of spillovers or crowding-out effects of waiting list management programs on untargeted interventions.

The paper is organized as follows. In the first section, we present a simple theoretical model that details the mechanisms leading to spillovers or crowding-out effects. The second section presents the data and empirical strategies. The results are presented in the third section, and discussed in the last section.

THEORETICAL MODEL

In this section, we present a simple theoretical model, which considers that the hospital maximizes a utility function depending of its revenues and patients' well-being. Both revenues and well-being depend on three types of activities: planned, unplanned, and extra-activity entitling to additional payments. In a first case, we assume that the capacity constraint is not binding, while it is binding in the second case; finally, the third case introduces a variable effort parameter, which enables for increasing the hospital capacity.

Let us thus consider that hospitals maximize a utility function as follows:

$$\max_{x,z} U = (pz + w - C(x+y+z))\gamma + \big(b(x+y) + f(z)\big)(1-\gamma) + (pz + w - C(x+y+z) - L)\delta \tag{1}$$

subject to

$$s.t\ K \geq x + y + z$$

where z is the extra activity, x the planned activity and y the emergency room activity, and w and p are the payments associated with planned and extra activity, respectively. The expression $b(x+y) + f(z)$ represents the patients benefits from treatment, while $c(x+y+z)$ represent the treatment costs. We postulate that benefit functions b and f are increasing and concave in all their arguments, while the cost function is increasing and convex, assuming diminishing marginal returns. The term $\gamma > 0$ reflects the respective weight of profits and patient benefit in the utility function of the hospital.

The term $\delta > 0$ captures the government pressure to achieve good financial results. In this model, we exclusively consider the case of public (NHS) hospitals, which the reform was addressed to. Public hospitals cannot be considered to be constrained by the budget as for-profit or not-for-profit private facilities, because losses are partly covered by the State, and do not lead to bankruptcy. Though, financial losses are penalized in some way, the most visible being the firing of administrators as it is systematically the case in the UK (Propper, Sutton, Whitnall, & Windmeijer, 2008). Finally, the expression $K \geq x + y + z$ represents the hospital capacity constraint.

1. *Non-binding capacity constraint*

We examine first the case of a non-binding capacity constraint. We get the following Lagrangian equation:

$$L = \big(pz + w - C(x + y + z)\big)(\gamma + \delta) + \big(b(x + y) + f(z)\big)(1 - \gamma) - \delta(L) \qquad (2)$$

for which the first-order conditions (FOC) are given by:

$$\text{z: } f'(z)(1 - \gamma) - (\gamma + \delta)(C' - p) = 0 \qquad (3)$$

$$\text{x: } b'(x + y)(1 - \gamma) - (\gamma + \delta)(C') = 0$$

From total differentiation of equations (3) and , we extract the impact on z and on x of an increase in p:

$$\begin{cases} (\gamma + \delta)\big(dp - C''(dx + dz)\big) + (1 - \gamma)f''dz = 0 \\ (\gamma + \delta)(-C''(dx + dz)) + (1 - \gamma)b''dx = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} [-(\gamma + \delta)C'' + (1 - \gamma)f'']dz - (\gamma + \delta)C''dx = -(\gamma + \delta)dp \\ [-(\gamma + \delta)C'']dz + [(1 - \gamma)b'' - (\gamma + \delta)C'']dx = 0 \end{cases}$$

Using the expressions above we are able to calculate the derivatives of z and x in order to p

$$\frac{dz}{dp} = \frac{1}{|H|}\{-(\gamma + \delta)[(1 - \gamma)b'' - (\gamma + \delta)C'']\} > 0$$

$$\frac{dx}{dp} = \frac{1}{|H|}[-(\gamma + \delta)^2 C''] < 0$$

where

$$H = [(1 - \gamma)b'' + (\gamma + \delta)(-C'')][(1 - \gamma)f'' + (\gamma + \delta)(-C'')] - (\gamma + \delta)^2(C'')^2 > 0$$

So, an increase in p will cause an increase in z but a decrease in x. The rationale is that x and z are substitutes in the hospital utility function, and z has a higher weight due to the extra payment; it is thus profitable to decrease planned activity as extra payment increases, even in the absence of a capacity constraint. This finding is also justified by the total activity level determining costs under constant returns to scale, so that the normal activity is independent of the price of the extra activity.

In the situation of an expansion in the emergency room activity (y), the impact on x and z is given by

$$\begin{cases} (\gamma + \delta)\big(-C''(dx + dy + dz)\big) + (1 - \gamma)f''dz = 0 \\ (\gamma + \delta)(-C''(dx + dy + dz)) + (1 - \gamma)b''(dx + dy) = 0 \end{cases}$$

$$\Leftrightarrow\begin{cases} [-(\gamma+\delta)C''+(1-\gamma)f'']dz-(\gamma+\delta)C''dx=(\gamma+\delta)C''dy \\ [-(\gamma+\delta)C'']dz+[(1-\gamma)b''-(\gamma+\delta)C'']dx=[(\gamma+\delta)C''-b''(1-\gamma)]dy \end{cases}$$

and, differentiating expressions, we obtain

$$\frac{dz}{dy}=\frac{1}{|H|}\{(\gamma+\delta)C''[(1-\gamma)b'']+[(\gamma+\delta)C''-b''(1-\gamma)](\gamma+\delta)C''\}\lesseqgtr 0$$

$$\frac{dx}{dy}=\frac{1}{|H|}\big[[-(\gamma+\delta)C''+(1-\gamma)f''][(\gamma+\delta)C''-b''(1-\gamma)] + (\gamma+\delta)C''^2(\gamma+\delta)\big]\gtreqless 0$$

The stability conditions provide no extra information about the signal of the effects.

An increase in emergency room activity has an ambiguous effect on both planned and extra activity. The increase in emergency visits is favorable for the patients' utility, but it encompasses an increase in costs that is not compensated by higher revenues. The impact on planned and extra activity will depend on the magnitude of these competing effects.

2. Binding capacity constraint

In this case, the Lagrangian is given by:

$$L=(p(K-x-y)+w-C(K))\gamma+\big(b(x+y)+f(K-x-y)\big)(1-\gamma)-\delta(C(K)-p(K-x-y)-w)-\alpha(K-(x+y+z)) \quad (4)$$

Deriving the Lagrangian with respect to x, we obtain

$$\frac{dL}{dx}=(\gamma+\delta)(-p)+(1-\gamma)[b'(x)-f'(K-x-y)]=0$$

from which we measure the impact of a variation of p on x

$$-(\gamma+\delta)dp+(1-\gamma)[b''-f'']dx=0$$

$$\frac{dx}{dp}=\frac{-(\gamma+\delta)}{-(1-\gamma)[b''+f'']}<0$$

If the capacity restriction is binding, x will again decrease with an increase in p. It is straightforward to demonstrate that the greater extra payment will also provoke a growth in extra activity. In this situation, an increase in emergency room activity will cause a reduction in x as well.

$$\frac{d^2L}{dxdy}=(1-\gamma)f''(K-x-y)<0$$

3. Introducing variable effort

We now consider that the hospital is able to exert some effort e that affects positively its capacity $K(e)$ but has a cost given by $h(e)$. The objective function is now

$$\max_{x,z} U = \big(pz + w - C(x + y + z, g)\big)(\gamma + \delta) + \big(b(x + y) + f(z)\big)(1 - \gamma) - h(e) \quad (5)$$

$$s.t\ K(e) \geq x + y + z$$

As we are testing how gains in efficiency translate into greater capacity, only the binding capacity situation is relevant. From the maximization problem, we get the following Lagrangian:

$$L = (pz + w - C(K(e), g))(\gamma + \delta) + \big(b(K(e) - z) + f(z)\big)(1 - \gamma) - h(e) \quad (6)$$

and the FOC are given by

$$\text{e:}[(\gamma + \delta)(-C') + (1 - \gamma)b']K' - h' = 0 \quad (7)$$

$$\text{Z: } (-b' + f')(1 - \gamma) + (\gamma + \delta)p = 0$$

We evaluate again whether an increase in extra payment incentivizes hospitals to increase activity, now with the possibility for hospitals to increase their capacity through enhanced effort. To do so, we totally differentiate equations (7):

$$(-b''(K'de - dz) + f''dz)(1 - \gamma) + (\gamma + \delta)dp = 0$$

$$\Leftrightarrow -\big((1 - \gamma)b'' + (1 - \gamma)f''\big)dz + b''(K')(1 - \gamma)de = (\gamma + \delta)dp$$

$$[(\gamma + \delta)(-C''K'de) + (1 - \gamma)b''(K'de - dz)]K' - h''de + K''(e)[h']de[(\gamma + \delta)(-C') + (1 - \gamma)b'] = 0$$

$$\Leftrightarrow -K'(1 - \gamma)b''dz + \Big(K''h'[(\gamma + \delta)(-C') + (1 - \gamma)b'] - h'' + K'^2\big((\gamma + \delta)(-C'') + (1 - \gamma)b''\big)\Big)\, de = 0$$

from which we obtain

$$\frac{dz}{dp} = (\gamma + \delta)\Big(K''h'[(\gamma + \delta)(-C') + (1 - \gamma)b'] - h'' + K'^2\big((\gamma + \delta)(-C'') + (1 - \gamma)b''\big)\Big)\frac{1}{|H|} > 0$$

$$\frac{de}{dp} = (1 - \gamma)b''K'(\gamma + \delta)\frac{1}{|H|} > 0$$

with $|H|$<0. The impact of the extra payment is thus positive both on extra activity and effort.

As regards the impact on the planned activity, we use

$$K(e) = (x + y + z) \Leftrightarrow x = K(e) - y - z$$

and computing the derivatives with respect to p, we get

$$\frac{dx}{dp} = \frac{dK}{de}\frac{de}{dp} - \frac{dz}{dp}$$

so that the impact on planned activity is ambiguous. If the price-related increase in effort is large enough to offset the greater z, x will increase. If by contrast effort does not change sufficiently so that it fully captured by the increase in z, x will decrease.

Finally, in order to evaluate the situation of an expansion in the emergency room activity (y) on x, z and e, let us consider the following Lagrangean:

$$L = (p(K-x-y)+w-C(K))\gamma + \big(b(x+y)+f(K-x-y)\big)(1-\gamma) - \delta(C(K) - p(K-x-y)-w) - \alpha\big(K(e)-(x+y+z)\big) - h(e) \quad (8)$$

The FOCs are:

$$Z\text{: } f'(z)(1-\gamma)-(\gamma+\delta)(C'-p)=0 \quad (9)$$

$$X\text{: } b'(x+y)(1-\gamma)-(\gamma+\delta)(C')=0$$

$$e\text{: } -h'-\alpha K'=0$$

and deriving these conditions, we extract the impact of an increase in y:

$$\begin{cases} (\gamma+\delta)\big(-C''(dx+dy+dz)\big)+(1-\gamma)f''dz=0 \\ (\gamma+\delta)(-C''(dx+dy+dz))+(1-\gamma)b''(dx+dy)=0 \\ -h''de-\alpha K''de=0 \end{cases}$$

$$\Leftrightarrow \begin{cases} [-(\gamma+\delta)C''+(1-\gamma)f'']dz-(\gamma+\delta)C''dx=(\gamma+\delta)C''dy \\ [-(\gamma+\delta)C'']dz+[(1-\gamma)b''-(\gamma+\delta)C'']dx=[(\gamma+\delta)C''-b''(1-\gamma)]dy \\ -de[h''+\alpha K'']=0 \end{cases}$$

so that

$$\frac{dz}{dy}=\frac{1}{|H|}\{(\gamma+\delta)C''[(1-\gamma)b''-(\gamma+\delta)C''] + [(\gamma+\delta)C''-b''(1-\gamma)](\gamma+\delta)C''\}[h''+\alpha K''] \lesseqgtr 0$$

$$\frac{dx}{dy}=\frac{1}{|H|}\big[[-(\gamma+\delta)C''+(1-\gamma)f''][(\gamma+\delta)C''-b''(1-\gamma)] + (\gamma+\delta)C''^2(\gamma+\delta)\big][h''+\alpha K''] \gtreqless 0$$

$$\frac{de}{dy}=0$$

The impact of variations in emergency stays is ambiguous on both planned and extra activity, while no impact is expected on effort.

The outcomes from the theoretical model are summarized in Table 1. An increase in the price of extra activity is expected to boost the volume of extra activity in all cases, and to decrease the volume of planned activity except when the effort increases so that spillovers occur. The rise in unplanned activity is expected to decrease the volume of extra and planned activity under binding capacity constraints, but again this decline is not certain when the effort is variable. The remaining of the paper is devoted to testing the impact of payment and emergency stays on effort, planned and unplanned activity, using Portuguese data.

DATA AND EMPIRICAL MODELS

We used data from all surgical stays at all National Health System (NHS) hospitals from 2005 until 2011, constituting a sample of 2,979,168 observations. Data included fully comparable information on patient's age, sex, Diagnosis Related Group (AP-DRG), type of admission (elective *versus* non-elective, SIGIC *versus* non-SIGIC), hospital and year of the intervention. All variables used in the model are summarized in Table 2, with their definition and representation in both theoretical and empirical models.

1. *Impact on effort*

In the empirical model, we considered as dependent variable the in-patient length of stay (LOS), as proxy for the hospital efficiency-enhancing effort (variable *e* in the theoretical model, see Table 2). Considering the maximization problem presented in the previous section, we modeled LOS as a function of patient benefits – proxied by age and gender –, and as a function of the extra revenue of extra activity – proxied by the patient being treated as SIGIC. Indeed, the marginal revenue of planned activity and emergency activity (non-SIGIC) being zero, the surgeries being classified as SIGIC can be assumed as the best proxy for receiving extra payment.

Additionally, we included as explanatory variables the percentage of SIGIC cases treated in the hospital, as a second proxy for the price effect, reflecting the cumulated incentive to enhance efforts. In the same line, the percentage of emergency admissions was included to reflect the cumulative cost of treating unplanned patients. Finally, year, hospital and DRG fixed effects were factored in. That is, we estimated:

$$LOS_{igth} = \beta_0 + \beta_1 age_{igth} + \beta_2 sex_{igth} + \beta_3 SIGIC_{igth} + \beta_4 \%SIGIC_{th} + \beta_5 VOL_{th}^{URGENT} + \sum_{t=2001}^{2011} YEAR_t + HOSP_h + DRG_g + e_{igth} \quad \text{(Model 1)}$$

where LOS_{igth} is the length of stay of patient *i* treated in AP-DRG g, hospital *h* in year *t*; age_{igth} and sex_{igth} are the patient's age and sex; $SIGIC_{igth}$ is a variable that values 1 if the patient received a SIGIC intervention, 0 otherwise; $\%SIGIC_{th}$ is the percentage of surgeries financed through the SIGIC program; and VOL_{th}^{URGENT} is the volume of non-elective admissions. The variables $YEAR$, DRG and $HOSP$ represent the year, the DRG, and the hospital fixed effects, respectively.

Then, we replicated the estimation for the SIGIC and non-SIGIC cases separately:

$$LOS_{igth}^{SIGIC} = \beta_0 + \beta_1 age_{igth} + \beta_2 sex_{igth} + \beta_3 SIGIC_{igth} + \beta_4 \%SIGIC_{th} + \beta_5 VOL_{th}^{URGENT} + \sum_{t=2001}^{2011} YEAR_t + HOSP_h + DRG_g + e_{igth} \quad \text{(Model 2)}$$

$$LOS_{igth}^{non-SIGIC} = \beta_0 + \beta_1 age_{igth} + \beta_2 sex_{igth} + \beta_3 SIGIC_{igth} + \beta_4 \%SIGIC_{th} + \beta_5 VOL_{th}^{URGENT} + \sum_{t=2001}^{2011} YEAR_t + HOSP_h + DRG_g + e_{igth} \quad \text{(Model 3)}$$

These separate estimations aimed at measuring the extent to which the efforts to enhance efficiency and capacity were also oriented towards untargeted interventions. In other terms, these equations allowed us test the extent to which the waiting list management program also impacted the management of untargeted surgeries.

Estimations were performed using normal OLS with LOS in logarithm to account for the non-negative and right-skewed distribution of this variable.

2. Impact on volume

As a second step, we tested whether the SIGIC program promoted an increase in the number of SIGIC and non-SIGIC surgeries, that is, the extent to which changes in efficiency translated into a higher volume of interventions. Following the theoretical model, we estimated the impact of the SIGIC payment on the total volume of interventions (VOL_{ght}^{TOT}):

$$VOL_{ght}^{TOT} = \beta_0 + \beta_1(\%SIGIC_{ht}) + \beta_2 VOL_{th}^{URGENT} + \beta_3 age_{ght} + \beta_4 sex_{ght} + \\ + \sum_{t=2001}^{2011} \beta_{1t}\, Year_t + DRG_g + HOSP_h \qquad \text{(Model 4)}$$

Note that contrary to the model on effort, the age and sex are average values for the Year/DRG/hospital group, because we now work with aggregate data. Next, we estimated the impact on the volume of SIGIC and non-SIGIC interventions, through simultaneous equations:

$$VOL_{ght}^{Non-SIGIC} = \beta_0 + \beta_1(\%SIGIC_{ht}) + \beta_2 VOL_{th}^{URGENT} + \beta_3 age_{ght} + \beta_4 sex_{ght} + \\ \sum_{t=2001}^{2011} Year_t + DRG_g + HOSP_h \qquad \text{(Model 5)}$$

$$VOL_{ght}^{SIGIC} = \beta_0 + \beta_1(\%SIGIC_{ht}) + \beta_2 VOL_{th}^{URGENT} + \beta_3 age_{ght} + \beta_4 sex_{ght} + \\ \sum_{t=2001}^{2011} Year_t + DRG_g + HOSP_h \qquad \text{(Model 6)}$$

By doing so, we tested again the spillover versus crowding-out hypotheses (non-SIGIC interventions). To estimate these models we used yearly data from 2005 until 2011, aggregated by hospital and DRG. The total number of observations was 66,653 for total interventions, 55,789 for SIGIC cases, and 66,653 for non-SIGIC cases. Estimations for Models 5 and 6 were performed using a seemingly-unrelated equation model, because the volume of SIGIC and non-SIGIC interventions is a joint decision, which thus cannot be treated as independent variables.

RESULTS

1. Descriptive analysis

The evolution of mean length of stay of all NHS hospitals between 2005 and 2011 is represented on Figure 1. We observe that the LOS has been decreasing continuously since the implementation of SIGIC, from 5.6 days in 2005 to 4.1 days in 2011. This decline was however previous to the waiting list management program, i.e., the average length of stay in 2000 was 7.3 days (OECD, 2014). The total volume of cases has increased from 2005 to 2011, from 44.8 cases per DRG to 57.2, also following the previous pattern. However, from 2009 to 2011 the percentage of SIGIC interventions experienced a decrease from around 10% to around 7%, while the volume of surgeries was still increasing. Also, the percentage of urgent cases decreased continuously until 2009, and remained stable thereafter. The volume of both SIGIC and non-SIGIC cases increased until 2008; thereafter, surgeries under SIGIC program decreased while surgeries not included on the SIGIC increased. Finally, the size of waiting lists increased between 2006 and 2007, and then decreased until 2010.

2. *Testing the efficiency impact*

As regards the efficiency impact, most estimates presented the expected signs (Table 3). The SIGIC stays were significantly shorter, and patients treated in hospitals with a higher percentage of SIGIC cases also experienced significantly shorter stays. An additional percentage point of SIGIC surgery was associated to a 0.2% lower length of stay. Also, the LOS was positively associated to the volume of urgent admissions. Possibly, the greater intensity of unplanned admissions reduced the hospital capacity to manage the stays in a more efficient way.

We then estimated the impact of the same variables on LOS considering SIGIC and non-SIGIC discharges separately. The signs of the variables were the same for both sub-samples. However the percentage of SIGIC cases was more influential on SIGIC cases, as indicated by the greater magnitude of the estimate. This indicates that efficiency gains were obtained on both SIGIC and non-SIGIC cases, corresponding to the spillover hypothesis, but to a larger extent on incentivized interventions.

2. *Testing the volume impact*

We observed that the percentage of SIGIC surgeries had a positive significant impact on the total volume of surgeries (Table 4). An additional percentage point of SIGIC surgery was associated to a 0.1% increase in the total volume of surgeries. To evaluate crowding-out *versus* spillover effects, we examined non-SIGIC cases separately. The percentage of SIGIC interventions had a positive impact on both SIGIC and non-SIGIC surgeries, but the effect was only significant (and greater) for non-SIGIC surgeries. This confirms again that the spillover effect was stronger than the crowding-out effect. To conclude, a higher volume of SIGIC interventions caused the volume of total and non-SIGIC surgeries to increase.

Note finally that the volume of urgent interventions was positively and significantly related to the total volume of treated cases. The model predicted that more urgent cases may signify a lower capacity to perform planned interventions and develop extra activity. However, this is not the case when the efficiency can be enhanced, as we demonstrated to be the case in hospitals with more unplanned interventions, which experience a lower LOS.

3. *Extension: the impact on waiting lists*

The final objective of the SIGIC program was to reduce the size of waiting lists, and the waiting times. We thus expected the greater efficiency and volume to achieve this aim. Consequently, we checked the extent to which hospitals with a higher efficiency and volume of SIGIC cases achieved shorter waiting lists and waiting times. To do so, we estimated the following model:

$$List_{th} = \beta_1 \% SIGIC_{th} + \beta_2 LOS_{th} + \beta_3 age_{th} + \beta_4 sex_{th} + \beta_4 prior_{th} + \sum_{t=2001}^{2011} Year_t + HOSP_h \quad \text{(Model 7)}$$

where $List_{th}$ denotes the size of the waiting list of year t at hospital h, and $prior_{th}$ the percentage of high-priority cases during year t at hospital h,

and

$$Time_{th} = \beta_1 \% SIGIC_{th} + \beta_2 LOS_{th} + \beta_3 age_{th} + \beta_4 sex_{th} + \beta_4 prior_{th} + \sum_{t=2001}^{2011} Year_t + HOSP_h \quad \text{(Model 8)}$$

where $Time_{th}$ denotes the average waiting time of year t at hospital h.

In order to estimate these effects, we merged the original data with a database including all NHS hospitals' average waiting times and lists for the 2006-2011 period (N=2,979,168). Data were aggregated by hospital and year, providing a final database of 183 observations. Estimations were performed using panel data analysis with hospital fixed effects.

Results indicated that a higher intensity on SIGIC is associated to smaller waiting lists, with an additional one-point percentage of SIGIC cases causing a 0.9% reduction of the waiting list (Table 5). Though, the size of the waiting list was not significantly associated to shorter stays, as proxy for the hospital efficiency. In contrast, a greater efficiency was significantly related to shorter waiting times for surgery; an additional day in the average LOS was related to a 6.5% longer waiting time.

Discussion

Likewise other NHSs, Portugal has faced during many years long waiting times and waiting lists for hospital treatments. Although waiting rations the demand in the absence of copayments, excess waiting times are not optimal because people's health and capacity to recover deteriorate, with damaging consequence in terms of quality of life, productivity, satisfaction, and support for the NHS. The waiting list management program, implemented in 2004, consisted in imposing maximum waiting times and offering incentives to hospitals to treat patients on waiting lists, through relatively generous fee-for-service payments.

In this paper, we postulated that incentives focusing on specific interventions could harm or benefit non-incentivized treatments, depending on the existence of crowding-out or spillover effects. This hypothesis largely stems from the multitasking problem, which is common to performance-based payments; because not all aspects of workers or institutions' behavior are rewarded, "agents emphasize only those aspects of performance that are rewarded", leading to "a reallocation of activities toward those that are directly compensated and away from the uncompensated activities" (Prendergast, 1999). Instead, the incentives may have promoted a greater efficiency diffusing to other non-incentivized dimensions, with larger benefits for the patient.

Our theoretical model demonstrated indeed that the extra payment favors the volume of extra activity, which substitutes for planned activity. Though, the extra payment was also shown to generate greater levels of effort to enhance efficiency. Thereby, we showed that planned activity would be crowded out by additional activity unless efficiency gains were large enough; in that case, the greater efficiency would lead to an increase in the hospital capacity, avoiding a reallocation of resources detrimental to planned activity. Finally, we emphasized the role of emergency visits, as the unplanned dimension of the hospital activity, which may subsequently reduce the capacity to increase the volume of planned and extra interventions.

The empirical model provided evidence that the extra payment was substantially associated to lower length of stay for both SIGIC and non-SIGIC interventions. The data also indicated the existence of spillovers, with a greater volume of non-SIGIC and total cases associated to extra payments. Thereafter, we measured a significant association of the greater intensity of SIGIC cases and shorter waiting lists, and a significant link between greater efficiency and shorter waiting times.

These results confirm those obtained in the literature as regards the positive impact of activity-based payments on the volume of interventions, which allows for reductions in waiting lists (Siciliani et al., 2013). Additionally, the findings are encouraging because they show that incentives promote changes over a larger scale at the hospital level, with beneficial effects on the volume of non-incentivized interventions. To our best knowledge, the existence of spillovers has not been addressed in the literature on hospitals; however, at the primary care level, there is now evidence that pay-for-performance also diffused to untargeted diseases and interventions, in line with our results (Langdown & Peckham, 2014). These findings highlight the potential of financial incentives to modify practices to a large extent, provided they are large enough, without the need to incentivize all dimensions of care and all health conditions.

This study suffers from several limitations. First, we lacked of a real counter-factual, as all hospitals were submitted to the reform, and pre-reform practices could not serve as control because interventions were not classified as SIGIC/non-SIGIC at that time. Consequently, we used as control the non-SIGIC interventions, which may not be fully comparable. We however minimized the risk of bias through controlling for DRGs, reflecting the differences in health conditions and procedures. Second, it would have been important to control for the hospital financial situation, because the ability to perform extra activity and improve efficiency might have been higher in wealthier settings. We included hospital fixed effects in all models to account for this possible confounding factor. Finally, other changes have been occurring across time possibly biasing results, such as the reinforcement of primary or the cuts in hospital budgets related to the economic recession. Once again, the inclusion of year fixed effect aimed at reducing this bias.

To conclude, this study highlights the potential of financial incentives to increase the activity, and subsequently to reduce waiting times and waiting lists. Importantly, findings are encouraging as regards the diffusion of more efficient practices to non-incentivized interventions.

Appendix

To see how the marginal costs react to an increase in p we use

$$C''_{zp} = \frac{d^2C}{dz^2}\frac{dz}{dp}$$

$$C''_{xp} = \frac{d^2C}{dx^2}\frac{dx}{dp}$$

$$\frac{d\,mgC}{dp} = \frac{d^2C}{dz^2}\frac{dz}{dp} + \frac{d^2C}{dx^2}\frac{dx}{dp}$$

By definition, as marginal costs are convex, the second derivative will be positive. For z we found that the derivative of z with respect to p is always positive, so the marginal costs of z will increase if p increases. The x case will be different since the variation of x, given an increase in p, may be negative or positive. However, even if the marginal cost of x decreases with an increase in p, as the variation of x is smaller than the variation of z, the marginal costs will always increase, given that the increase in the marginal costs of z will offset the decrease in the marginal costs of x.

REFERENCES

Barros, Pedro, Cristovão, R, & Gomes, Pedro. (2013). Portugal. In L. Siciliani, M. Borowitz & V. Moran (Eds.), *Waiting time policies in the health sector: what works?, OECD Health Policy Studies* (pp. 237-262): OECD publishing.

Cullis, J, Jones, P, & Propper, C. (2000). Waiting lists and medical treatment: analysis and policies. *Handbook of Health Economics, 1*, 1201-1249.

Iversen, T. (1993). A theory of hospital waiting lists. *Journal of Health Economics, 12*(1), 55-71.

Langdown, Carwyn, & Peckham, Stephen. (2014). The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review. *Journal of Public Health, 36*(2), 251-258.

OECD. (2014). OECD Health Statistics 2014, online. Retrieved 3-2-2015, from OECD Publishing http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT

Prendergast, Canice. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 7-63.

Propper, Carol, Sutton, Matt, Whitnall, Carolyn, & Windmeijer, Frank. (2008). Did'targets and terror'reduce waiting times in England for hospital care? *The BE Journal of Economic Analysis & Policy, 8*(2).

Schut, F, & Varkevisser, M. (2013). Tackling hospital waiting times: The impact of past and current policies in the Netherlands. *Health Policy, 113*(1), 127-133.

Siciliani, L, Borowitz, M, & Moran, V. (2013). *Waiting Time Policies in the Health Sector: What Works?* : OECD Publishing.

Siciliani, L, & Hurst, J. (2005). Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. *Health Policy, 72*(2), 201-215.

Table 1. Expected effects from the theoretical model

	Increase in the price of extra activity	**Increase in unplanned interventions**
1. Non-binding capacity constraint	Increase in extra activity Decrease in planned activity	Ambiguous effects on extra and planned activity
2. Binding capacity constraint wo/ variable effort	Increase in extra activity Decrease in planned activity	Decrease in extra activity Decrease in planned activity
3. Binding capacity constraint w/ variable effort	Increase in effort Increase in extra activity Ambiguous effect on planned activity	No impact on effort Ambiguous effects on extra and planned activity

Table 2. Description of the variables used in the empirical model.

Empirical specification	Definition	Theoretical specification	Mean (S.D.)
LOS	Length of stay	Effort (e)	4.7 (15.9)
Age	Age	Patient benefit (b)	52.0 (22.1)
Sex	Female sex (0/1)	Patient benefit (b)	58.7%
SIGIC	Intervention under the SIGIC program (0/1)	Price of extra activity (p)	7.7%
%SIGIC	Percentage of interventions under the SIGIC program	Price of extra activity (p)	7.7%
VOL^{URGENT}	Volume of unplanned interventions	Unplanned intervention (y)	23.8%
$VOL^{Non\text{-}SIGIC}$	Mean volume of interventions out of the SIGIC program, by year-DRG	Planned activity intervention (z)	48.7 (136.5)
VOL^{SIGIC}	Mean volume of interventions under the SIGIC program, by year-DRG	Extra activity intervention (z)	4.1 (38.2)
YEAR	Year dummy		.
HOSP	Hospital dummy		.
DRG	DRG dummy		.

Table 3. Determinants of length of stay[1].

Variables	Estimate (Standard Error)
Complete sample	
SIGIC patient	-0.246** (0.002)
Percentage of SIGIC patients	-0.266** (0.012)
Volume of urgent surgeries	0.108** (0.013)
Age	0.006** (0.000)
Female	-0.029** (0.001)
R-square	55.41%
SIGIC cases	
SIGIC patient	.
Percentage of SIGIC patients	-0.544** (0.045)
Volume of urgent surgeries	-0.040 (0.093)
Age	0.005** (0.000)
Female	-0.009** (0.003)
R-square	66.71%
Non-SIGIC cases	
SIGIC patient	.
Percentage of SIGIC patients	-0.240** (0.012)
Volume of urgent surgeries	0.107** (0.004)
Age	0.006** (0.000)
Female	-0.030** (0.001)
R-square	65.66%

Notes. [1]Models include year, hospital, and DRG fixed effects. Significance: *p-value<0.05; **p-value<0.01.

Table 4. Determinants of volume[1].

Variables	Estimate (Standard Error)
Total volume	
Percentage of SIGIC patients	0.113** (0.048)
Volume of urgent surgeries	0.008** (0.001)
Mean age	0.001 (0.000)
Percentage women	0.005 (0.016)
R-square	69.25%
Volume of SIGIC cases	
Percentage of SIGIC patients	0.122 (0.129)
Volume of urgent surgeries	0.002** (0.001)
Mean age	0.001 (0.000)
Percentage women	0.057* (0.031)
R-square	53.93%
Volume of non-SIGIC cases	
Percentage of SIGIC patients	0.228** (0.091)
Volume of urgent surgeries	0.010** (0.001)
Mean age	-0.002** (0.001)
Percentage women	0.038 (0.050)
R-square	74.53%

Notes. [1]Models include year, hospital, and DRG fixed effects. Significance: *p-value<0.10; **p-value<0.05.

Table 5. Determinants of the size of waiting lists and waiting times.

Variables	Estimate (Standard Error)
Waiting lists	
Percentage of SIGIC patients	-0.868** (0.427)
Mean length of stay	-0.026 (0.036)
Mean age	0.062** (0.019)
Percentage women	0.063** (0.028)
Percentage high-priority cases	-0.859 (0.957)
R-Squared	96.20%
Waiting times	
Percentage of SIGIC patients	-0.143 (0.378)
Mean length of stay	0.065* (0.033)
Mean age	-0.041** (0.016)
Percentage women	-0.020 (0.014)
Percentage high-priority cases	-0.555 (1.147)
R-Squared	92.09%

Notes. [1]Models include year and hospital fixed effects. Significance: *p-value<0.10; **p-value<0.05.