Does Patient Health Behaviour respond to Doctor's Effort?

Eleonora Fichera, James Banks, Luigi Siciliani, Matt Sutton

No. 62 /17

BATH ECONOMICS RESEARCH PAPERS

Department of Economics

Department of Economics | UNIVERSITY OF BATH

Does Patient Health Behaviour respond to Doctor's Effort?

Eleonora Fichera James Banks Luigi Siciliani Matt Sutton *

March 7, 2017

Abstract

Incentive pay systems have been introduced in public sectors such as education and healthcare. In these organisations where the outcome (health or education) is a joint product between different agents, it is unclear what the effects of these incentives are onto the behaviour of untargeted agents. We focus on patient health as a joint product of patient effort, through lifestyle and behaviour, and doctor effort, through diagnosis and treatment. Patient response to doctor effort is *a priori* ambiguous and depends on the degree of complementarity or substitution between doctor and patient effort. We use data on the physical activity, drinking and smoking behaviours of over 2,000 patients aged over 50 with cardiovascular diseases in England. Through a new data linkage and an instrumental variable approach, we test whether changes in doctors' treatment efforts triggered by changes in their payment system between 2004 and 2006 had an impact on patient behaviour. Doctors working in primary care practices increased the proportion of patients with controlled disease from 76% to 83% in response to the payment change. Patients responded by reducing the frequency of drinking alcohol and their cigarette consumption. This suggests that patient efforts are complements to doctor effort. The results have implications for the effectiveness of pay-for-performance schemes which encourage higher doctor effort, and the design of such incentive schemes.

Keywords: Health behaviours; price change; instrumental variables.
JEL classification: I1

*Eleonora Fichera is at the Department of Economics, University of Bath; Matt Sutton is at the Manchester Centre for Health Economics, University of Manchester. James Banks is at the Institute for Fiscal Studies, London and Economics, University of Manchester. Luigi Siciliani is at the Department of Economics and Related Study at the University of York. Correspondence: 3 East, 2.11 University of Bath, Claverton Down Bath, BA2 7AY. Phone: +44 (0)1225 386021. Email: e.fichera@bath.ac.uk.
Acknowledgements: We are grateful for seed-corn funding from the Manchester Institute for Collaborative Research on Ageing (MICRA). Eleonora Fichera acknowledges financial support from the MRC Early Career Fellowship in Economics of Health (MR/K021583/1). She has written this paper when working at the Manchester Centre for Health Economics, University of Manchester. We are grateful to Tommy Allen for assistance with the linkage of ELSA to general practices. We thank participants in the Royal Economic Society conference in Manchester, the Econometrics of Ageing and Longevity organised session at the 10^{th} World Congress in Dublin, the 15^{th} European Health Economics Workshop in Lausanne and seminar participants at the University of Bristol. We also thank Kurt Brekke, Shiko Maruyama, Mario Pezzino and Marcos Vera-Hernández for their helpful comments. ELSA data were obtained from the UK Data Archive. ELSA was developed by a team of researchers from University College London, the Institute for Fiscal Studies and the National Centre for Social Research and funded by the National Institute of Aging and a consortium of UK government departments co-ordinated by the ESRC. Responsibility for interpretation of the data and any errors is the authors' alone.

1 Introduction

The use of performance incentives are common in the private sector, but recently explicit incentive contracts have been adopted in public sectors such as education and healthcare to name a few (Lavy (2009); Propper et al. (2010); Propper and Wilson (2012)). Most of this research has investigated the incentive effects on targeted agents (Prendergast (1999)) or teams (Burgess et al. (2010)). However, in complex organisations agents interact with each other and we might expect spillover effects of these incentives onto the actions of untargeted agents. One example of such complex organisations are schools where there are multiple stakeholders, teams and tasks. The outcome, education, is a joint product of teacher effort, through teaching style and lecture content, and pupil effort, through homework completion and participation to classes. In this context, incentives that aim to improve teacher performance might have spillovers onto pupil effort. Some studies have found that performance pay has negative spillovers onto untargeted agents and actually reduce pupil GPA and daily attendance rates, especially when performance pay is related to pupil feedback (Elberts et al. (2002)). Some other studies have found a positive association between teacher incentives and pupil performance (Glewwe et al. (2010); Lavy (2009)), but were unable to determine whether this was due to better schools selecting into incentive pay systems or teacher incentives exerting more effort from teachers (Figlio and Kenny (2007)).

This study focuses on an outcome measure, health, that is also a joint product between two agents, doctors and patients, where doctors provide medical treatment and patients choose their lifestyle behaviours. Our identification strategy exploits a natural experiment generated by the largest pay for performance incentive for English primary care practices. To motivate our empirical analysis, we first provide a stylised model of a patient-doctor interaction. We show that patient's response to doctor's effort depends on the degree of complementarity and substitution between doctor and patient efforts. We then investigate doctor response to a change in the payment rule. Whereas doctors always increase their effort in response to a price increase, patient response depends on the degree of complementarity or substitution between her effort and doctor effort. This has implications for the effectiveness of the scheme which encourage doctor's effort. If patient and doctor efforts are complements, by incentivising doctor's effort, the purchaser also incentivises patient's effort.

We estimate the effect of doctor's effort on patient's effort using a large sample of individuals diagnosed with cardiovascular conditions (CVDs) from the English Longitudinal Survey of Ageing (ELSA) between 2004 and 2006. Individuals with long-term conditions (i.e. diabetes, hypertension and heart disease) represent over 15 million people in the UK and are predicted to rise by over a third over the next ten years (Department of Health (2012)). Best practice management and prevention for these diseases suggest a combination of medical treatment and behaviour change interventions, through reduction of smoking and drinking, and increase in physical activity (NICE (2009); NICE (2014); NICE (2015)). A new strategy encouraging people to engage in their own health and healthcare was described by Wanless (2002) and the Department of Health (2013).

We are the first to empirically investigate how patient lifestyle behaviours respond to doctor effort and we do this by implementing a unique linkage between ELSA respondents and administrative data on primary care practice treatment rates. One difficulty in estimating this effect is the potential endogeneity and reverse causality between the doctor's decision to supply medical treatments and the doctor's expectation of the patient's health behaviours. We overcome this issue by means of a two-stage least squares (2SLS) approach, in which doctor's effort is instrumented with the price change induced by changes in a large pay-for-performance (P4P) scheme, the Quality and Outcomes Framework (QOF). This incentive system was introduced in the U.K. in 2004 to reward primary care practices for providing high-quality care in a number of disease areas. Specifically, our identification exploits the fact that between 2004 and 2006 primary care practices face different price incentives from the QOF, depending on their initial level of performance in 2004. We use this change in price schedule to determine the financial incentive that was generated by the QOF and we use this as an instrument for doctor effort. Our identifying assumption is that the change in *individual* patient behaviour is uncorrelated with the *aggregated* initial practice performance. We subject our estimation strategy to a battery of sensitivity analyses including a larger set of performance indicators and falsification tests that use unrelated outcomes and indicators.

Our key finding is that patient's effort is complementary to doctor's effort. Higher doctor disease control reduces alcohol and cigarette consumption, while no effect is found for physical activity. The results may be explained by alcohol and cigarette consumption being more observable to the doctor and likely to generate the highest health benefits to the patient. The quantitative effect is large. An average practice which increases its rate of disease control from 76% to 83% triggers a reduction in cigarettes consumption by 1.7

per day per smoker and a reduction in patient's frequency of alcohol drinking by 0.5 days per patient per month, though the latter is only significant at 10% level.

The results have two policy implications. First, for a given payment scheme, the presence of complementarities increases the effectiveness of such schemes: there are additional benefits on top of those generated by doctors' effort which are triggered by improved patients' lifestyle. Second, in terms of the design of the incentive scheme, our theoretical framework in Section 2 suggests that the optimal price per unit of doctor effort is higher in the presence of complementarities with patient effort since there are larger benefits from incentivising doctor effort. In other words, the presence of spillovers makes the incentive scheme higher powered.

Our study contributes to two distinct strands of the literature, one examining the determinants of health behaviours and the other investigating the effect of price shocks on the quality of healthcare. Firstly, we contribute to a number of studies that have analysed the determinants of health behaviours (see Cawley and Ruhm (2011) for a review), with only a few focusing on the potential complementarity or substitutability between these behaviours and healthcare. A first group of studies examines seemingly exogenous changes in medical treatment to determine its effects on health behaviours (Slade (2012); Fichera and Sutton (2011); Fichera et al. (2016)). For instance, Slade (2012) used nine waves of the Health and Retirement Study (HRS) to examine the relation between medication and the lifestyle behaviours of diabetics and found improvements in health behaviours, but only in the short-term. In order to address potential selection bias, Slade (2012) compared diabetic to pre-diabetic patients (i.e. those who have impaired glucose metabolism) using HRS biomarker data. Fichera and Sutton (2011) used three cross-sections of the Health Survey for England (HSE) to determine the effect of lipid-lowering drugs and smoking cessation advice on quitting smoking behaviour. In a trivariate probit regression they adopted an exclusion restriction involving the individuals' level of cholesterol and type of heart disease. They found evidence that doctors use medical treatment to compensate for patients' unobserved propensity to quit smoking. Using the same data, Fichera et al. (2016) found that the sharp increase in the supply of medical treatment induced by the introduction of the QOF was associated with a reduction in cigarettes consumption amongst individuals with targeted conditions. A second group of studies exploits price shocks induced by insurance to determine whether individuals complement or substitute cheaper healthcare with healthier lifestyle behaviours (Dave and Kaestner (2009); de Preux (2011); Spenkuch (2012); Yilma et al. (2012)). For instance, Dave and Kaestner (2009) used the

HRS data and compared uninsured with insured (for whom the price of healthcare is lower) pre-65 year-olds with a difference-in-differences method. They found statistically weak evidence of substitution for some lifestyle factors (i.e. physical exercise, quit smoking and alcohol consumption) but only for males. de Preux (2011) used the same data with a double robust difference-in-differences estimation to examine if individuals engage in unhealthier lifestyle behaviours in anticipation of insurance eligibility. She found evidence of a ten-percentage point reduction in physical activity but no effect on smoking or alcohol drinking.

A strong limitation of these studies is that the identification strategy hinges on the plausibility of their exclusion restrictions or comparison groups. The identification strategy adopted by Fichera and Sutton (2011) is based on an essentially untestable assumption of no direct effect of the severity of CVD on smoking behaviour. Fichera et al. (2016) assume that individuals on either side of the QOF-induced discontinuity are similar in their unobserved propensity to engage in healthier lifestyle behaviours. Dave and Kaestner (2009) assume that insured and uninsured do not differ in their unobservable characteristics.

Secondly, we contribute to the literature investigating the effect of price shocks on the quality of healthcare and on the effort of doctors. Some of these studies investigate the effect of price changes on the quality of care provided by hospitals finding an increase in incentivised procedures, a reduction in long stays and an increase in the number of patients treated (see for example, Allen et al. (2016) and Januleviciute et al. (2016)) but no effect on mortality (Seshamani et al. (2006)), volume of admissions or intensity of care (Dafny (2005)). Several studies examine the effect of price changes induced by financial incentives on the quality of primary healthcare (see for example the systematic reviews by Scott et al. (2011) and Gillam et al. (2012)) and on the effort of doctors (Flodgren et al. (2011); Gravelle et al. (2010); Sutton et al. (2010)). With regards to the QOF, the P4P scheme we examine in this paper, most studies have found an improvement in the quality of care for the incentivised areas in the first few years following its introduction after which performance reached a *plateau*. Sutton et al. (2010) found evidence of substantial increases in the recording of the incentivised risk factors when the QOF was introduced and positive spillovers onto the recording of unincentivised risk factors (such as alcohol consumption and BMI) for patients with the targeted conditions.

In the next section we provide a simple theoretical motivation for our empirical analysis. Section 3 describes the data sets. The empirical strategy is outlined in section 4. Section 5 describes the results and section 6 concludes.

2 A model of doctor and patient behaviour

We provide a stylised model of a doctor providing medical treatment and diagnosis (doctor effort), and the patient choosing her lifestyle behaviours (patient effort). We assume that the representative patient has been diagnosed with CVD and has already made the choice of visiting the doctor (therefore abstracting from modelling her participation decision).

The patient has two possible health states, H_s, after treatment is received: either the patient is healthier, $s = 1$ or not, $s = 0$ with $H_1 > H_0$. There is uncertainty in the effectiveness of medical treatment. A better health state is realised with probability $\varphi \in (0, 1)$.

Patients' utility in state s is equal to $U(H_s, I_s)$, where $I_s = I(H_s) - \tau$ denotes patient's gross income and τ a tax used to finance health care. Utility is increasing in both health and income, and concave in both arguments, $U_H > 0$, $U_I > 0$, $U_{HH} < 0$ and $U_{II} < 0$. For simplicity we assume that utility is separable in health and income, and $U_{HI} = 0$. We therefore assume that health affects utility both directly and indirectly through its income I, and health is both a consumption and an investment good, which is consistent with the Grossman (1972) model. Patients are risk-averse.

Critically, we assume that the probability of being healthier $\varphi(y, a)$ depends on patient's lifestyle behaviours (patient effort), y, and medical treatment or advise supplied by the doctor (doctor effort), a, with $\varphi_y > 0$, $\varphi_a > 0$, $\varphi_{yy} < 0$ and $\varphi_{aa} < 0$. Both increase the probability of a good state at a decreasing rate. Patient effort and doctor effort are observable and can be either complements or substitutes. Complementarity ($\varphi_{ya} > 0$) occurs when patients' healthier lifestyles increase the effectiveness of doctors' medical effort. Substitutability ($\varphi_{ya} < 0$) occurs when patients' lifestyles make doctors' medical effort less effective. For instance, medical treatment may increase life expectancy, thus inducing positive returns to patient investments. Alternatively, it is precisely because doctor's effort increases patient health that the scope for further improvements reduce. As an illustrative example, consider the prescription of lipid lowering drugs to a CVD patient. This patient might complement doctor's effort by improving her lifestyle behaviours (i.e. by reducing drinking, smoking, improving diet and physical activity). Otherwise, she might reduce her engagement in healthier lifestyle behaviours knowing that her life expectancy has anyway been increased by doctor's effort.

Patient non-monetary costs from healthier lifestyle behaviours are $g(y)$ with $g_y > 0$ and $g_{yy} > 0$.[1] These might include time costs to engage in physical activity or to cook at home. Patient's expected utility is:

$$EU(y,a) = \varphi(y,a)U(H_1, I_1) + [1 - \varphi(y,a)]U(H_0, I_0) - g(y). \tag{1}$$

For a given level of doctor's effort a, patient's optimal lifestyle $y^*(a)$ satisfies the following First Order Condition:

$$EU_y(y^*, a) \equiv \varphi_y(y^*, a) \times [U(H_1, I_1) - U(H_0, I_0)] - g_y(y^*) = 0. \tag{2}$$

The marginal benefit from higher health is equal to the marginal cost of lifestyle behaviours. The Second Order Condition is: $EU_{yy} = \varphi_{yy} [U(H_1, I_1) - U(H_0, I_0)] - g_{yy} < 0$, which is always satisfied.

The patient response to higher doctor's effort depends on the degree of complementarity and substitution of doctor's and patient's efforts. Analytically:

$$\frac{dy^*}{da} = -\frac{EU_{ya}(y^*, a)}{EU_{yy}(y^*, a)} = -\varphi_{ya}(y^*, a)\frac{U(H_1, I_1) - U(H_0, I_0)}{EU_{yy}(y^*, a)}. \tag{3}$$

The doctor is assumed to be risk-neutral, and her expected utility depends on income and the medical treatment she provides. In addition to a fixed payment $\omega > 0$, she receives a linear performance payment $q(a) = p(a - \overline{a})$ for $a \geq \overline{a}$.[2] The cost of providing medical treatment is $c(a)$ with $c_a > 0$ and $c_{aa} > 0$. We assume the doctor is altruistic and that $\alpha \in [0, 1]$ proxies the degree of doctor's altruism in relation to patient's health (as in Ellis and McGuire (1986)).

The doctor's expected utility is equal to:

$$EW(y,a) = \omega + p(a - \overline{a}) + \alpha[\varphi(y,a)H_1 + (1 - \varphi(y,a))H_0] - c(a) \tag{4}$$

For a given level of patient lifestyle behaviours, the optimal level of doctor's effort $a^*(y)$ satisfies the following First Order Condition:

[1] We could model such costs as monetary. This would make the presentation more cumbersome but would not alter the key insights of the model. The marginal cost would now be the expected marginal cost across health states.

[2] In reality, doctors' remuneration is more complex and depends on their contractual arrangement (i.e. partner or salaried doctor) and on how the QOF income is split within the practice (see for example Gravelle et al. (2010)). This simplification is appropriate for our empirical analysis. Moreover, primary care doctors have also a lower threshold below which they do not receive additional revenues. In our empirical sample hardly any GP practice is below the minimum threshold, and therefore this simplification has no implications for our empirical analysis.

$$EW_a(y, a^*) \equiv p + \alpha\varphi_a(y, a^*) \times (H_1 - H_0) - c_a(a^*) = 0. \tag{5}$$

The marginal monetary and non-monetary benefit from higher doctor's effort is equal to its marginal cost. The Second Order Condition is $EW_{aa} = \alpha\varphi_{aa}(H_1 - H_0) - c_{aa} < 0$. If the doctor has strictly positive altruism, the doctor responds to an increase in patient lifestyle behaviours by increasing or reducing effort depending on their degree of complementarity or substitution. Analytically,

$$\frac{da^*}{dy} = -\frac{EW_{ay}(y, a^*)}{EW_{aa}(y, a^*)} = -\alpha\varphi_{ya}(y, a^*) \times \frac{H_1 - H_0}{EW_{aa}(y, a^*)}. \tag{6}$$

We assume that patient's lifestyle and doctor's efforts are chosen simultaneously and we interpret this as the outcome of a long-run relationship between the patient and the doctor (which seems plausible for patients with chronic conditions where patient and doctor know each other well).

The equilibrium is denoted by the pair (y^*, a^*) which simultaneously satisfies respectively the first order condition of the patient and the doctor:

$$EU_y(y^*, a^*) = 0, \tag{7}$$

$$EW_a(y^*, a^*) = 0. \tag{8}$$

In the next section we conduct comparative statics with respect to the change in price p.

2.1 Comparative statics with respect to price

Applying Cramer's rule, we obtain the optimal response of doctor's and patient's effort to a price change:

$$\frac{da^*}{dp} = -\frac{EU_{yy}(y^*, a^*)}{\Delta} > 0, \tag{9}$$

$$\frac{dy^*}{dp} = \frac{EU_{ya}(y^*, a^*)}{\Delta} = \varphi_{ya}(y^*, a^*) \times \frac{U(H_1, I_1) - U(H_0, I_0)}{\Delta}, \tag{10}$$

where

$$\Delta := [EU_{yy}(y^*, a^*) \times EU_{aa}(y^*, a^*)] - [EU_{ya}(y^*, a^*) \times EW_{ay}(y^*, a^*)] \tag{11}$$

and $\Delta > 0$ is satisfied under minimal regularity conditions. The results are intuitive. The doctor responds to an increase in price by always increasing effort or medical treatment. Whether the patient responds by increasing or reducing effort depends on whether patient's and doctor's effort are respectively complements or substitutes.

2.2 The purchaser of health services

We assume that the purchaser of health services maximises patient health $EU(y,a)$ and medical care is financed through the tax τ, subject to the participation constraint and a limited liability constraint of the provider:

$$
\begin{aligned}
EW(y,a) &\geq 0, & (12)\\
\omega + p(a-\overline{a}) - c(a) &\geq 0. & (13)
\end{aligned}
$$

The latter ensures that the doctor does not make a negative profit. In the presence of altruism, the participation constraint is always satisfied when the limited liability constraint is satisfied, ie when $\omega+p(a-\overline{a}) = c(a)$. We therefore set the tax equal to the minimum transfer necessary to satisfy such constraint: $\tau = c(a)$, which susbstituted in patient's utility gives:

$$
EU(y,a) = \varphi(y,a)U(H_1, I(H_1) - c(a)) + [1-\varphi(y,a)]U(H_0, I(H_0) - c(a)) - g(y). \quad (14)
$$

The level of patient's and doctor's efforts which maximise patient expected utility, denoted with y^f and a^f, are given by:

$$
\begin{aligned}
EU_y(y^f, a^f) &\equiv \varphi_y(y^f, a^f) \times \Delta U(a^f) - g_y(y^f) = 0, & (15)\\
EW_a(y^f, a^f) &\equiv \varphi_a(y^f, a^f) \times \Delta U(a^f) - EU_I(y^f, a^f)c_a(a^f) = 0, & (16)
\end{aligned}
$$

where

$$
\Delta U(a^f) := [U(H_1, I(H_1) - c(a^f)) - U(H_0, I(H_0 - c(a^f)))], \quad (17)
$$

is the gain from being in a better health state evaluated at doctor's optimal effort, and

$$EU_I(y^f, a^f) := \left\{\varphi(y^f, a^f)U_I(H_1, I(H_1) - c(a^f)) + [1 - \varphi(y^f, a^f)]U_I(H_0, I(H_0) - c(a^f))\right\} \tag{18}$$

is the expected marginal utility of income across the two health states evaluated at optimal efforts.

The optimality condition for the patient is analogous to the one derived above and requires no further comments. The optimal level of doctor's effort is such that the marginal health gain is equal to the marginal cost of medical treatment weighted by the expected marginal utility of income.

Comparing the optimality condition of doctor's effort under a pay-for-performance incentive scheme,

$$EW_a(y, a^*) \equiv p + \alpha\varphi_a(y, a^*) \times (H_1 - H_0) - c_a(a^*) = 0, \tag{19}$$

with the one desired by the purchaser of health services, we obtain the price level which implements the optimal patient's and doctor's effort:

$$p^f = \varphi_a(y^f, a^f) \times \left[\Delta U(a^f) - \alpha(H_1 - H_0)\right] + c_a(a^f) \times \left\{1 - EU_I(y^f, a^f)\right\}. \tag{20}$$

The condition suggests that the purchaser sets the price equal to the marginal utility gain from higher medical treatment (first term in the square bracket). If the doctor is altruistic, then the purchaser needs to motivate less the provider, and this tends to reduce the price (second term in the square bracket). Finally, if the patient and the doctor differ in the marginal utility of income, an additional adjustment is made. For example, suppose that the marginal utility of income is higher for the patient in equilibrium. Then the marginal cost taken into account by the doctor is lower than the one of the patient, and the optimal amount of medical care and the corresponding price is also lower (third term).

The optimal price depends on the degree of complemetarity or substitution in doctor and patient effort, i.e. on φ_{ay}. If patient and doctor efforts are complements then the optimal price is higher: by incentivising doctor effort, the purchaser also incentivises patient effort. The opposite holds when efforts are substitutes. The benefit from incentivising doctor effort is crowded out by the reduction in patient effort.

The optimal pricing rule highlights that the estimation of the degree of complementarity or substitution in doctor and patient effort can help design policies aimed at incentivising doctors.

3 Data and Descriptive Statistics

Examining the patient-doctor interaction requires data on these agents' efforts. We therefore link two data sources: the English Longitudinal Study of Ageing (ELSA) and the National Health Service Quality Management and Analysis System (QMAS) database. In this section we describe each data source in turn and provide some simple descriptive statistics of our sample before discussing the way in which we utilise information on doctor payment incentives.

3.1 English Longitudinal Study of Ageing (ELSA)

The English Longitudinal Study of Ageing (ELSA) is a biannual survey and the first study in the UK to include a range of topics necessary to understand the economic, social, psychological and health elements of the ageing process. Our analysis uses waves 2 and 3 of ELSA corresponding to years 2004 and 2006, respectively, since they cover a period over which an exogenous change in doctor remuneration has occurred.

ELSA is designed to be a representative sample of those aged 50 or over and living in private households in England. For the purpose of our analysis we use data from the "core" ELSA interview questionnaires on diagnosis of diseases, health behaviours, demographic characteristics, and wealth. We match over 7,000 individuals in 2004 and over 6,000 in 2006 to the primary care practices they are registered with (more details on the matching are described in sub-section 3.2). ELSA participants were asked whether they have been diagnosed by a doctor with one of the following conditions: diabetes, high blood pressure (hypertension), angina, heart attack, heart failure, heart murmur, irregular heart rhythm and other heart problems. We choose this subset of conditions because they are covered by the payment change that we consider for doctor performance and involve a large number of patients. In order to avoid picking up changes in CVD diagnosis due to the change in the payment incentive, we select individuals who reported CVD in the 2004 wave of ELSA (i.e. the first year the payment incentive was introduced). From this sample, we obtain 3,300 individuals with CVD who are observed both in 2004 and 2006.

We consider a number of socioeconomic characteristics including household size, and whether the respondent is married or cohabiting as opposed to being divorced or separated, widowed or never married. We also consider whether the respondent is employed or self-employed as opposed to unemployed, disabled, looking after home or family or retired. We use total wealth to measure socio-economic status since it has a higher correlation with health than income in older populations (see Demakakos et al. (2008)). Total (non-pension) wealth is defined as the sum of financial wealth, physical wealth and housing wealth after deducting debts. This variable is measured in pounds sterling and deflated by the Consumer Price Index with 2005 as the base year.

As lifestyle behaviours we consider physical activity, smoking and drinking behaviour because they are measured in both the 2004 and 2006 waves. All lifestyle behaviours have been coded to be increasing in health effort. We use a measure of physical activity which is self-reported and categorised as follows: 1 (none) - not working or sedentary occupation, or engages in only mild exercise; 2 (low) - working in a job that involves standing and/or engaging in moderate activity; 3 (moderate) - working in a job involving physical work and/or engaging in vigorous activity once a week to 1-3 times a month; 4 (high) - engaged in heavy manual work and/or doing vigorous leisure activity more than once a week. In each wave of the survey respondents are asked the number of cigarettes smoked on a weekday or weekend. We calculate a weighted average of the two to determine the average number of cigarettes smoked per day. We recode it to be negative in order to indicate a lifestyle behaviour increasing in health effort. Alcohol consumption is defined as frequency of consumption in the past year. It is categorised as follows: 1=Daily; 2=Frequently: once per week or more; 3=Rarely: once/twice per week or once every two months; 4=Never.

Table 1 reports descriptive statistics for 2,249 CVD patients (top panel) of which 238 are smokers (bottom panel). The sample of CVD patients used in the analysis is smaller than the total sample of individuals with CVD because of missing values in individual characteristics. The distribution of physical activity shifts towards lower intensity from 2004 to 2006. The proportion of people drinking either daily or frequently decreases from 2004 to 2006. There is also a reduction in the average number of cigarettes per day from 14 in 2004 to 12 in 2006. We note that the balanced sample may suffer from attrition bias. However, comparing the individual characteristics and behaviours of the balanced sample with the unbalanced sample we find little difference between the two (available from the authors on request).

3.2 The Quality Management and Analysis System database

The data on quality of care for over 8,000 primary care practices in England is stored in the National Health Service Quality Management and Analysis System (QMAS) database[3]. We use this database to

[3]The data is freely available at: http://www.hscic.gov.uk/qof.

obtain the codes and addresses of all practices in England. We describe here how we matched this data to the English Longitudinal Study of Ageing (ELSA).

As part of the nurse visits carried out in 2004 and 2008, ELSA respondents were asked the name and address of their doctor. The initial sample of ELSA respondents for whom we had some information on the primary care practice they were registered with was 7,332 in 2004 and 8,138 in 2008 (as these were the years when nurse visits were implemented). After a two-stage imputation process, we have successfully matched to practices about 82% and 80% of the sample of initial ELSA respondents in 2004 and 2008. Of the 9,168 individuals who did not move between 2004 and 2008 we successfully matched about 73% to practices in 2006. The majority of these respondents were uniquely matched to practices but due to incomplete postcode or address information, there were multiple potential matches for about 6% and 5% of respondents in 2004 and 2008 respectively. In this case we use information from all the potential practices, constructing a sampling weight that equals the share of registered patients (i.e. practice j's list size) that respondent i's matched practice represents to the total list size of all the matched practices. This share is equal to one if respondent i is uniquely matched to practice j and it is less than one if she is matched to multiple practices[4].

The geographical coverage of ELSA is good. ELSA contains at least one person registered with 32% and 31% of all practices in England in 2004 and 2006, respectively. Practices are grouped geographically into 151 Primary Care Trusts. ELSA contains at least one person from each of the 151 Primary Care Trusts.

Having linked ELSA respondents to their doctor practices we are in a position to bring in information on the characteristics and performance of that practice and the behaviours of the doctors within it. General practices in the UK are a group of one to six doctors responsible for a pool of patients. We measure their deprivation by the Low Income Scheme Index (LISI), a measure of practice list deprivation from prescription data. It indicates the proportion of prescribing expenditure that was dispensed to people exempt from prescription charges on grounds of low income. We measure the performance of the practice by the Quality and Outcomes Framework (QOF) which determines about 20% of each doctor's income. We use QOF data at the practice level to obtain proxies for doctors' effort and to measure the exogenous change in their remuneration.

The QOF was officially introduced on 1^{st} April 2004 with the first measurement of performance taking place on 31^{st} March 2005. Practices are rewarded on the basis of their performance on a number of indicators.

[4]Our results are unaffected by the inclusion of multiple matches. More details on the data matching process can be made available to the reader upon request.

These indicators cover how well the practice is organised and the quality of clinical care that the practice provides to its patients. The assessment of the quality of clinical care focuses on the practice's measurement of risk factors, delivery of appropriate treatments and success in keeping the levels of risk factors under control.

The practice's "achievement", $a \in [0,1]$[5], is measured by the proportion of eligible patients who meet the quality indicator. Practices are awarded points based on these levels of achievement. Lower and upper thresholds for achievement are specified for each indicator. No points are awarded if the practice's level of achievement is less than or equal to the lower threshold. Maximum points are awarded if the practice's level of achievement is on or above the upper threshold (see Gravelle et al. (2010); Doran et al. (2011) for further details). The number of points awarded to the practice for their achievement on each indicator varies linearly between the lower and upper thresholds. The payment schedule is therefore piece-wise linear, with flat segments below the lower threshold and above the upper threshold, and a constant per-unit price in-between.

Practices are paid on the basis of the number of points they are awarded. More points are on offer for activities that require more effort. Levels of payment per point are adjusted for case-mix, such that practices with more eligible patients receive a higher level of payment per point awarded.

Since the QOF was introduced, the levels of the lower and upper thresholds and the maximum points that are available for particular indicators have been amended to stimulate better performance. To capture the change in incentive for doctor's effort, we focus on the four CVD indicators for which there was a change in the threshold levels and/or in the maximum number of points available between 2004 and 2006. Table 2 reports a description of these four indicators. There were increases in the lower thresholds from 25% to 40% for all indicators. There were increases of either 5% or 10% in the upper thresholds. There were increases in the maximum number of points available for two of the indicators and in the price per point between 2004 and 2006.

We then aggregate these four indicators into a Disease Control indicator accounting for 50 and 52 points, respectively in 2004 and 2006[6]. Disease Control indicates whether cholesterol and blood pressure are under the recommended levels.

[5] Note that a indicates the average doctor's effort in a practice as described in section 2.

[6] We aggregate them because they are highly correlated to each other where the correlation coefficient ranges between 0.3 and 0.6.

We weight the achievement rates for disease control by the proportion of total points available for each of the four indicators and we report them in Table 3. The points weighted average achievement rates of an average practice is 76% and 83% in 2004 and 2006, respectively. As the national average price per point in 2004 was £75, increasing to £126 by 2006, an average practice could earn up to £2,850 in 2004 and up to £5,438 in 2006 for its effort in controlling CVD.

We define the change in price as the "averted loss", implied by changes in the payment functions for doctor's effort, with her effort fixed at the 2004 level. Formally, $\Delta p = w^{04}(a^{04}) - w^{06}(a^{04})$ where w indicates the payment rate for doctor's effort. A graphical representation of this price change is provided in Figure 1. On the horizontal axis, we report the proportion of treated patients - the achievement rate a^t with $t = 2004, 2006$. This is indexed with 'L' or 'U' indicating lower and upper thresholds of achievement, respectively. On the vertical axis, we report the change in price between 2006 and 2004, Δp. This change in price is zero for practices with achievement below 2004 lower threshold because they will not be remunerated for this level of performance either in 2004 or in 2006[7]. If achievement rate is between 2004 and 2006 lower thresholds, the change in price is positive because with the new payment function practices would need to perform better in 2006 compared to 2004 in order to avoid a loss. The change in price then becomes progressively smaller for rates of achievement between 2006 lower threshold and 2004 upper threshold. This change in price becomes even smaller for some rates of achievement between 2004 and 2006 upper thresholds because of the change in the maximum available points. It then becomes zero if the rate of achievement is above 2006 upper threshold because practices achieve the maximum points both in 2004 and in 2006.

The price change reported in Table 3 is on average positive as practices could avoid a larger loss in 2006 than in 2004 if they increased their performance. On average, practices would avoid losing around 5% of the available payment, and they could avoid losing as much as 39% if, maintaining 2004 performance in 2006, they were above 2006 lower payment threshold.

In order to examine the representativeness of our sample of practices, we compare the practices in ELSA to the full sample of practices in England. In Table A.1 we report the number of practices in England (left panel) and those in the ELSA sample (right panel) with an achievement rate which is: i) below the lower threshold, ii) between the lower and the upper threshold, and iii) above the upper threshold in 2004 and in 2006, respectively. The proportion of total practices in ELSA that fall in each threshold interval is very

[7]Note that in our data we do not have practices below the lower threshold in 2006.

similar to that of England. In Table A.2 we report the points weighted average achievement rates and the price change of the full sample of English practices. Our sample of practices in ELSA adequately represents that of England: the mean and minimum values, and the standard deviation of the price change in the last row of Table 3 do not differ much from those in Table A.2.

4 Empirical Strategy

4.1 Main models

The main aim of the empirical analysis is to estimate patient's response to doctor's effort as defined by Eq. (1) and the comparative statics described by Eqs. (9-10). In order to do so, we first consider a linear model for *each* lifestyle behaviour of patient i registered with practice j in year $t = 2004; 2006$:

$$y_{ijt} = \beta_0 x_{ijt} + \gamma_0 a_{jt} + c_{ij} + \epsilon_{ijt} \tag{21}$$

where y_{ijt} indicates the lifestyle behaviours, x_{ijt} includes a set of socio-economic characteristics such as marital status, household size, employment status and wealth, a_{jt} indicates the practice j's points weighted average achievement rates for disease control, c_{ij} is the time-invariant unobserved component of patient i registered in practice j and ϵ_{ijt} is the error term.

We then take the first-difference between 2006 and 2004:

$$\Delta y_{ij} = \beta_0 \Delta x_{ij} + \gamma_0 \Delta a_j + \Delta \epsilon_{ij} \tag{22}$$

where $\Delta y_{ij} = y_{ijt=2006} - y_{ijt=2004}$ indicates the difference in lifestyle behaviours between 2006 and 2004 of patient i registered with practice j[8]. As a result, patient i registered in practice j time-invariant unobserved component c_{ij} is differenced out. Δx_{ij} includes changes in socio-economic characteristics.

Δa_j indicates the change in practice j's points weighted average achievement rates for disease control between 2004 and 2006. If patient and doctor efforts are complements (substitutes), the patient responds to

[8]We account for multiple practices matched to patients as follows. We create a dataset with 10 random draws of practices within each stratum (i.e. an individual-wave observation) where the probability of drawing is given by the constructed sampling weights. Note that if an ELSA respondent is uniquely matched to a single practice then she will be duplicated 10 times. Then we weight the observations in the regression using the constructed sampling weights.

doctor's effort by increasing (reducing) her effort and $\gamma_0 > 0$ ($\gamma_0 < 0$). However, estimates of γ_0 are biased if patient's response to doctor's effort is generated by the doctor's anticipation of the patient behaviour. In order to estimate the patient's response to doctor's effort that has been exogenously induced by a change in the payment system, we use an instrumental variable approach. In the first stage we predict doctor's effort from a reduced form estimation of the change in the average doctor's disease control effort on the price change (Eq. (23)). We then include this predicted effort by the doctor in the second stage equation of the change in lifestyle behaviours (Eq. (24)).

Analytically, the first stage regression of the change in the average doctor's disease control effort between 2004 and 2006 can be written as follows:

$$\Delta a_j = \zeta \Delta p_j + \lambda \Delta x_{ij} + \Delta \eta_{ij} \tag{23}$$

where Δp_j is the instrument and measures the price change as defined in the previous section, and η_{ij} is the error term. As a result, ζ is the change in doctor's effort in response to a change in price as defined in Eq. (9). We argue that this instrument is valid since, even if patients were aware of the price change (which is unlikely) they would not be aware of how the price change affected specific practices or healthcare indicators.

In addition to the socio-economic characteristics described above, Δx_{ij} includes the Low Income Scheme (LISI), a deprivation measure of practice j's list of patients.

The second-stage consists of first differenced linear models of lifestyle behaviours as in Eq. (22), but with the inclusion of the first-stage predicted doctor effort:

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \gamma_1 \Delta \hat{a}_j + \Delta u_{ij} \tag{24}$$

Patient's causal response to doctor's effort in Eq. (1) is represented by γ_1. After including the first stage predicted doctor's effort, the change in effort $\Delta \hat{a}_j$ is one exogenously induced by the payer via the change in price. As a result, the coefficient γ_1 is an unbiased estimate of patient i's response to doctor's effort and it is positive (negative) if she responds to doctor's effort by engaging in healthier (unhealthier) lifestyle behaviours.

Finally, we examine the comparative statics derived in the theory section in Eq. (10) and estimate patient's lifestyle response to the price change induced by the QOF price as follows:

$$\Delta y_{ij} = \vartheta \Delta p_j + \lambda \Delta x_{ij} + \Delta \eta_{ij} \tag{25}$$

if $\vartheta > 0$ ($\vartheta < 0$) we expect that $\gamma_1 > 0$ ($\gamma_1 > 0$) because a change in patient's lifestyle behaviour can only occur through the doctor's effort.

All models for physical activity and alcohol drinking are run on the full sample. Cigarette consumption is modelled on the sample of smokers in 2004. Lifestyle behaviours are examined separately because we found correlations below one percent between changes in cigarette and alcohol consumption, and physical activity between 2004 and 2006. Cutler and Glaeser (2005) examine changes in health behaviours (i.e. smoking, drinking heavily, being overweight, or not exercising) using the Health and Retirement Study from 1992 to 2002. They also find that behavioural changes are not highly correlated with a correlation coefficient lower than one percent.

4.2 Robustness checks

One potential concern is that practices may vary in the way they respond to the change in price depending on the initial performance. For example, practices with lower performance have more room for improvement. In our main models we control for time-variant changes in the patient pool by including the LISI. However, there still could be an omitted variable bias if unobserved factors affecting individuals' changes in health behaviours are related to the initial practice performance. It is unlikely that the individual patient underlying propensity to healthier behaviours is correlated with the *aggregated* changes in practice performance. Nevertheless, as a robustness check we modify Eq. (24) and include the initial performance of the practice in 2004:

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \gamma_1 \Delta \hat{a}_j + \gamma_2 a_j^{04} + \Delta u_{ij}. \tag{26}$$

An additional concern is the focus on only four of the total 33 indicators available for CVD. This choice was driven by the focus on those indicators that were directly affected by the price change. However, as the 33 indicators available for CVD make up for 40% of the 550 total points available for clinical care in the first year of the scheme[9], there could be effort diversion issues as worse practices might put more effort on indicators for which no change in the incentive occurred. We modify Eq. (24) and consider the full set of 33 indicators.

A final concern is that the complementarity effect we find might result from a general improvement of practices after the QOF was introduced and not from the specific CVD indicators we consider. In order to address this concern we implement falsification tests using alternative outcomes and indicators. If the complementarity effect is attributable to the disease control effort of doctors, then we should not find any effect on outcomes unrelated to patient lifestyle effort nor should we find an effect of indicators unrelated to disease control. We estimate the same 2SLS models as in Eq. (24). For the first falsification test, we use the change in non single marital status and in household size as alternative outcome variables. For the second falsification test, we use practice points weighted achievement rates for cervical screening[10] and instrument it with the price change induced by the QOF. We select this indicator because its lower threshold of achievement was raised from 25% in 2004 to 40% in 2006 and because it is completely unrelated to CVD disease control (i.e. we should not expect spillover effects).

5 Results

In Table 4 the OLS regressions provide weak evidence that patient's effort responds to doctor's effort. Although the effects are not statistically significant, the point estimates suggest that an increase in disease control would reduce the number of cigarettes per day, it would increase intensity of physical activity, but it would also increase alcohol consumption. This differential effect might be due to the direction of the bias induced by doctor's expectation of patient's effort.

As an initial step in our 2SLS approach, we examine the power of the instrument and find that the price change is a strong predictor of changes in the average doctor's disease control effort[11] as reported in Table 5. A one percentage point increase in the price shock increases disease control by about 0.5 percentage points. This means that the higher the avoided loss for practice j, i.e. the higher is the effective price that the

[9] More specifically, 15 indicators were available for Coronary Heart Diseases (CHD) with a maximum of 121 points and 18 indicators were available for diabetes with 99 points in total. Note that CVD encompasses both CHD and diabetes.

[10] This measures the proportion of patients aged 25-64 years whose notes record that a cervical smear has been performed in the last three to five years.

[11] We run the first stage regressions separately on each sample of lifestyle behaviours.

practice is facing, the more it would do to manage its patients' disease. This positive relation between the average doctor's effort and the change in price is the one we expected from our comparative statics in Eq. (9). The first stage F-statistics (not reported on the Table) take values between 11.59 and 19.93 that are above both the critical values outlined in Stock and Yogo (2005) and the conventional minimum value of F=10 (Stock et al. (2002)), meaning that we can reject the null that the price change is a weak instrument. Given that we cluster standard errors at the practice level, errors cannot be assumed to be i.i.d. and so we also check the Kleibergen-Paap Wald rank F-statistics, robust to non-i.i.d errors, and we confirm the results of the previous test.

In the second stage in Table 6, we provide our key results and find that patients complement more effective disease control effort by the doctor with better lifestyle behaviours. An average practice which increases its rate of disease control from 76% to 83% triggers a reduction in patient's frequency of alcohol drinking by 0.5 days per patient per month and a reduction in cigarettes consumption by 1.7 per day per smoker[12]. The former is however only weakly statistically significant at 10% level. This indicates that doctor and patient efforts are complements, i.e. $\varphi_{ya}(y^*, a) > 0$ in Eq. (1).

Given this complementarity between patient and doctor efforts ($\varphi_{ya}(y^*, a) > 0$), we would also expect that patient's health behaviour responds to an increase in price, i.e. $dy^*/dp > 0$ in Eq. (10). We report the estimates of $\vartheta \equiv dy^*/dp$ in Table 7 and find that the higher the practice j's avoided loss from a change in the payment function, the more effort patient i would put in engaging in healthier lifestyle behaviours (i.e. $\vartheta > 0$). In other words, for an average practice facing a change in price as much as 5%, there would be a decrease of drinking equal to 0.2 days per patient per month and a decrease of 0.7 cigarettes per day[13].

The results of the robustness checks are displayed in Table 8. We find that the complementary effect of the average doctor's effort with the patient's cigarette and alcohol consumption is robust to the inclusion of the initial practice performance in 2004. Although this effect is more statistically significant than in the main models, its size is around half the one in Table 6, for cigarette and alcohol consumption, respectively. For an average practice increasing its rate of disease control from 76% to 83% there would be a reduction of the frequency of alcohol drinking by 0.2 days per patient per month and a reduction of one cigarette per

[12]These are calculated as the average change in doctor's effort in practice j times the average change in patient i's alcohol drinking (i.e. $(83-76)*0.07$) and smoking (i.e. $(83-76)*0.25$).

[13]These are calculated as the average price change of practice j times the average change in patient i's alcohol drinking (i.e. $(5)*0.04$) and smoking (i.e. $(5)*0.14$).

day per smoker[14]. Baseline performance is statistically significant in the first stage, but it is not displayed on the table.

We also report the average doctor's effort in all indicators that apply to CVD. The sample is slightly smaller than in previous tables because some practices did not report their achievement in at least one of the additional indicators that we consider. This might explain why the effect on alcohol consumption is not statistically significant. The size and statistical significance of the relation between average doctor's effort and patient's cigarettes consumption is slightly larger the one in Table 6 mitigating concerns of effort diversion. For an average practice increasing its rate of disease control from 76% to 83% there would be a reduction of two cigarettes per day per smoker.

Finally, in Tables 9 and 10 we report the coefficients of the falsification tests, the second stage models using alternative outcomes and indicators. We find no evidence of a statistically significant effect of the change in disease control on the change in marital status or household size. We also find no statistically effect of the change in cervical screening effort on patient lifestyle behaviours[15].

6 Conclusions

The use of performance pay incentives is increasing in public sectors such as education and healthcare. Whilst numerous studies have investigated the effect of these incentives on targeted agents, not much is known about their spillovers onto untargeted agents. This is particularly important in complex organisations where the outcome (health or education) is a joint product between different agents. In such contexts, incentives onto targeted agents might increase or reduce the effort provided by untargeted agents.

This is the first paper that examines the interaction between doctor and patients in the joint production of health. Lack of data and reasonable instruments to determine patient's response to doctor's provision of treatment are two reasons for this gap in research, which this study fills.

Our stylised model has obtained predictions of a patient-doctor interaction to motivate our empirical strategy. It shows that patient's response to doctor's effort depends on the degree of complementarity and substitution of doctor and patient efforts. Comparative statics suggests that doctors respond to an increase

[14] These are calculated as the average change in doctor's effort in practice j times the average change in patient i's alcohol drinking (i.e. $(83-76)*0.03$) and smoking (i.e. $(83-76)*0.14$).

[15] The first stage models (not reported on the table) indicate that the price change is a strong instrument for doctor's effort in cervical screening.

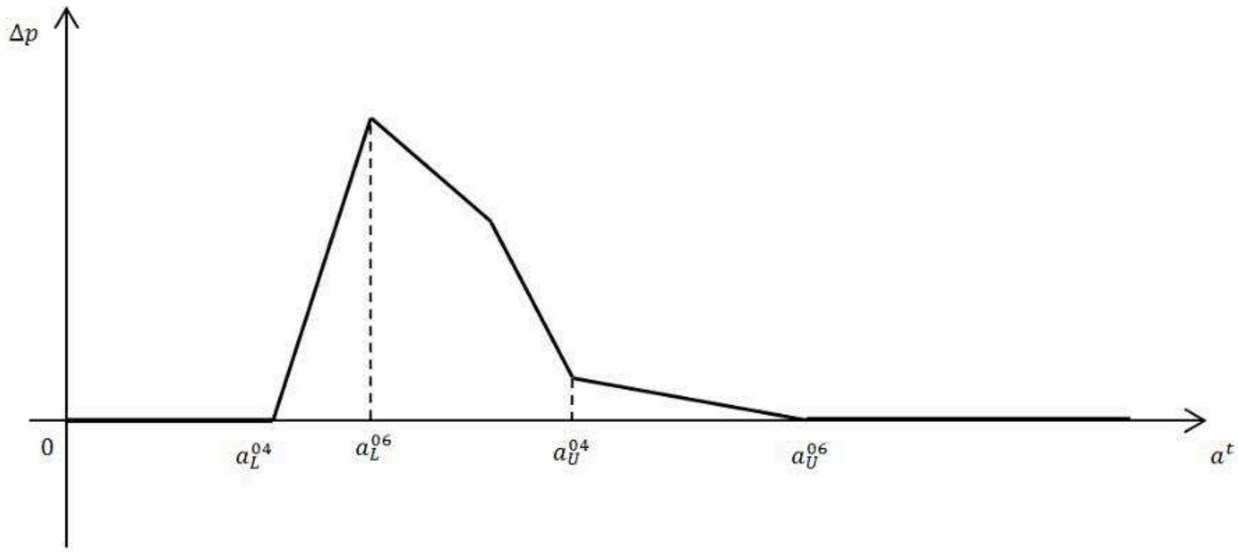
in price in their payment system by increasing effort. The extent to which patients will increase or decrease their effort in response to a price change also depends on the degree of complementarity and substitution between patient's and doctor's efforts. This matters for welfare and the optimal design of incentive schemes: if patient's and doctor's efforts are complements then the optimal price, i.e. the power of the incentive scheme, is higher. This is because by incentivising doctor's effort, the purchaser also incentivises patient's effort, a form of positive spillover effect. Conversely, if efforts are substitutes, then the spillover is negative and the optimal price is lower.

Our empirical analysis suggests that patient's and doctor's efforts are complements, as patients increase their efforts in healthier lifestyle behaviours when doctors increase their treatment efforts. More precisely, for an average primary care practice increasing its CVD control rate of their patients from 76% to 83% there is a reduction of frequency of drinking by 0.5 days per patient per month and cigarettes consumption by 1.7 cigarettes per day per smoker. We find no effect on physical activity. Our identification strategy relies on a 2SLS approach which instruments doctor's effort with an exogenous change in price induced by the QOF incentive scheme: one percentage point increase in this price shock increased doctor's disease control effort by about 0.5 percentage points.

One limitation of this study is the lack of data on individual doctor's disease control effort. We use instead the average performance of the practice. This is advantageous in that it avoids selection biases caused by matching of patients and individual doctors, but it does induce a form of measurement error. Additionally, the small sample size prevents us from analysing practice heterogeneities. Despite picking up a local average treatment effect, our results are relevant for all those practices that were not always under or over performing throughout the sample period (over 80% of the 2006 practices). A final limitation is the use of self-reported data on lifestyle behaviours which might also generate a measurement error. We assume that this error is time invariant and therefore it is differenced out in our empirical specification. For instance, Clark and Etilé (2002) find that in the British Household Panel Survey smokers report cigarettes consumption in fractions and multiples of a packet (a measurement error called "heaping" in self-reported job tenure data). Instead of aggregating the data around "heaping" points for the case of cigarette smoking, a process that would reduce variation in our data, we assume that in the short time period between interviews individuals will always make the same approximation when asked about the average number of cigarettes smoked per day.

Nevertheless, our results suggest that neglecting interactions between doctors and patients can lead to underestimating the impact of health care reforms which incentivise doctors' effort on patients' health behaviours because by incentivising doctors' effort purchasers can also incentivise patients' effort. In terms of policy design, our empirical results also suggest that doctors' incentive schemes could be higher powered since there are higher overall benefits from setting higher prices. On the methodological side, our analysis demonstrates the potential value of research exploiting matched data on doctors and their patients. Such data are becoming more readily available for researchers both in the UK and elsewhere, and we would expect this development to lead to other empirical insights on health behaviours and their interaction with healthcare.

Figure 1: *The Quality and Outcomes Framework price change function*



Note: a^t indicates the proportion of treated patients and Δp is the change in price.

Table 1: *Definition of the variables used in the models by year*

All sample		**2004**		**2006**	
	Definition	**Mean**	**S.D.**	**Mean**	**S.D.**
	Outcome variables:				
Physical activity	Physical activity on a 4-point scale:	(2,249)		(2,249)	
	1=no physical activity	0.18%		0.18%	
	2=low physical activity	15.61%		18.19%	
	3=moderate physical activity	55.45%		54.87%	
	4=high physical activity	28.77%		26.77%	
Alcohol drinking	Alcohol drinking in 4 points scale:	(2,080)		(2,080)	
	1=daily	19.57%		17.74%	
	2=frequently	41.49%		41.35%	
	3=rarely	29.23%		29.28%	
	4=never	9.71%		11.63%	
	Individual characteristics:				
Household size	Number of household members	1.9	0.8	1.9	0.8
		(2,249)		(2,249)	
Non single	Married/cohabitee/civil partner	67.045%		68.8%	
		(2,249)		(2,249)	
Ln(wealth)	Natural logarithm of real equivalised household wealth in £	11.87	1.6	11.87	1.6
		(2,249)		(2,249)	
Employed	Whether individual is employed vs. unemployed, disabled, retired, home work	26.85%		24.41%	
		(2,249)		(2,249)	

Smokers in 2004		**2004**		**2006**	
	Definition	**Mean**	**S.D.**	**Mean**	**S.D.**
	Outcome variable:				
Number of cigarettes per day	Average no. cigarettes per day (negatively coded)	-13.94	8.67	-12.24	9.34
		(238)		(238)	
	Individual characteristics:				
Household size	Number of household members	1.83	0.77	1.78	0.76
		(238)		(238)	
Non single	Married/cohabitee/civil partner	62.18%		57.98%	
		(238)		(238)	
Ln(wealth)	Natural logarithm of real equivalised household wealth in £	11.06	2.18	11.13	2.06
		(238)		(238)	
Employed	Whether individual is employed vs. unemployed, disabled, retired, home work	26.05%		22.27%	
		(238)		(238)	

Notes: descriptive statistics on the balanced sample of individuals 50+ with CVD for whom practice characteristics and individual characteristics are observed. The sample for individual characteristics is conditional on physical activity being also observed. Number of observations in (); S.D.=standard deviation. Wealth is measured in pounds sterling and deflated by the Consumer Price Index with 2005 as the base year. Physical activity is coded as follows: 1 (none) - not working or sedentary occupation, or engages in only mild exercise; 2 (low) - working in a job that involves standing and/or engaging in moderate activity; 3 (moderate) - working in a job involving physical work and/or engaging in vigorous activity once a week to 1-3 times a month; 4 (high) - engaged in heavy manual work and/or doing vigorous leisure activity more than once a week. The sample of CVD patients (smokers and non-smokers) used in the analysis is lower than the sample of individuals with CVD because of missing values in the individual characteristics. We also note that the balanced sample may suffer from attrition bias. We compare the individual characteristics and behaviours of the balanced sample with the unbalanced sample and we find little difference between the two. This additional table can be made available from the authors on request.

Table 2: *Description of indicators of doctors' disease control*

Indicator name	Description	2004 LT	2004 UT	2004 Points	2004 Price per point	2006 LT	2006 UT	2006 Points	2006 Price per point
CHD8	The percentage of patients with coronary heart disease whose last measures cholesterol (measured in the last 15 months) is 7mmol/l or less.	25	60	16	£75	40	70	17	£126
DM7	The percentage of patients with diabetes in whom the last HbA1C is 10 or less (or equivalent test/reference range depending on local laboratory) in last 15 months.	25	85	11	£75	40	90	11	£126
DM12	The percentage of patients with diabetes in whom the last blood pressure is 145/85 or less.	25	55	17	£75	40	60	18	£126
DM17	The percentage of patients with diabetes whose last measured total cholesterol within previous 15 months is 5 or less.	25	60	6	£75	40	70	6	£126

LT=Lower Threshold; UT=Upper Threshold.

Table 3: *Summary statistics on the average achievement rates and price change*

	Mean	Min.	Max.	Std. Dev.
Disease Control in 2004 $[a^{04}]$	0.76 (2,642)	0.004	1	0.08
Disease Control in 2006 $[a^{06}]$	0.83 (2,523)	0.08	1	0.05
Price change $[\Delta p]$	0.05 (2,269)	0	0.39	0.07

Note: Statistics weighted by the constructed sampling weights. The achievement rates are weighted by the number of points. Sample sizes in () represent practices.

Table 4: *Coefficients of first differenced linear models of health behaviours*

	Model I: Increase in physical activity	**Model II: Reduction in alcohol drinking**	**Model III: Reduction in no. cigarettes**
Change in Disease Control (γ_0)	0.28	-0.10	8.66
	(0.24)	(0.23)	(7.27)
Change in non single status	0.01	0.12	1.88
	(0.11)	(0.10)	(1.81)
Change in employed status	0.02	0.08*	1.29
	(0.06)	(0.04)	(2.05)
Change in household size	-0.01	0.04	-1.07
	(0.03)	(0.03)	(0.93)
Change in Ln(equivalised wealth)	0.02	0.02	-0.16
	(0.02)	(0.02)	(0.57)
Constant	-0.07***	0.07***	1.08
	(0.02)	(0.02)	(0.76)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD. Model III on sample of smokers in 2004. Lifestyles increasing in health effort: intensity of physical activity, rarity of alcohol drinking and reduction in no. cigarettes.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 5: *First stage coefficients of first differenced linear models of practice achievement on the price change*

	Change in Disease Control on physical activity sample	**Change in Disease Control on alcohol sample**	**Change in Disease Control on no. cigarettes sample**
Price change (ζ)	0.52***	0.51***	0.55***
	(0.03)	(0.03)	(0.04)
Change in non single status	0.0004	-0.004	-0.001
	(0.01)	(0.01)	(0.02)
Change in employed status	-0.01	-0.01	0.003
	(0.004)	(0.005)	(0.01)
Change in household size	0.0005	0.001	-0.001
	(0.002)	(0.003)	(0.01)
Change in Ln(equivalised wealth)	0.001	0.0002	0.0001
	(0.002)	(0.002)	(0.004)
Change in LISI	-0.003*	-0.002*	-0.001
	(0.0001)	(0.001)	(0.001)
Constant	0.05***	0.05***	0.05***
	(0.002)	(0.002)	(0.004)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD.
$^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$.

Table 6: *Second stage coefficients of first differenced linear models of health behaviours*

	Model I: Increase in physical activity	**Model II: Reduction in alcohol drinking**	**Model III: Reduction in no. cigarettes**
Change in Disease Control (γ_1)	0.13	0.68*	23.81**
	(0.39)	(0.40)	(10.22)
Change in non single status	0.01	0.13	2.02
	(0.11)	(0.10)	(1.97)
Change in employed status	0.02	0.08*	1.20
	(0.06)	(0.04)	(1.99)
Change in household size	-0.006	0.04	-1.08
	(0.03)	(0.03)	(1.02)
Change in Ln(equivalised wealth)	0.02	0.02	-0.15
	(0.02)	(0.02)	(0.58)
Constant	-0.05*	0.01	-0.15
	(0.03)	(0.03)	(0.95)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD. Model III on sample of smokers in 2004. Lifestyles increasing in health effort: intensity of physical activity, rarity of alcohol drinking and reduction in number of cigarettes.
$^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$.

Table 7: *Coefficients of first differenced linear models of health behaviours with price change*

	Model I: Increase in physical activity	**Model II: Reduction in alcohol drinking**	**Model III: Reduction in no. cigarettes**
Price change (ϑ)	0.09	0.38*	13.29**
	(0.20)	(0.20)	(5.53)
Change in non single status	0.01	0.12	1.99
	(0.11)	(0.10)	(1.70)
Change in employed status	0.02	0.08*	1.28
	(0.06)	(0.04)	(2.00)
Change in household size	-0.01	0.04	-1.10
	(0.03)	(0.03)	(0.85)
Change in Ln(equivalised wealth)	0.02	0.02	-0.14
	(0.02)	(0.02)	(0.57)
Constant	-0.05***	0.05***	1.03*
	(0.02)	(0.02)	(0.57)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD. Model III on sample of smokers in 2004. Lifestyles increasing in health effort: intensity of physical activity, rarity of alcohol drinking and reduction in no. cigarettes.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 8: *Tests of robustness using second stage models*

	Model I: Increase in physical activity	**Model II: Reduction in alcohol drinking**	**Model III: Reduction in no. cigarettes**
Test A. Including Disease Control in 2004:			
Change in Disease Control (γ_1)	0.10	0.29***	13.92***
	(0.31)	(0.10)	(3.37)
Constant	-0.05*	0.04***	0.65**
	(0.03)	(0.008)	(0.30)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247
Test B. Including All CVD indicators:			
Change in Disease Control (γ_1)	-0.26	0.39	28.32**
	(0.47)	(0.45)	(13.16)
Constant	-0.04	0.04	0.10
	(0.03)	(0.03)	(0.87)
No. observations	21,268	19,660	2,261
No. individuals	2,135	1,973	227
No. practices	1,408	1,303	235

Statistics weighted by the constructed sampling weights and clustered std. errors in (). Model III on sample of smokers in 2004. Lifestyles increasing in health effort: intensity of physical activity, rarity of alcohol drinking and reduction in number of cigarettes. Second stage regressions displayed. All models include marital status, whether employed, household size, and wealth. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 9: *Tests of falsification using second stage models with alternative outcomes*

	Model I: Change in non single status	**Model II: Change in household size**
Change in Disease Control	-0.03	0.25
	(0.08)	(0.23)
Change in non single status	-	0.87
		(0.14)
Change in employed status	-0.0003	0.01
	(0.01)	(0.04)
Change in household size	0.11***	-
	(0.03)	
Change in Ln(equivalised wealth)	0.01*	-0.01
	(0.01)	(0.01)
Constant	-0.01	-0.05***
	(0.01)	(0.02)
No. observations	22,467	22,467
No. individuals	2,249	2,249
No. practices	1,501	1,501

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD. All models on the sample with no missings in physical activity.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 10: *Tests of falsification using second stage models with alternative indicator*

	Model I: Increase in physical activity	**Model II: Reduction in alcohol drinking**	**Model III: Reduction in no. cigarettes**
Change in Cervical Screening	0.43	0.17	4.67
	(0.74)	(0.80)	(15.54)
Change in non single status	0.01	0.12	1.82
	(0.11)	(0.10)	(1.79)
Change in employed status	0.02	0.08*	1.38
	(0.06)	(0.04)	(2.10)
Change in household size	-0.006	0.04	-1.06
	(0.03)	(0.03)	(0.91)
Change in Ln(equivalised wealth)	0.02	0.02	-0.15
	(0.02)	(0.02)	(0.58)
Constant	-0.05***	0.06***	1.76***
	(0.02)	(0.01)	(0.51)
No. observations	22,467	20,782	2,376
No. individuals	2,249	2,080	238
No. practices	1,501	1,394	247

Statistics weighted by the constructed sampling weights and clustered std. errors in (). No. observations for the 10 imputed datasets on the sample of people aged 50+ with CVD. Model III on sample of smokers in 2004. Lifestyles increasing in health effort: intensity of physical activity, rarity of alcohol drinking and reduction in number of cigarettes.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

A Additional results

Table A.1: *Sample of practices in England and ELSA by threshold levels for Disease Control*

	England:				**ELSA/Practice linkage**			
	LT	***(LT, UT)***	***UT***	***Total***	***LT***	***(LT, UT)***	***UT***	***Total***
2004/05	3 (0.04%)	2,675 (32.28%)	5,609 (67.68%)	8,287	1 (0.04%)	665.69 (25.18%)	1,977.54 (74.82%)	2,643.23
2006/07	1 (0.01%)	2,077 (24.85%)	6,279 (75.13%)	8,357	0 (0.00%)	428.63 (17.00%)	2,092.63 (83.01%)	2,521.25

Note: Statistics weighted by the constructed sampling weights LT=Upper Threshold level and UT=Lower Threshold level; (LT, UT) between lower and upper threshold. Achievements from ELSA practices are calculated on the sample of individuals 50+. Proportion of total sample in ().

Table A.2: *Summary statistics on the points weighted average achievement rates and price change (England)*

	Mean	**Min.**	**Max.**	**Std. Dev.**
Disease Control [a^{04}]	0.76 (8,287)	0.004	1	0.09
Disease Control [a^{06}]	0.83 (8,357)	0.08	1	0.06
Price change [Δp]	0.06 (8,055)	0	0.40	0.08

Note: Statistics weighted by the constructed sampling weights. Sample sizes in () represent practices.

References

Allen T., Fichera E. and Sutton M. Can payers use prices to improve quality? Evidence from English hospitals. Health Economics, 2016; 25: 56-70.

Burgess S., Propper C., Ratto M., von Hinke Kessler Scholder S., Tominey E. Smarter task assignment or greater effort: the impact of incentives on team performance. The Economic Journal, 2010; 120(547): 968-989.

Clark A. and Etilé F. Do health changes affect smoking? Evidence from British panel data. Journal of Health Economics, 2002; 21(4): 533-562.

Cawley J. and Ruhm C. The economics of risky health behaviours. IZA Discussion Paper Series No. 5728, 2011.

Cutler D.M. and Glaeser E. What Explains Differences in Smoking, Drinking, and Other Health-Related behaviours? The American Economic Review; 95(2): 238-242.

Dafny LS. How do hospitals respond to price changes? The American Economic Review, 2005; 95: 1525-1547.

Dave, D., Kaestner, R. Health insurance and ex ante moral hazard: Evidence from Medicare. Journal of Health Care Finance and Economics, 2009; 9 (4): 1-24.

Department of Health. Long Term Conditions Compendium of Information: Third Edition 2012. https://www.gov.uk/government/publications/long-term-conditions-compendium-of-information-third-edition (accessed 12.21.16).

Department of Health. Cardiovascular Disease Outcomes Strategy. Improving outcomes for people with or at risk of cardiovascular diseases, 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217118/9387-2900853-CVD-Outcomes_web1.pdf (accessed 12.21.16).

de Preux L. Anticipatory ex ante moral hazard and the effect of Medicare on prevention. Health Economics, 2011; 20: 1056-72.

Demakakos P., Nazroo J., Breeze E., Marmot M. Socioeconomic status and health. The role of subjective social status. Social Science & Medicine 2008; 67: 330-340.

Doran T., Kontopantelis E., Valderas J.M., Campbell S., Roland M., Salisbury C. and Reeves D. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ 2011; 342:1-12.

Elberts R., Hollenbeck K. and Stone J. teacher performance incentives and student outcomes. The Journal of Human Resources, 2002; 37 (4):913-927.

Ellis R. and McGuire T. Provider behavior under prospective reimbursement: Cost sharing and supply. Journal of Health Economics, 1986; 5(2): 129-151.

Fichera E. and Sutton M. State and self investments in health. Journal of Health Economics, 2011; 30(6): 1164-1173.

Fichera E., Gray E. and Sutton M. How do individuals' health behaviours respond to an increase in the supply of health care? Evidence from a natural experiment. Social Science & Medicine, 2016; 159: 170-179.

Figlio D.N. and Kenny L.W. Individual teacher incentives and student performance. Journal of Public Economics, 2007; 91(5-6): 901-914.

Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. Database of Systematic Reviews 2011, Issue 7.

Gillam S.J., Siriwerdena A. N. and Steel N. Pay-for-Performance in the United Kingdom: impact of the Quality and Outcomes Framework: a systematic review. Annals of Family Medicine, 2012; 10(5): 461-468.

Glewwe P., Ilias N. and Kremer M. Teacher Incentives. American Economic Journal: Applied Economics, 2010; 2: 205-227.

Gravelle H., Sutton M. and Ma A. Doctor behaviour under a pay for performance contract: treating, cheating and case finding? The Economic Journal, 2010; 120(February) F129-156.

Grossman M. On the concept of health capital and the demand for health. Journal of Political Economy, 1972; 80 (2): 223-255.

Januleviciute J., Askildsen J.E., Kaarboe O., Siciliani L. and Sutton M. How do hospitals respond to price changes? Evidence from Norway. Health Economics, 2016; 25: 620-636.

Lavy V. Performance pay and teachers' effort, productivity and grading ethics. The American Economic Review, 2009; 99(5): 1979-2021.

National Institute for Health and Care Excellence. Type 2 Diabetes: the Management of Type 2 Diabetes, 2009. Retrieved from: http://www.nice.org.uk/guidance/cg87/ifp/chapter/cardiovascular-disease.

National Institute for Health and Care Excellence. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular diseases, 2014. Retrieved from: http://www.nice.org.uk/guidance/cg181/chapter/introduction.

National Institute for Health and Care Excellence. Type 2 Diabetes in Adults: Management, 2015. Retrieved from: https://www.nice.org.uk/guidance/ng28/chapter/1-recommendations.

Prendergast C. The provision of incentives in firms. Journal of Economic Literature, 1999; 37: 7-63.

Propper C., Sutton M., Whitnall C. and Windmeijer F. Incentives and targets in hospital care: Evidence from a natural experiment. Journal of Public Economics, 2010; 94(3-4): 318-335.

Propper C. and Wilson D. The use of performance measures in health care systems. The Elgar Companion to Health Economics, Second Edition, 2012; 350-358.

Scott A., P. Sivey, D.A. Ouakrim, L. Willenberg, L. Naccarella, J. Furler and Young D. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database Syst Rev, 2011.

Seshamani M., Schwartz J.S., Volpp K.G.. The Effect of Cuts in Medicare Reimbursement on Hospital Mortality. Health Services Research, 2006; 41: 683-700.

Slade A.N.. Health investment decisions in response to diabetes information in older Americans. Journal of Health Economics 31, 2012: 502-520.

Spenkuch J.L. Moral hazard and selection among the poor: Evidence from a randomized experiment. Journal of Health Economics, 2012; 31(1): 72-85.

Stock, J.H., Wright, J.H., Yogo, M. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. J. Bus. Econ. Stat. 20, 2002: 518?529.

Stock, J.H., Yogo, M. Testing for Weak Instruments in Linear IV Regression, in: Identification and Inference for Econometric Models. Cambridge University Press, New York, 2005: 80?108.

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

Yilma, Z., Kempen, L., Hoop, T. A perverse "net" effect? Health insurance and ex-ante moral hazard in Ghana. Social Science & Medicine, 2012; 75(1): 138-47.

Wanless D. Securing our Future Health: Taking a Long-Term View. HM Treasury 2002.