



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

**GPs and hospital expenditures.
Should we keep expenditure
containment programs alive?**

Gianluca Fiorentini
Matteo Lippi Bruni
Cristina Ugolini

Quaderni - Working Paper DSE N° 829



GPs and hospital expenditures. Should we keep expenditure containment programs alive?

Gianluca Fiorentini, Matteo Lippi Bruni and Cristina Ugolini

Department of Economics, University of Bologna

Correspondence to:

Matteo Lippi Bruni, Piazza Scaravilli 2, 40126, Bologna, Italy.

Tel. +39-051-2098145 - Fax +39-051-2098040.

matteo.lippibruni2@unibo.it

Acknowledgements

The paper is part of the research project “*The role of economic incentives in the governance of primary care*” financed by the Health Department of Emilia-Romagna, Italy. The authors wish to thank the Regional Agency for Healthcare Services of Emilia Romagna for providing the data. Elisa Iezzi granted valuable research assistance throughout the entire project. We are also grateful to Stefania Bortolotti and Rinaldo Brau for useful suggestions and to Rosella Verzulli who commented an earlier version of the paper. Helpful indications were provided also by participants to the Annual Conference of the Italian Association of Health Economics held in Naples. The usual disclaimers apply.

22 May 2012

Abstract

Pay-for-performance programs offering additional payments to GPs can be used not only to improve the quality of care but also for cost containment purposes. In this paper, we analyse the impact of removing financial incentives in primary care that were aimed at containing hospital expenditure in the Italian region of Emilia Romagna during the period 2002-04. Our analysis draws on regional databanks linking GPs' characteristics to those of their patients (including all sources of public payments made to GPs), together with information on the utilisation of hospital services. We employ a difference-in-difference specification to assess changes in expenditures for avoidable and total hospital admissions. We identify the treatment group with GPs operating in districts where the program is withdrawn during the observation period ("*Leavers*"). Their performance is compared to that of two separate control groups, namely: GPs working in districts that grant incentives for the entire period ("*Stayers*"), and those working in districts that never introduced measures for the containment of hospitalisations ("*Non Participants*").

The comparison between treatment and control groups shows that removing incentives does not result in a worse performance by *Leavers* compared to both control groups. This supports the policy of removing incentives, as such entail extra payments to GPs which, however, do not seem capable of significantly influencing their behaviour in the desired ways. Our findings complement previous evidence from the same institutional context showing that only those programs that aim to improve disease management for specific conditions - rather than to simply contain expenditure - have proven successful in reducing avoidable admissions for the target population.

Key words: Health economics, primary care, hospital expenditure, ambulatory care sensitive conditions, economic incentives.

JEL classification: I11, I18, C31

1. Introduction

Publicly-funded health systems are increasingly struggling to improve the efficiency of resource allocation through cost-conscious decisions by health professionals and institutions, because of the widening gap between health needs and available resources. At the same time, cost containment must be weighed against its possible adverse effects on health outcomes. Consequently, initiatives for improving the incentives offered to healthcare providers are currently a central concern for health-policy makers.

Ensuring the effectiveness of the design of incentives first involved the hospital sector, due to its organisational complexity, the concentration of (often irreversible) physical and human capital investment, the severe case-mix, all of which calling for the careful planning and implementation of service provision. More recently, a number of other critical areas have emerged, including the coordination between primary and secondary care. This is a consequence of hospitals' increasing specialisation in the provision of acute care, which has implied the transfer of responsibility for low-intensity treatments to the district level. Moreover, General Practitioners (GPs) provide assistance in areas such as preventive care, chronic diseases and post-acute follow ups. Consequently, their role as both providers and gatekeepers is essential to the appropriate utilisation of the different levels of care and to the reconciling of cost-containment policies with successful outcomes.

GPs' remuneration schemes, based exclusively on capitation, are considered not to be fully effective in inducing optimal professional effort on the part of the GPs themselves [Iversen & Luras 2006]. Given that educational strategies alone, such as guidelines and protocols, have not proven completely successful [Grilli et al. 2000], economic incentives such as Pay-for-Performance (P4P) programs often represent additional means by which to reinforce the governance of healthcare delivery. Usually, they add up to capitation and reward high quality care, as well as the achievement of specific policy targets.

In the present paper, we study the impact of incentives provided to GPs for containing levels of hospital expenditure in the Italian region of Emilia Romagna. The main purpose of these P4P programs is cost containment rather than improvement in the quality of care, and they are grounded in the belief that capitation can result in suboptimal professional performance when GPs see patients at the ambulatory level. This may result in referral to secondary-care facilities even if patients' conditions do not strictly require hospitalisation. As long as additional rewards increase GPs' efforts, one would expect a reduction in the utilisation of hospital services, especially for conditions that can be effectively treated in a primary care setting.

In Emilia Romagna, primary care incentives are managed independently by each district, which can decide whether to use such incentives, which clinical areas or targets to prioritise, and how much funding to give to each area. As a matter of fact, their use varies greatly among districts but once the Health District (HD) and the GPs' organisations agree upon a particular program, all GPs working in a district become eligible. This institutional feature ensures that our empirical analysis does not suffer of individual voluntary selection into the programs.

At the start of our observation period, we document a number of ongoing initiatives designed to reward GPs for containing the recourse to hospital services by their listed patients. The incentives in question constitute only a fraction of the additional payments payable to GPs, since different objectives (e.g. the improved management of chronic illness) are in many cases incentivised by means of other programs. Following the pioneering period when P4P programs were first introduced, our data span a subsequent period characterised by the reorganisation of some of these schemes. One of the reasons for this reorganisation was the increasing scepticism regarding the effectiveness of programs specifically designed to contain hospitals' spending. Such scepticism led certain HDs to end these programs during the period of observation. Others did likewise in later years, and no such program is ongoing at present. Unfortunately, we only have a full dataset for a limited number of years, during which such programs were ended in a subset of HDs. Among other things, these programs were ended due to the fact that focusing on the financial implications of service utilisation, rather than on health improvements and disease management, was seen to generate negative feedback on GPs' motivations and involvement, which could have undermined the effectiveness of the programs themselves. The present work exploits this policy change in order to evaluate the implications of the removal of financial incentives to containing hospitalisations in Emilia Romagna during the period 2002-04. We use this policy shift as a natural experiment in order to empirically assess the influence of financial incentives on GPs' behaviour and to provide useful policy indications for the design of payment schemes.

Our study draws on administrative data linking GPs' characteristics (including all sources of professional income) with the utilisation of hospital services by registered patients. This enables us to establish whether programs for reducing hospitalisation levels were in fact ongoing in a given district and year. Consequently, we identify all those GPs within our sample who were eligible for an incentive program aimed at containing hospital expenditure. We use a difference-in-difference (DID) approach in order to remove potential sources of bias when assessing the influence of P4P programs on physicians' activities.

Two specific features distinguish our case from standard natural experiments recently used to investigate the responses of primary care providers to changes in the incentive system (e.g. Nolan 2008, Dumont et al. 2008, Layte et al. 2009). Firstly, instead of the *introduction* of an incentive-based program, we consider the consequences of its *closure*. Behavioural evidence, supported by lab and field experiments, suggests that penalties and premiums may not always affect individual choices symmetrically (Gneezy and Rustichini 2000). Within our specific framework, non-symmetric reactions to the introduction/removal of incentives may stem from clinical practices that are costly to change once they have been adopted. Therefore, removing financial incentives may not necessarily see physicians return to previous styles of practice. A less optimistic hypothesis suggests that if the use of financial incentives crowds out intrinsic motivations, their subsequent removal could further reduce intrinsic motivations, causing a reduction in effort and productivity below the pre-existing level (Camerer, 2010). Nowadays, many programs have been operating for several years, and this raises questions about how existing schemes can be modified in order to improve their effectiveness and to bring them in line with new policy priorities. Consequently, gaining an insight into the specific implications of removing incentive-based programs - or parts thereof - is of particular interest to policymakers.

The second distinguishing feature of our analysis is that while the treatment group is exclusively identified with those GPs working in districts where the program had been closed (“*Leavers*”), two alternative control groups are taken into consideration. The first of these consists of GPs working in districts that provide the aforesaid incentives for the entire period (“*Stayers*”), while the second includes GPs working in districts that did not introduce any specific program for containing hospitalisations (“*Non Participants*”). In principle, it is always possible that the units characterised by the same initial conditions as the treated ones, representing the natural control group (*Stayers*), may not fully control for unobservable factors. Given this possibility, the addition of a second control group, which is usually not available in policy evaluation studies, can reinforce the robustness of empirical analysis, provided that the two comparisons lead to similar conclusions.

2. Background Literature

Due to the fact that financial incentives in primary care can improve practice behaviour, a variety of pay-for-performance programs (P4P) have been established worldwide [e.g. Rosenthal et al. 2005; Mullen et al. 2010; Lester et al., 2010 for the US; Li et al., 2011 for Canada; Scott et al., 2009 for Australia]. In the British NHS, studies have focused on the UK’s Quality and Outcomes Framework, in order to assess trends in quality indicators before and

after the introduction of this program [see Campbell et al. 2008, Gravelle, Sutton and Ma 2010; Sutton et al. 2010; Dusheiko et al., 2011]. Overall, such P4P programs have led to a broad debate over the impact of existing schemes and possible design improvements [Christianson et al., 2008; Nolan et al., 2011; Cromwell et al., 2011; Scott et al., 2011].

To the extent that GPs operate in a setting characterised by incomplete contracts and benefit of informational advantages, it has been recognised that remuneration schemes can affect GPs' behaviour towards alignment with the general policy goals set by the healthcare authorities [Dumont et al., 2008]. However, others have countered that the presence of strong medical ethics may dilute the influence of incentive programs [Dixit, 2002; Benabou and Tirole, 2003; Siciliani, 2009].

Economic theory has outlined the strengths and weaknesses of different schemes, but the question of their effectiveness ultimately remains an empirical one. Until now, the evidence has been far from conclusive, and shows a conflicting evidence on the impact of P4P on different measures of the quality of care [Scott et al, 2011; Nolan et al, 2011]. This may be the consequence of poor policy design, but also of the lack of sufficiently long and detailed datasets required to identify genuine causal relations. Additional challenges stem from the organisation of general practices, where many confounding factors may influence physician's response, including inter- and intra- country differences in remuneration systems, practice arrangements, non-financial incentives and case-mix [Boyden and Carter, 2000].

Actually, P4P schemes differ according to a variety of factors, including the identified targets, the monitoring of activities and the size of financial transfers. Moreover, one should also take into account non-pecuniary factors such as professional autonomy, altruistic concerns regarding patients' health, and other demographic, socioeconomic and organisational characteristics [Chaix-Couturier et al. 2000].

Most programs are designed to improve the quality of care, but some have also been used to contain health expenditure. Among the studies that investigate the question of whether P4P in primary care can influence hospital use, Lee et al. (2010) find in Taiwan a significant reduction in inpatient admissions and diabetes-related hospital expenditure for patients whose physicians were enrolled in a P4P disease management program; similar results are obtained by Chen et al. (2010) in Hawaii, whereas Mullen et al.(2010) in California show no significant change in avoidable admissions following the introduction of an incentive program covering a range of clinical quality measures. Dusheiko et al. (2011) analyse cross-sectional and panel data to examine whether the improved management of ten chronic diseases in primary care lead to any

reduction in hospital costs in the UK. They find a significant drop in total hospital expenditure only for practice stroke care.

As for the policy measures specifically analysed in this paper, there has been very little empirical evidence regarding the effects of removing financial incentives from physicians' payment schemes. One exception is the study by Lester et al. (2010) on the effects of removal of a number of financial incentives offered to medical facilities, rather than to individual physicians, in the HMO Kaiser Permanente in California. In this case, the closure of certain programs was not expected to modify the total funding available for the facilities in question. Results show that removing a set of incentives results in a reduction in performance levels compared to those reached when operating under the incentive scheme.

Finally, the present work ought to be compared to a number of related papers assessing the role of financial incentives within the same institutional context considered here. Lippi Bruni et al. (2009) and Iezzi et al. (2011) both study the impact of diabetes management incentives in Emilia Romagna offered for programs promoting the assumption of responsibility of chronic patients by GPs. Both works outline a significant reduction in avoidable hospitalisations for the target population. Fiorentini et al. (2011) do not focus on a specific set of incentives as we do here, but analyse the overall impact of the total extra payments received by GPs on hospital referral patterns. They consider several indicators of appropriateness of care, and the only significant reduction in the use of hospital services is that recorded with regard to 27 medical DRGs that the Emilia Romagna region identifies as at risk of inappropriateness in primary care.

3. Primary care and hospital expenditures in Emilia-Romagna

In 1978, Italy set up its National Health System (NHS), a nationwide public healthcare system in which Regional Governments have progressively expanded their powers regarding healthcare service provision. More recently, fiscal decentralization has increased tax autonomy, whereby Regional Governments are now more financially accountable than in the past with regard to health spending, with substantial implications in terms of interregional redistribution (Ferrario and Zanardi, 2011).

Nationwide, primary care is organized according to a single payer, list-based system where family physicians do not to face multiple insurers or organizations but are directly contracted with the NHS. Consultations are free of charge and citizens must register with a GP who regulates access to public specialist and hospital care. The maximum number of listed patients each GP may have is 1,500, in the case of full-time, and 800 in the case of part-time physicians,

although exceptions are allowed for those practitioners exceeding the aforesaid threshold at the time of its introduction.

Healthcare Districts (HDs), the smallest institutional unit within the NHS, are responsible for coordinating primary and outpatient care. HDs are aggregated within Local Health Authorities (LHAs), whose managers are directly appointed by Regional Government. LHAs and HDs are granted a large degree of autonomy when it comes to drafting primary care policies, including the possible introduction of financial incentives. Over the last decade there has been a generalised shift from hospital to community care, in recognition of the fact that certain conditions, corresponding to the principal chronic illnesses, do not always require hospitalisation. Although all patients are registered with a specific GP, LHA and HDs have also promoted measures favouring arrangements among family physicians in order to create networks for the sharing of both facilities and knowledge (Fattore et al. 2009).

The first, and most important, part of a GP's remuneration is represented by capitation, as negotiated between national government and the physicians' organisations. A second, smaller component is the variable part (fee-for-service) awarded for specific forms of treatment provided at practice level, including minor surgery and immunisation up-take for selected groups of patients. Both components are uniform throughout the country. Following regional- and district-level agreements, a third additional block can be introduced. This consists of financial incentives aimed at promoting specific goals established by regional and district health authorities. Such additional payments top up capitation often following a P4P scheme, and may vary considerably both between, and within, regions. These incentives are designed in particular to encourage cooperation between GPs and other public healthcare providers, to promote a better quality of care, to contain costs and to reduce the inappropriate use of hospital resources.

As a consequence of those objectives pursued locally, HD agreements may include various remunerated activities, as well as payments of a varying entity. Interestingly, once the agreement has been signed, all GPs are eligible for the incentives without any individual voluntary selection into the program. In some cases, targeted activities refer to conditions for which GPs are capable of influencing the quality, and the appropriate setting, of care. For example, some local contracts provide financial incentives in order to increase the number of protected hospital discharges of patients needing follow-up care, to strengthen home and community care for the elderly, to reward the direct provision of certain forms of treatment (e.g. immunisation uptake), to promote GPs' assumption of responsibility for chronic patients (e.g. diabetes, hypertension), and to encourage the adoption of organisational routines designed to

improve cooperation among providers, such as participation in medical networks, or adherence to evidence-based guidelines.

A number of other programs have also addressed expenditure targets. In this paper we actually consider the effects of contracts rewarding the containment of hospital expenditure, as implemented in certain regional districts and later removed by a subset of the latter.

4. Data and estimation issues

4.1 The data

Our datasets draw on the population of Emilia Romagna (Italy), with a total population of 4.5 million. The study population consists of all regional citizens between 18-65 years of age, observed in the period 2002-2004. The resulting dataset includes 2,936,834 patients, 3,229 GPs and 39 districts belonging to 11 LHAs. During the period the average number of GPs active each year amounts to 3,187 (std. dev. 58).

The 39 HDs can be divided into three groups: the 8 districts providing incentives for the entire period (“*Stayers*”); the 12 districts that stopped the program during the observational period (“*Leavers*”); the 19 districts that never introduced a specific program for containing hospitalisations (“*Non Participants*”). Our estimation strategy is based on a comparison between “*Leavers*”, “*Stayers*” and “*Non Participants*”. The first is our treatment group, while the second and third are alternative control groups.

Since our aim is to study the links between financial incentives in primary care and the use of hospital services by listed patients, we have conducted our analysis mainly on the basis of episodes the occurrence of which is expected to be influenced by quality of primary care. For this purpose, we have constructed a measure for the utilisation of hospital services that could have been avoided. Avoidable admissions are identified by adopting the list of Ambulatory Care Sensitive Conditions (ACSCs) coded using the International Classification of Diseases (ICD-9-CM) and developed by Billings et al. (1993) and Caminal et al. (2004), which has recently also been used in empirical studies that evaluate how primary care policies affect the use of hospital facilities [Nolan, 2011; Fiorentini et al. 2011]. We classify hospitalisations as inappropriate if at least one of the ICD-9-CM codes referring to ACSCs is recorded as the primary reason for admission. Table 1 shows all ACSCs and the associated ICD-9-CM codes. We focus primarily on ACSCs, since they represent the subgroup of clinical conditions that are more likely to be affected by GPs’ behaviour. On the contrary, for most other conditions, hospital admissions are deemed to be beyond the control of general practitioners. By pooling together those conditions

under and beyond the control of the family doctor, one may excessively dilute the effect of policies designed to influence practice style, which tend to induce changes in hospital admissions for a limited subset of conditions only, namely for those conditions or cases which can be treated effectively treated in a general practice setting.

Ambulatory Care Sensitive Condition ICD-9-CM Codes

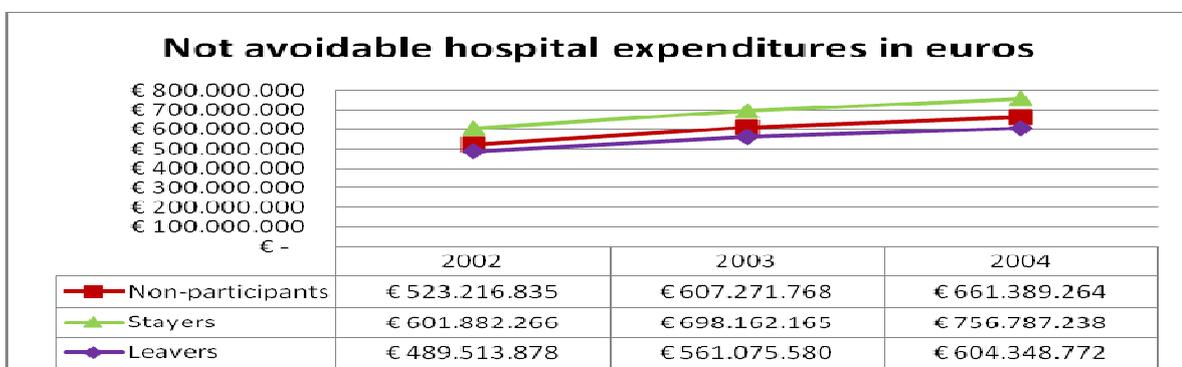
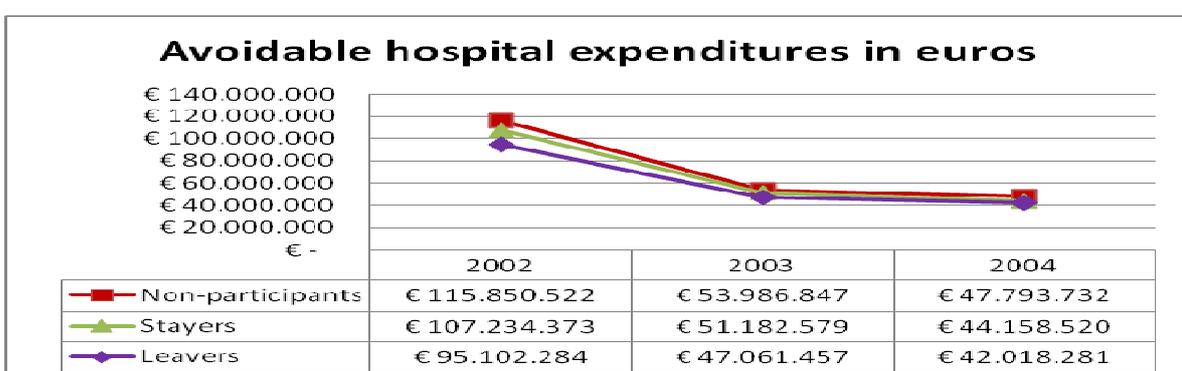
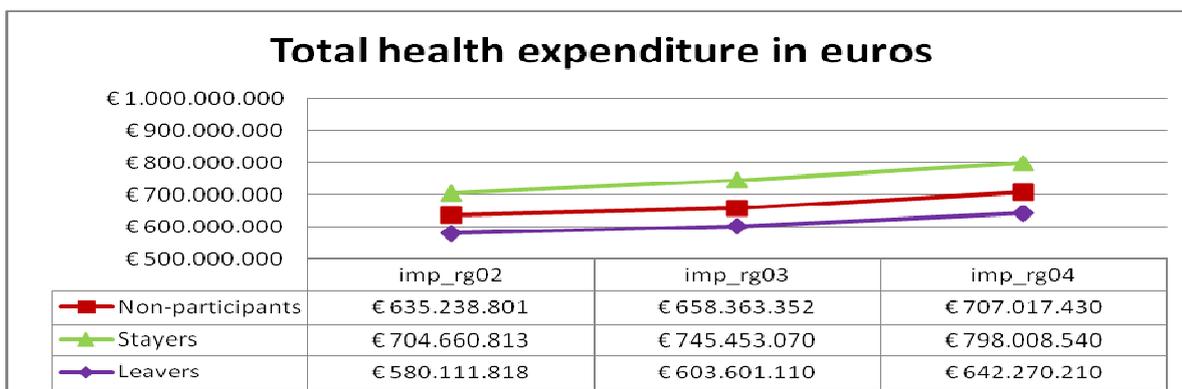
| | |
|--|---|
| Angina | 411.1, 411.8, 413. Excludes cases with procedure codes [01-86.99] |
| Asthma | 493 |
| Bacterial pneumonia | 481, 482.2, 482.3, 482.9, 483, 485, 486. Excludes cases with secondary diagnosis of sickle cell [282.6]. |
| Cellulites | 681, 682, 683, 686. Excludes cases with any procedure codes except 860 where it is the only procedure |
| Chronic obstructive pulmonary disease | 491, 492, 494, 496, 466.0 |
| Congestive heart failure | 428, 402.01, 402.11, 402.91, 518.4 |
| Dehydration - volume depletion | 276.5 |
| Diabetes | 250.1, 250.2, 250.3, 250.8, 250.9, 250.0, 251 |
| Gangrene | 785.4 |
| Gastroenteritis | 558.9 |
| Grand mal status and other epileptic convulsions | 345, 780.3 |
| Hypertension | 401.0, 401.9, 402.00, 402.10, 402.90. Excludes cases with procedures 36.01, 36.02, 36.05, 36.1, 37.5, 37.7. |
| Hypoglycemia | 251.2 |
| Hypokalemia | 276.8 |
| Immunization-related and preventable conditions | 032, 033, 037, 045, 055, 072, 320.0, 390, 391 |
| Kidney/urinary infection | 590, 599.0, 599.9 |
| Pelvic inflammatory disease | 614 (Excludes 68.3-68.8) |
| Peptic ulcer | [531, 532, 533] |
| Pulmonary tuberculosis and other tuberculosis | 011, 012-018 |
| Pyelonephritis | 590 |
| Ruptured appendix | 540.0, 540.1 |
| Severe ear, nose, and throat infections | 382, 462, 463, 464, 465, 472.1 |
| Skin grafts with cellulitis | DRG 263, DRG 264 |

Table 1

The dependent variable that proxies utilisation of hospital resources, is the log of expenditure calculated from the diagnostic group assigned to each (avoidable) hospital admission, and the corresponding DRG tariff set by the Regional Health Authority, which remains fixed for the entire period (*log of avoidable hospital expenditure*).

Figure 1 shows trends in total hospital expenditure, together with its disaggregation into avoidable and unavoidable expenditure. The validity of the DID estimator relies on the assumption that the trend in the dependent variable is the same for both treatment and control groups. Inspection of Figure 1 confirms that for our data the assumption of common trends is reasonable.

Figure 1 – Trend of total hospital expenditure and its disaggregation in reasonably avoidable hospital expenditure and not avoidable hospital expenditure, year 2002-2004.



The policy focus of the paper is on *financial incentives* for the containment of hospitalisations. For this purpose, we have identified, the physicians working in districts that in 2002 and/or 2004 provide incentives of the kind mentioned above. We should bear in mind that in Emilia Romagna, incentive-based programs are designed at the HD level, and all GPs operating within a particular district are eligible for the additional bonuses. Of course, for programs where payment is conditional on performance, some GPs may receive the bonus while others may not according to their results even if they work in the same district. Given this, using payments at the individual level – or even payment size - to indentify the impact of financial incentive on GPs activity, can be misleading due to the reverse causality effect. In fact, a positive bonus paid to a particular GP is the consequence of that GP's meeting the predefined target, but we cannot infer from this whether the existence of the program caused any change in his behaviour.

To address this problem, we utilise observed payments in order to single out those districts that adopted P4P programs for containing hospital expenditure, in 2002 and 2004. This allows us to identify if a GP operates under an incentive scheme or not, irrespectively of whether he was able to meet the agreed target. By doing so, we can ultimately separate treatment (*Leavers*) from control groups (*Stayers* and *Non Participants*).

Figure 2 shows incentives for containing hospital expenditure, by group and by year. We present group averages calculated as a share of GPs' annual income paid by the Regional Health Authority. Figures 3 and 4 show the distribution of financial incentives and the fraction of GPs receiving financial incentives by districts, distinguishing between *Stayers* and *Leavers* for 2002 and 2004.

Figure 2 – Financial incentives as a percentage of GP's annual income by groups, 2002-2004.

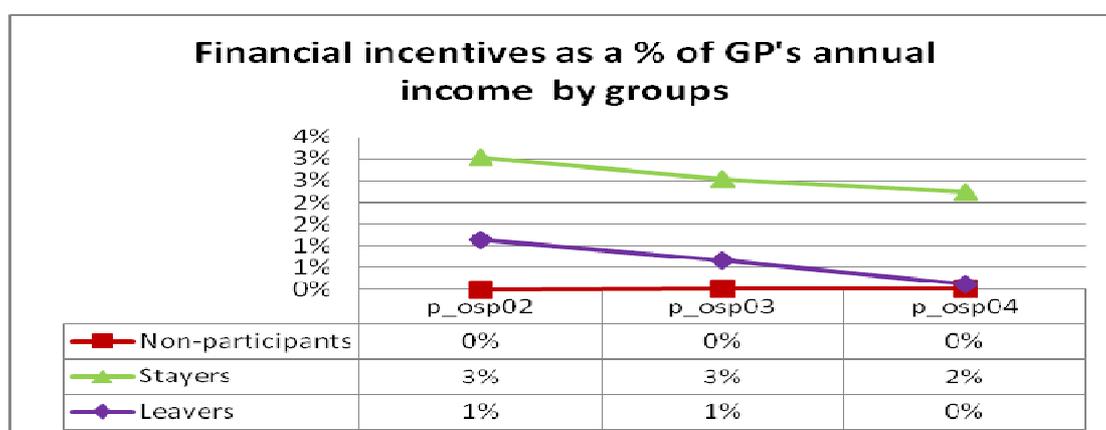


Figure 3 – Financial incentives as a percentage of GP annual income by districts (Stayers and Leavers), 2002-2004.

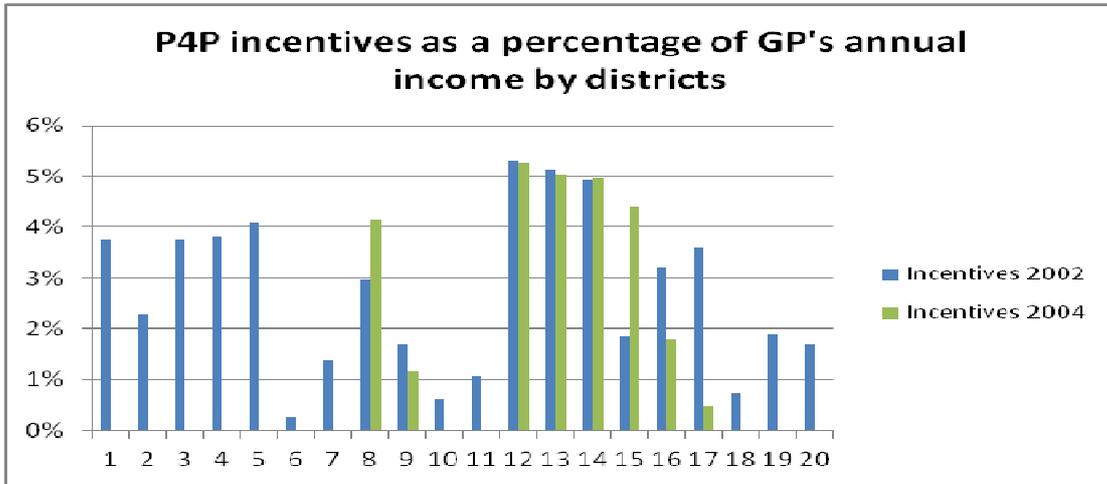
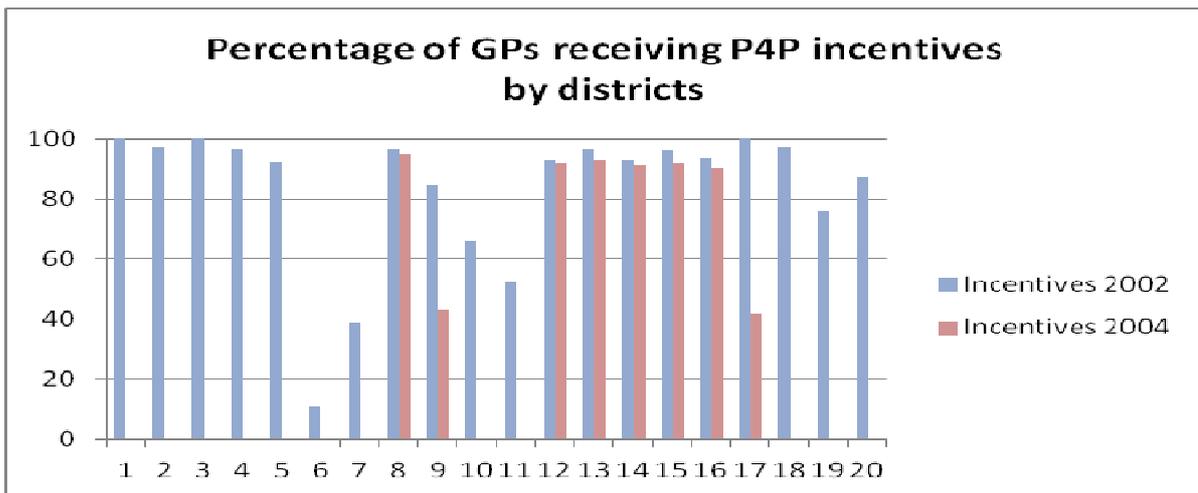


Figure 4 – Percentage of GPs receiving financial incentives by districts (Stayers and Leavers), 2002-2004.



We model the *log of avoidable hospital expenditure* as dependent upon a set of controls regarding the characteristics of both the physician and the practice. The first set of covariates consist of the GP's gender, age, age squared and seniority. We also control for the type of practice, distinguishing between individual and group practices.

To the extent that patients may choose the physician they wish, there is a possibility of some degree of self-selection, by patients with similar characteristics (e.g. a particular chronic disease) in the same list. Still, self-selection is not expected to be correlated with eligibility for financial incentives, since the choice of physician usually involves those GPs operating in the same area, whereas eligibility only varies across districts. In order to control for selection on

relevant, observable list characteristics, we include a set of patients' characteristics extracted from the list such as: the share of male patients; the average age of listed patients; and the average Charlson index calculated from hospitalised patients. The Charlson index is a weighted index of comorbidity, computed from the ICD-9-CM diagnostic and procedure codes available in administrative datasets, according to their potential influence on the risk of mortality (Charlson et al., 1987; Romano et al., 1993). By doing so, we address potential biases due to (observable) patients' characteristics which may affect hospital admission rates.

In order to account for supply-side characteristics, we have included the number of *ordinary hospital beds* per 1,000 inhabitants. In addition, as an indicator of market structure, we construct a *competition index*, calculated as the number of GPs per 1,000 inhabitants (Kann, Biørn, Luras, 2010), and an index of *population dependency* on district hospitals, while geographical accessibility is proxied by *population density*. All these controls are entered at the district level. Table 2 presents the descriptive statistics for our estimating sample. We observe only slight differences in average values between the treatment and control groups and this supports the validity of the identification strategy adopted.

| <i>Variable</i> | <i>Treatment group (Leavers)</i> | | <i>1 control group (Stayers)</i> | | <i>2 control group (Non participant)</i> | |
|--------------------------------------|----------------------------------|-----------|----------------------------------|-----------|--|-----------|
| | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> | <i>Mean</i> | <i>SD</i> |
| GP gender (Male=1) | 74% | 0.439 | 71% | 0.455 | 74% | 0.436 |
| GP age | 52 | 6.348 | 51 | 5.852 | 51 | 6.623 |
| GP age squared | 2701 | 701.870 | 2615 | 630.456 | 2630 | 716.554 |
| GP seniority | 17 | 7.702 | 18 | 7.763 | 17 | 7.615 |
| Practice type (associated=1) | 50% | 0.500 | 53% | 0.499 | 57% | 0.495 |
| List size | 1156 | 495.153 | 1119 | 463.587 | 1163 | 506.661 |
| List proportion male | 48% | 0.038 | 48% | 0.036 | 48% | 0.036 |
| Average patient age | 49 | 4.959 | 49 | 4.546 | 47 | 5.363 |
| Charlson index | 0.615 | 0.718 | 0.526 | 0.377 | 0.570 | 0.472 |
| Hospital beds by population | 0.0002 | 0.000 | 0.0002 | 0.000 | 0.0002 | 0.000 |
| District population density | 2.044 | 2.181 | 0.352 | 0.441 | 0.457 | 1.227 |
| Competition index | 0.476 | 0.277 | 0.381 | 0.187 | 0.210 | 0.129 |
| Pop_ dependency to district hospital | 77.213 | 10.932 | 80.971 | 2.677 | 80.575 | 8.158 |

Table 2 - **Descriptive Statistics: Treatment group and control groups, year 2002-2004**

3.2 Statistical analysis

Our aim is to estimate the effect of P4P incentives on (avoidable) hospital expenditure, by comparing variations across time between the treatment group and each control group. More precisely, we look at expenditure for avoidable hospital treatments (before and after) the ending

of incentive programs in the 12 districts that experienced such policy change. We specify the following equation:

$$y_{ijt}^h = \beta_0 + \beta_1' X_{it} + \beta_2' Z_{jt} + \gamma T_t + \rho D^h + \delta T_t \cdot D^h + \mu_{ijt}^h \quad (1)$$

Districts ($j= 1, \dots, 39$) are grouped according to the presence of financial incentives that are aimed at containing hospital expenditure during the period of investigation. Given the feature of our dataset, they are aggregated into three groups: *Leavers*, *Stayers* and *Non Participants* ($h= L, S, NP$). The dependent variable y_{ijt}^h is the log of (avoidable) hospital expenditure of physician i in year t , operating in district j which is included in group h . Vector X_{it} contains covariates referring to physician i and to his list at time t ; vector Z_{jt} includes covariates referring to the district where each GP operates at time t ; T_t is a dummy equal to 1 for observations in 2004, and equal to 0 otherwise (2002); D^h is a dummy equal to 1 if the physician belongs to the treatment group, or 0 otherwise; $T_t * D^h$ is the interaction term that takes value 1 if physician i is eligible for the incentives at time t . The coefficients β_1 and β_2 measure the marginal effects of changes in GP and district variables on the log of hospital expenditure. The coefficients γ and ρ estimate the average impact on the log of expenditures of the observation being included the post treatment period and in the treatment group, respectively. The coefficient δ captures the most relevant policy indicator in our analysis because it measures the effect of removing the incentives during the second period. It multiplies the interaction term and functions as a dummy variable equal to 1 for observations jointly belonging to the treatment group and the second period; μ_{ijt}^h is the idiosyncratic error term. Equation (1) is estimated by a pooled linear panel data model.

We estimate hospital expenditure for the three groups: districts that provide incentives for the entire period (y_{ijt}^S for “*Stayers*”); districts that ended the program (y_{ijt}^L for “*Leavers*”); districts that never introduced any specific program for containing hospitalisations (y_{ijt}^{NP} for “*Non Participants*”). The expected difference in hospital expenditure changes between *Leavers* and *Stayers* is:

$$DD^{L-S} = E(\Delta y_{ijt}^L) - E(\Delta y_{ijt}^S), \quad (2)$$

while the expected difference between *Leavers* and *Non-Participants* is:

$$DD^{L-NP} = E(\Delta y_{ijt}^L) - E(\Delta y_{ijt}^{NP}). \quad (3)$$

We estimate an empty and full multivariate specification of (2) and (3). To mitigate the over-rejection problem for DID estimates when the inference of the regular *t*-statistic is based on unadjusted standard errors, we use robust standard errors clustered at district level [Bertrand, Duflo and Mullainathan, 2004; Li et al. 2011].

5. Results

Table 3 shows estimates for the DID specification with the log of avoidable hospital expenditures as the dependent variable. Firstly, we compared *Leavers* to *Stayers*, and then *Leavers* to *Non Participants*, and in each case we consider two specifications. The first one consists of the empty model comparing average group expenditure, while the second one is the multivariate specification obtained by including the controls presented in Table 2. Adding covariates improves the precision of the estimates, but results are in line with the empty model.

Table 3 – Difference in difference results for avoidable hospital expenditure

| | <i>Treatment group Leavers</i> | | <i>Control group Stayers</i> | | <i>Between group difference</i> | <i>Treatment group Leavers</i> | | <i>control group Non Participant</i> | | <i>Between group difference</i> |
|--------------------|--------------------------------|-----------|------------------------------|-----------|---------------------------------|--------------------------------|-----------|--------------------------------------|-----------|---------------------------------|
| | <i>Diff</i> | <i>SD</i> | <i>Diff</i> | <i>SD</i> | | <i>Diff</i> | <i>SD</i> | <i>Diff</i> | <i>SD</i> | |
| Empty model | | | | | | | | | | |
| 2002 | 11,078 | (0.104) | 11,016 | (0.089) | 0.062 (0.137) | 11,078 | (0.103) | 11,245 | (0.079) | -0.167 (0.129) |
| 2004 | 10,099 | (0.145) | 10,066 | (0.107) | 0.033 (0.180) | 10,099 | (0.144) | 10,307 | (0.059) | -0.208 (0.156) |
| Diff-in-diff | -0.029 (0.140) | | | | | -0.040 (0.106) | | | | |
| R ² | 0.08611 | | | | | 0.08710 | | | | |
| Full model | | | | | | | | | | |
| 2002 | 3,995 | (2.008) | 3,937 | (1.996) | 0.058 (0.120) | 1,997 | (1.657) | 2,275 | (1.586) | -0.278 (0.115) |
| 2004 | 3,019 | (1.999) | 2,888 | (2.021) | 0.132 (0.074) | 0,997 | (1.654) | 1,344 | (1.611) | -0.347 (0.121) |
| Diff-in-diff | 0.074 (0.148) | | | | | -0.069 (0.102) | | | | |
| R ² | 0.47550 | | | | | 0.50954 | | | | |
| | Coeff | SD | pvalue | | | Coeff | SD | pvalue | | |
| GP gender | -0.194 | 0.083 | 0.029 | | | -0.198 | 0.087 | 0.031 | | |
| GP age | -0.022 | 0.055 | 0.701 | | | 0.050 | 0.051 | 0.340 | | |
| GP age squared | -0.000 | 0.000 | 0.879 | | | -0.001 | 0.000 | 0.186 | | |
| GP seniority | 0.011 | 0.006 | 0.105 | | | 0.005 | 0.004 | 0.276 | | |
| Practice type | -0.046 | 0.045 | 0.318 | | | -0.012 | 0.042 | 0.778 | | |
| List size | 0.001 | 0.000 | 0.000 | | | 0.001 | 0.000 | 0.000 | | |
| Proportion male | 2.528 | 1.298 | 0.066 | | | 2.168 | 1.177 | 0.076 | | |
| Av_patient age | 0.120 | 0.009 | 0.000 | | | 0.122 | 0.009 | 0.000 | | |
| Charlson index | 0.008 | 0.064 | 0.905 | | | 0.024 | 0.057 | 0.682 | | |
| Hospital beds | -0.001 | 0.000 | 0.031 | | | -0.000 | 0.000 | 0.280 | | |
| Dist pop_density | -0.009 | 0.051 | 0.867 | | | -0.032 | 0.032 | 0.325 | | |
| Competition index | 0.157 | 0.254 | 0.545 | | | 0.234 | 0.192 | 0.235 | | |
| Pop_dependency | -0.004 | 0.009 | 0.674 | | | -0.005 | 0.005 | 0.331 | | |

Standard errors are reported in parenthesis.

We start by comparing *Leavers* and *Stayers*. The empty model indicates that both groups reduce avoidable expenditures over time. The reduction in the dependent variable amounts to 9.7% for *Leavers* and to 8.6% for *Stayers*. In the multivariate model, the within-group difference displays similar trends over time although the estimated percentage reduction is larger for *Stayers*. In both specifications, the DID is not significant, however, and thus the removal of financial incentives does not seem to have impacted on the *Leavers*' behaviour differently from what has happened to GPs working in districts where the incentives were maintained.

The comparison between *Leavers* and *Non-Participants* confirms the general trend of falling expenditure on avoidable hospitalisations. In the empty model, the within-group time difference shows that *Non-Participant* districts reduce the resources spent on ACSCs hospitalisations by 8.3%, whereas this reduction was slightly larger among *Leavers* (9.7%). If we consider between-group differences, *Non-Participant* districts display a higher expenditure level in both years and the gap slightly increases over time. Similar findings hold for the multivariate model. Consistently with the previous case, the DID estimation shows that removing financial incentives has not significantly changed the difference between treated observations and controls.

Overall, the impact produced by the ending of P4P programs, does not display any significant differential effect between the districts affected by the change in the incentive programs and any of the control group considered. We observe a generalised decreasing trend in avoidable expenditures but such variations are similar between groups despite the change in the incentive structure over time. Such evidence indicates that the programs that were removed do not seem to have effectively influenced physicians' behaviour.

However, referrals for ACSCs represent only a small fraction of total hospital admissions (Figure 1), and HDs usually set targets according to overall use of hospital services, without distinguishing between avoidable and unavoidable cases. Therefore, as a robustness check, we have re-estimated the model by including expenditure for all hospitalizations as a dependent variable and results are reported in table 4.

Table 4 – Difference in difference results for total hospital expenditure

| | <i>Treatment group Leavers</i> | | <i>Control group Stayers</i> | | <i>Between group difference</i> | <i>Treatment group Leavers</i> | | <i>Control group Non participant</i> | | <i>Between group difference</i> |
|--------------------|--------------------------------|-----------|------------------------------|-----------|---------------------------------|--------------------------------|-----------|--------------------------------------|-----------|---------------------------------|
| | <i>Diff</i> | <i>SD</i> | <i>Diff</i> | <i>SD</i> | | <i>Diff</i> | <i>SD</i> | <i>Diff</i> | <i>SD</i> | |
| Empty model | | | | | | | | | | |
| 2002 | 13.144 | (0.020) | 13.192 | (0.028) | -0.048 (0.035) | 13.144 | (0.036) | 13.153 | (0.028) | -0.009 (0.046) |
| 2004 | 13.173 | (0.020) | 13.286 | (0.028) | -0.113 (0.034) | 13.173 | (0.040) | 13.161 | (0.030) | 0.012 (0.050) |
| Diff-in-diff | -0.065 (0.049) | | | | | 0.021 (0.025) | | | | |
| R ² | 0.00461 | | | | | 0.00023 | | | | |
| Full model | | | | | | | | | | |
| 2002 | 8.032 | (0.335) | 8.054 | (0.312) | -0.023 (0.041) | 8.653 | (0.463) | 8.692 | (0.454) | -0.039 (0.032) |
| 2004 | 8.070 | (0.338) | 8.122 | (0.318) | -0.051 (0.032) | 8.689 | (0.466) | 8.723 | (0.455) | -0.034 (0.032) |
| Diff-in-diff | -0.029 (0.018) | | | | | 0.005 (0.016) | | | | |
| R ² | 0.84489 | | | | | 0.84593 | | | | |
| | Coeff | SD | pvalue | | | Coeff | SD | pvalue | | |
| GP gender | -0.087 | 0.019 | 0.000 | | | -0.072 | 0.024 | 0.006 | | |
| GP age | 0.035 | 0.012 | 0.010 | | | 0.018 | 0.015 | 0.230 | | |
| GP age squared | -0.000 | 0.000 | 0.004 | | | -0.000 | 0.000 | 0.085 | | |
| GP seniority | 0.000 | 0.001 | 0.727 | | | 0.001 | 0.001 | 0.346 | | |
| Practice type | -0.002 | 0.014 | 0.901 | | | 0.004 | 0.012 | 0.739 | | |
| List size | 0.001 | 0.000 | 0.000 | | | 0.001 | 0.000 | 0.000 | | |
| Proportion male | 0.684 | 0.309 | 0.039 | | | 0.373 | 0.320 | 0.254 | | |
| Av_ patient age | 0.056 | 0.002 | 0.000 | | | 0.058 | 0.003 | 0.000 | | |
| Charlson index | 0.027 | 0.015 | 0.086 | | | 0.034 | 0.015 | 0.030 | | |
| Hospital beds | -0.000 | 0.000 | 0.734 | | | 0.000 | 0.000 | 0.123 | | |
| Dist pop_ density | -0.012 | 0.006 | 0.051 | | | -0.022 | 0.008 | 0.012 | | |
| Competition index | -0.002 | 0.042 | 0.960 | | | 0.013 | 0.040 | 0.746 | | |
| Pop_dependency | 0.002 | 0.001 | 0.264 | | | -0.001 | 0.002 | 0.634 | | |

Standard errors are reported in parenthesis.

For both *Stayers* and *Leavers*, the within-group difference in the empty model reveals a slight positive trend in total hospital expenditure. Opposite time trends between total and avoidable expenditure indicate that the rate of inappropriate use of hospital resources has fallen over time. The between-group difference shows that in 2002 the *Leavers*' districts had a lower level of hospital expenditure than the *Stayers* did, and this difference increased over time: removing financial incentives widened the gap between the two groups, although in the opposite direction to what one would have expected, since the rate of growth in spending in the group that ends the incentives (*Leavers*) is lower than that of the group that was still granted additional remuneration for curbing the use of hospital services (*Stayers*). Nevertheless, once again we find that the effect is not significant. Introducing covariates to control for differences in observable characteristics between the two groups does no modify the results in a qualitative sense.

The comparison of *Leavers* with *Non-Participants* confirms the presence of a common upward trend in total expenditure over time., Shifting the focus from avoidable to overall hospitalisations confirms the non-significance of the effect of ending the incentive programs, in both the empty and full specifications. This is not surprising since the latter dependent variable is expected to be relatively less influenced by GPs' decisions than the former..

In conclusion, according to our DID estimates, the adoption of total hospital expenditure as a dependent variable does not modify our main policy conclusion, according to which financial incentives seem to exert no influence on GPs' behaviour. Considering a dependent variable that cover a larger set of conditions displays changes in the time trend of expenditures, but, once again, no significant difference is detected across groups.

6. Discussion and conclusions

The intense debate over the role of monetary incentives in improving the quality of primary care is a consequence of the growing number of P4P programs worldwide. Relatively less attention has been paid to the implications of programs explicitly designed to contain costs. Moreover, empirical studies have almost exclusively focused on physicians' responses following the *introduction* of incentives. Now that several programs have been operating for nearly a decade or even more, the need to improve policy design requires an insight into the effects of *removing* some programs, or parts thereof, as well. Such removal may be stimulated by changes in policy priorities, or by the belief that certain aspects of the programs are not very effective, or indeed may produce unintended, dysfunctional consequences.

We have analysed the impact of removing programs offering cash bonuses to GPs in exchange for containing hospital expenditure in Italy's Emilia Romagna region over the period 2002-04. We have estimated DID models comparing changes in hospital expenditure, both for ambulatory care-sensitive conditions and for overall hospitalisations. GPs operating in districts that removed their incentives during the period of observation constitute our treatment group. Changes in expenditure for their listed patients are compared with those for patients residing in districts that maintained their programs and also with expenditure on patients residing in districts where no similar program was ever introduced.

We show that the removal of incentives designed to contain hospitalisations did not produce any significant difference in performance between the treatment and control groups in terms of either of the dimensions considered here (avoidable and overall hospitalisations). The empirical evidence lends support to the decision of those DHAs that closed down their programs, as these

programs guaranteed monetary transfers to GPs without influencing their behaviour in the desired direction.

This finding is noteworthy especially when compared with the impact produced by other measures introduced in Emilia Romagna during that same period, which were expressly designed to improve the quality of service provision. For example, it has been shown [Lippi Bruni et al. 2009, Iezzi et al. 2011] that incentives rewarding the assumption of responsibility (and a set of related tasks) for patients affected by chronic diseases such as type *II* diabetes, ultimately reduced the probability of recourse to hospital care by the target population for avoidable conditions. As has been seen in other countries (Li et al., 2011, Mullen et al. 2010), our findings suggest that GPs react to certain incentives, but not to others. One possible lesson that can be drawn here with regard to future policy design, is that incentives explicitly designed to contain costs prove to be less effective than those promoting improvements in care delivery by rewarding additional patient follow-up. This result is consistent with the idea that the effectiveness of additional remuneration may suffer if such bonus payments, instead of reinforcing the physician-patient relationship, are perceived as shifting physicians' attention from patients' health towards the pursuit of general financial targets.

References

- Benabou, R., Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70, 489-520.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *Quarterly Journal of Economics*, 119 (1), 249-75.
- Billings, J., Zeitel, L., Lukomnik, J, Carey, TS, Blank, AE, & Newman, L. (1993). Impact of Socioeconomic Status On Hospital Use In New York City. *Health Affairs*, 12, 162-173.
- Boyden, A., & Carter, R. (2000). *The appropriate use of financial incentives to encourage preventive care in general practice*. Centre for Health Program Evaluation, Research Report 18.
- Camerer, C.F. (2010). Removing financial incentives demotivates the brain. *Proceedings of the National Academy of Sciences USA*, 107, 49, 20849-50.
- Caminal, J., Starfield B., Sanchez E., Casanova C., & Morales M. (2004). The role of primary care in preventing ambulatory care sensitive conditions. *European Journal of Public Health*, 14 (3), 246-51.
- Campbell, S.M., McDonald, R., & Lester, H. (2008). The experience of pay for performance in English family practice: a qualitative study. *Annals of Family Medicine*, 6, 228–234.
- Chaix-Couturier, C., Durand-Zaleski, I., Jolly, D., & Durieux, P. (2000). Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues. *International Journal for Quality in Health Care*, 12, 133-42.
- Charlson, M., Pompei, P., Ales, K., & McKenzie, C. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronical Disease*, 40, 373-83.
- Chen, J.Y., Tian, H., Taira, J.D., Hodges, K.A. et al. (2010). The effect of a PPO pay-for-performance program on patients with diabetes. *The American Journal of Managed Care*, 16, e11-9.
- Christianson, J.B., Leatherman, S., & Sutherland, K. (2008). Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. *Medical Care Research and Review*, 5, 5S-35S.
- Cromwell, J., Trisolini, M., Pope, G., Mitchell, J., & Greenwald, L. (2011). *Pay for Performance in Health Care: Methods and Approaches*, New York: RTI Press.

- Dixit, A. (2002). Incentives and organizations in the public sector: an interpretative review. *Journal of Human Resources*, 37, 696-727.
- Dumont, E., Fortin, B., Jacquemet, N., & Shearer, B. (2008). Physicians' multitasking and incentives: empirical evidence from a natural experiment. *Journal of Health Economics*, 27, 1436-50.
- Dusheiko, M., Gravelle, H., Martin, S., Rice, N., & Smith, P.C. (2011). Does better disease management in primary care reduce hospital costs? Evidence from English primary care. *Journal of Health Economics*, 30, 919-32.
- Fattore G., Forini F., Salvatore D. & Tozzi V. (2009). Social network analysis in primary care: the impact of interactions on prescribing behaviour. *Health Policy*, 92, 141-148.
- Ferrario, C., & Zanardi, A. (2011). What happens to interregional redistribution upon fiscal decentralization? Evidence from the Italian NHS. *Health Policy*, 100, 71-80
- Fiorentini, G., Iezzi, E., Lippi Bruni M., & Ugolini C. (2011). Incentives in primary care and their impact on potentially avoidable hospital admissions. *European Journal of Health Economics*, (12), 297-309.
- Gneezy U. & Rustichini A. (2000). A fine is a price. *Journal of Legal Studies*, 29, 1-17.
- Gravelle, H., Sutton, M., & Ma, A. (2010). Doctor behaviour under a pay for performance contract: treating, cheating and case finding? *Economic Journal*, 120, F129- F156.
- Grilli, R., Magrini, N., Penna, A., Mura, G, & Liberati, A. (2000). Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet*, 355 (9198), 103-06.
- Iezzi, E., Lippi Bruni, M., & Ugolini, C. (2011). The role of GP's compensation schemes in diabetes care: evidence from panel data, Department of Economics, University of Bologna, WP. n. 766.
- Iversen, T., Luras, I. (2006). Capitation and incentives in primary care. In Jones A.M. (ed.) *The Elgar Companion to Health Economics*, Edward Elgar, Cheltenham, UK.
- Kann, I.G., Biørn, E., & Luras, H. (2010), Competition in general practice: Prescription to the elderly in a list patient system. *Journal of Health Economics*, 29, 751-764.
- Layte, R., Nolan, A., McGee, H., O'Hanlon, A. (2009). Do consultation charges deter general practitioner use among older people? A natural experiment. *Social Science & Medicine*, 1432-1438.

- Lee, T., Cheng, S., Chen, C., & Lai, M. (2010). A pay-for-performance program for diabetes in Taiwan: a preliminary assessment. *American Journal of Managed Care*, 16, 5-9.
- Lester, H., Schmittiel, J., Selby, J. et al. (2010). The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *British Medical Journal*, 340, c1898.
- Li, J., Hurley, J., DeCicca, P., & Buckley, G. (2011). *Physician response to pay-for-performance: evidence from a natural experiment*, NBER WP 16909.
- Lippi Bruni, M., Nobile, L., & Ugolini C. (2009). Economic incentives in general practice: the impact of pay-for-participation and pay-for-compliance programs on diabetes care. *Health Policy*, 90 (2-3), 140-148.
- Mullen, K.J., Frank, R.G., & Rosenthal, M.B.(2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *Rand Journal of Economics*, 41 (1), 4-91.
- Nolan, A. (2008). Evaluating the eligibility for free care on the use of general practitioner (GP) services: a difference-in-difference matching approach. *Social Science & Medicine*, 67, 1164-1172.
- Nolan, A. (2011). An extension in eligibility for free primary care and avoidable hospitalisations: A natural experiment. *Social Science & Medicine*, 73, 978-985.
- Nolan, A., O'Reilly, J., Smith, S., & Brick, A. (2011). The potential role of Pay-for-performance in Irish Health Care. Renewal Series paper 4, ESRI.
- Romano, P.S., Roos, L.L., & Jollis, J.G. (1993). Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *Journal of Clinical Epidemiology*, 46, 1075-1079.
- Rosenthal, M.B., Frank, R.G., & Epstein, A.M. (2005). Early experience with pay-for-performance: from concept to practice. *Journal of the American Medical Association*, 294 (14), 1788-93.
- Scott, A., Schurer, S., Jensen, P.H., & Sivey, P. (2009). The effects of an incentive program on quality of care in diabetes management. *Health Economics*, 18, 1091-108.
- Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L. Et al. (2011). The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Review*, issue 9.

Siciliani, L. (2009). Paying for performance and motivation crowding out. *Economic Letters*, 103, 8-71.

Sutton, M., Elder, R., Guthrie, B., Watt, G. (2010). Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Economics*, 19, 1-13.



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
<http://www.dse.unibo.it>