UNIVERSITY OF
BATH

University of Bath

Does Patient Health Behaviour respond to Doctor Effort?

Eleonora Fichera, James Banks, Luigi Siciliani, Matt Sutton [*]

October 20, 2018

Abstract

Incentive pay systems have been introduced in public sectors such as education and health care. In these sectors the output (education or health respectively) depends on the actions of different agents and it is unclear what the effects of such incentive systems are on the behaviour of untargeted agents. In this study we focus on patient health, modelled as a joint product of patient effort (through lifestyle and behaviour) and doctor effort (through diagnosis and treatment). Patient response to doctor effort is shown to be *a priori* ambiguous and depends on the degree of complementarity or substitution between doctor and patient effort. We build an empirical model to estimate the effect of doctors' treatment effort on patient behaviour. To address the endogeneity of doctor effort we exploit a change in payments to doctors in the U.K. that led to incentive changes that varied by practice, depending on their prior performance levels. We use panel data on the physical activity, drinking and smoking behaviours of over 2,000 cardiovascular disease patients aged over 50 in England and link these data to their primary care practice performance data. Our results indicate that 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, suggesting that patient efforts are complements to doctor effort. Understanding such complementarities has implications for assessing the design and effectiveness of pay-for-performance schemes which encourage higher doctor effort.

Keywords: Health care incentives, health behaviours, pay-for-performance.
JEL classification: I1

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 health care to name a few (Lavy (2009); Propper et al. (2010); Propper and Wilson (2012)). Most research in this area 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 the incentive effects on targeted agents to be complemented or substituted by 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 in classes. In this context, incentives that aim to improve teacher performance may be enhanced by pupils' complementary actions. Some studies have found that performance pay may actually reduce pupil GPA and daily attendance rates, especially when performance pay is related to pupil feedback (Elberts et al. (2002)). Others 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 health, that, like education, is a joint product between two agents. In this case the agents are doctors and patients, where doctors provide medical treatment and patients choose their lifestyle behaviours. Our contribution is to investigate whether patient effort is complementary to doctor effort using an arguably-exogenous variation in doctor effort generated by the largest pay for performance incentive for English primary care practices, the Quality and Outcomes Framework (QOF). There is very little evidence on how the QOF has affected patients, particularly with respect to their health behaviours and their health outcomes (Gillam et al. (2012); Sutton et al. (2010)). 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 wider implications for the effectiveness of schemes which encourage doctor's

effort. If patient and doctor efforts are complements then by incentivising doctor's effort the purchaser also incentivises patient's effort. In this case, incentive based schemes will have additional value to the one that might have been expected *ex ante*.

We estimate this potential complementarity between doctor and patient actions specifically in the context of patients with cardiovascular disease in England. This is an area of health care where best practice management and prevention recommends 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)) and where a new strategy encouraging people to engage in their own health and health care was recently introduced (Wanless (2002) and the Department of Health (2013)). To estimate our model we use a large sample of individuals diagnosed with cardiovascular diseases (CVDs) from the English Longitudinal Survey of Ageing (ELSA) between 2004 and 2006 and implement a new linkage between ELSA respondents and administrative data on primary care practice treatment rates.

One clear difficulty in estimating the effect of doctor behaviour on patients actions is the potential endogeneity of the doctor's decision. We address this issue by means of a two-stage least squares (2SLS) approach in which doctor's effort is instrumented with the potential change in revenue induced by changes in a large pay-for-performance (P4P) scheme, the QOF. This incentive system was introduced in the U.K. in 2004 to reward primary care practices for providing high-quality care (relative to pre-defined thresholds) in a number of disease areas. Specifically, our identification exploits the fact that between 2004 and 2006 primary care practices faced 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 (practice-specific) financial incentive that was generated by the QOF and we use this variation as an instrument for doctor effort. Our identifying assumption is that the potential change in revenue induced by changes in the QOF prices is independent of subsequent changes in all *individual* patient lifestyle determinants and changes in *aggregate* practice performance, *conditional* on a rich set of patient characteristics and practice-level area-based deprivation measures. We provide evidence in support of our identification strategy and subject our estimation to a battery of sensitivity analyses including a larger set of performance indicators and falsification tests that use unrelated outcomes, patient groups and indicators.

Our key finding is that patient's effort is complementary to doctor's effort in some important dimensions. Higher doctor disease control reduces alcohol and cigarette consumption, although 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 corresponds to 7-12% of the sample mean of cigarette smoking. An average practice which increases its rate of disease control from 76% to 83% triggers a reduction in cigarette consumption by 1-1.7 per day per smoker and a reduction in patient's frequency of alcohol drinking by 0.05 points (about 2% of the sample mean), though the latter is only statistically significant at 10% level.

In terms of the design of incentive schemes, the policy implication of our study is 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, interactions between agents make the incentive scheme higher powered.

Our study relates to two distinct strands of the health care literature, one examining the determinants of health behaviours and the other investigating the effect of price shocks on the quality of health care. Most directly, we add to the body of studies that have analysed the determinants of health behaviours (see Cawley and Ruhm (2011) for a review) and have also focused on the potential complementarity or substitutability between these behaviours and health care arrangements. These studies are grounded in economic models by Ehrlich and Becker (1972) and Peltzman (1975) where "state investments" that provide an insurance role against the magnitude of negative effects of diseases and a protection role against the probability that such diseases occur, may be complemented or substituted by "self-investments".

A first group of such studies examine changes in medical treatment to determine its effects on health behaviours (Bennett (2012); Slade (2012); Fichera and Sutton (2011); Fichera et al. (2016); Kaestner et al. (2014); Wilson et al. (2014)). 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

4

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 cigarette consumption amongst individuals with targeted conditions. Kaestner et al. (2014) used the gradual penetration of statins in the U.S. market since their introduction in 1987 as an instrument for statin use. They found that an increase in statins was associated with a 0.3-0.5 point increase in BMI for females and males and an increase of 15% of the mean in moderate alcohol consumption by males. Their results provide evidence for a strong substitutability of healthier behaviours and health care. However, they found no consistent evidence of a decrease in smoking as a result of statin use.

A second group of studies exploits price shocks induced by insurance to determine whether individuals complement or substitute cheaper health care with healthier lifestyle behaviours (Dave and Kaestner (2009); de Preux (2011); Spenkuch (2012); Yilma et al. (2012); Lakdawalla et al. (2006); Bhattacharya et al. (2011)). For instance, Dave and Kaestner (2009) used the HRS data and compared uninsured with insured (for whom the price of health care 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. Lakdawalla et al. (2006) showed that breakthroughs in HIV treatment coupled with cheaper access to medical technologies increases HIV-infections. This happens because HIV+ patients who get treatment in U.S. states with more generous Medicaid eligibility rules are more likely to engage in risky sexual behaviours.

In addition to not modelling the links between doctor and patient effort, these studies are empirical studies where identification 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 but plausible 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.

Since our 2SLS strategy requires us to model doctor effort as well as patient behaviour, our study also

contributes directly to the literature on the effect of price shocks on the quality of health care 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 health care (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 and results are outlined in section 4. Section 5 and 6 describe the robustness checks and mechanisms, respectively. Section 7 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 effect on 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 $\varphi(y, a)$, which is a key determinant of the *health production* function, depends on patient's lifestyle behaviours (patient effort), y, and medical treatment or advice 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 in the probability of being healthy.

The function $\varphi(y, a)$ gives the production function through which different efforts translate into a different probability of falling ill. Complementarity ($\varphi_{ya} > 0$) occurs when patients' healthier lifestyles increase the effectiveness of doctors' medical effort. If the patient has a healthier lifestyle (an increase in y), an increase in doctor's effort a becomes more effective in improving the health of the patient. Substitutability ($\varphi_{ya} < 0$) occurs when patients' lifestyles make doctors' medical effort less effective or productive. Therefore, the function $\varphi(y, a)$ can be interpreted as the technology through which different efforts translate into the probability of being healthy.

As an illustrative example, consider the prescription of lipid lowering drugs to a CVD patient. If the patient has better lifestyle behaviours (i.e. less drinking, smoking, better diet and more physical activity) the effect of the drug in reducing the probability of falling ill can be higher (because the treatment is more effective in patients who do not smoke or drink) or it might be lower (if a higher health makes it more difficult to achieve further health gains).

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.

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). We first derive the optimal level of patient effort for a given doctor effort. Second, we derive the optimal doctor effort for a given patient effort. Third, we derive the simultaneously determined doctor and patient effort.

---

[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.

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 into the health production function. 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 $R(a) = p(a - a_L)$ conditional on performance being above a lower threshold a_L and below a higher threshold a_U, i.e. $a_L \leq a \leq a_U$.[2] The performance payment is instead zero, i.e. $R(a) = 0$, if performance is below the lower threshold, i.e. $a < a_L$. The performance payment is $R(a) = p(a_U - a_L)$ if performance is above the upper threshold, i.e. $a > a_U$.

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)).

For performance values which are above the lower threshold and below the higher threshold, the doctor's expected utility is equal to:

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

---

[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.

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

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

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 into the health production function. Analytically,

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

The equilibrium is denoted by the pair (y^*, a^*), where we assume that $a_L \leq a^* \leq a_U$, and 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 section 2.1, we conduct comparative statics with respect to the change in price p. In section 2.2, we introduce provider heterogeneity, which allows us to investigate a change in the upper threshold. Both these changes are relevant for our empirical analysis.

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

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

9

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 Comparative statics with respect to the upper threshold

We now assume that providers are heterogeneous and differ in the degree of altruism $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, which is uniformly distributed with a density function equal to one. We assume that the provider with lowest altruism has a level of effort which is above the lowest threshold, so that $a(\underline{\alpha}) > a_L$. This assumption is consistent with the data used in our empirical analysis, and allows us to focus on the effect of changing the upper threshold.

Since we are interested in the effect of varying the upper threshold, provider heterogeneity is required so that some providers are above the upper threshold a_U and others are below. Define a^{**} the equilibrium doctor effort in the absence of the price, i.e. $p = 0$ (and y^{**} as equilibrium patient effort):

$$EW_a(y^{**}, a^{**}) \equiv \alpha \varphi_a(y^{**}, a^{**}) \times (H_1 - H_0) - c_a(a^{**}) = 0. \tag{12}$$

Moreover, define $\underline{\alpha}_U$ as the level of altruism such that optimal provider effort when the price is *positive* is equal to the upper threshold, $a^*(\underline{\alpha}_U) = a_U$; and define $\overline{\alpha}_U$ as the level of altruism such that the optimal provider effort when the price is zero is equal to the upper threshold, $a^{**}(\overline{\alpha}_U) = a_U$, where $\overline{\alpha}_U > \underline{\alpha}_U$. Then, the following result holds:

Proposition 1. *If $\alpha \in [\underline{\alpha}, \underline{\alpha}_U)$, provider effort is equal to $a^*(\alpha)$; if $\alpha \in [\underline{\alpha}_U, \overline{\alpha}_U]$, provider effort is equal to the upper threshold a_U; and if $\alpha \in (\overline{\alpha}_U, \overline{\alpha}]$ then provider effort is equal to $a^{**}(\alpha)$.*

The proof is in Appendix A. Figure 1 illustrates. Providers with relatively low altruism, i.e. $\alpha \in [\underline{\alpha}, \underline{\alpha}_U)$, are motivated by the positive price incentive to provide the higher level of effort $a^*(\alpha)$, relative to the lower effort $a^{**}(\alpha)$ when the price is absent. Providers with intermediate levels of altruism, i.e. $\alpha \in [\underline{\alpha}_U, \overline{\alpha}_U]$, bunch at the upper performance threshold a_U: the price incentive is strong enough to induce them to provide effort at the upper threshold but no more. Providers with highest altruism, i.e. $\alpha \in (\overline{\alpha}_U, \overline{\alpha}]$, have incentive to provide effort above the upper threshold despite the lack of the price incentive due to their altruistic

concern.

Figure 2 illustrates the change of an increase in the upper threshold a_U. Analytically, both $\underline{\alpha}_U$ and $\overline{\alpha}_U$ increase (as $d\underline{\alpha}_U/da_U = 1/(\partial a^*/\partial \alpha) > 0$ and $d\overline{\alpha}_U/da_U = 1/(\partial a^{**}/\partial \alpha) > 0$). Providers who, before the change, were bunching at the old upper threshold now increase their effort to a level strictly above the old threshold (and either below or at the new upper threshold). Following the increase in the threshold, more providers respond to the price since additional activity is rewarded with additional revenues. Moreover, providers now bunch at a higher level of performance.

Figure 3 illustrates the case of a simultaneous increase in the price p and the upper threshold a_U. The effect is similar to Figure 2, but in addition providers with low altruism also increase the effort.

2.3   The purchaser of health services

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

$$EW(y,a) \geq 0, \tag{13}$$

and a limited liability constraint of the provider:

$$\omega + p(a - a_U) - c(a) \geq 0. \tag{14}$$

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, i.e. when $\omega + p(a - a_U) = c(a)$. 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:

$$EU(y,a) = \varphi(y,a)U(H_1, I(H_1) - c(a)) + [1 - \varphi(y,a)]U(H_0, I(H_0) - c(a)) - g(y). \tag{15}$$

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

$$EU_y(y^f, a^f) \equiv \varphi_y(y^f, a^f) \times \Delta U(a^f) - g_y(y^f) = 0, \tag{16}$$

$$EW_a(y^f, a^f) \equiv \varphi_a(y^f, a^f) \times \Delta U(a^f) - EU_I(y^f, a^f)c_a(a^f) = 0, \tag{17}$$

where

$$\Delta U(a^f) := [U(H_1, I(H_1) - c(a^f)) - U(H_0, I(H_0 - c(a^f)))], \tag{18}$$

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

$$EU_I(y^f, a^f) := \left\{ \varphi(y^f, a^f)U_I(H_1, I(H_1) - c(a^f)) + [1 - \varphi(y^f, a^f)]U_I(H_0, I(H_0) - c(a^f)) \right\} \tag{19}$$

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,

$$EW_a(y, a^*) \equiv p + \alpha\varphi_a(y, a^*) \times (H_1 - H_0) - c_a(a^*) = 0, \tag{20}$$

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:

$$p^f = \varphi_a(y^f, a^f) \times \left[\Delta U(a^f) - \alpha(H_1 - H_0) \right] + c_a(a^f) \times \left\{ 1 - EU_I(y^f, a^f) \right\}. \tag{21}$$

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 the provider less, 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 U.K. 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 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). More details are available in Appendix C.

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 carried out). 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

---

[3]The data is freely available at: https://qof.digital.nhs.uk.

[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.

practices in the U.K. 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 1st April 2004 with the first measurement of performance taking place on 31st March 2005. Practices are rewarded on the basis of their performance on a number of indicators. 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", $a \in [0, 100]$[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 (Guthrie et al. (2006)).

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

---

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

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 percentage points 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 illustrate the marginal revenue function arising from changes in achievement thresholds and prices in Figure 4. The practice marginal revenue function increases for achievement rates between the new lower and upper thresholds in 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.

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 potential change in revenue as the change in performance payments for doctor's effort, with her effort fixed at the 2004 level. Formally for each indicator ι, $R_\iota^{04}(a^{04}, a_L^{04}, a_U^{04}) - R_\iota^{04}(a^{04}, a_L^{06}, a_U^{06}) = \min\left\{1, \max\left\{\frac{(a^{04}-a_L^{04})}{(a_U^{04}-a_L^{04})}, 0\right\}\right\} - \min\left\{1, \max\left\{\frac{(a^{04}-a_L^{06})}{(a_U^{06}-a_L^{06})}, 0\right\}\right\}$. R represents the revenue for disease control in which each indicator is weighted by the price per point and the proportion of total points available, $p_\iota^t * \frac{\pi_\iota^t}{\bar{\pi}^t}$. So, $\Delta R = \left\{\frac{[\sum_{\iota=1}^k p_\iota^{04} * \pi_\iota^{04} * R_\iota^{04}(p^{04}, a^{04}, a_L^{04}, a_U^{04})]}{\bar{\pi}^{04}}\right\} - \left\{\frac{[\sum_{\iota=1}^k p_\iota^{06} * \pi_\iota^{06} * R_\iota^{04}(p^{06}, a^{04}, a_L^{06}, a_U^{06})]}{\bar{\pi}^{06}}\right\}$. Intuitively, ΔR is the difference between doctor's remuneration in 2004 (evaluated at 2004 prices, points and thresholds) for her level of effort in 2004 and the hypothetical doctor's remuneration in 2004 (evaluated at 2006 prices, points and thresholds) for her level of effort in 2004. As an illustrative example, assume that practice j's achievement in indicator CHD8 in 2004 is $a^{04} = 65$. In other words, 65% of patients with CHD have had a measurement of cholesterol below 7mmol/l. This value is above 2004 upper threshold but below 2006 upper threshold. Following Table 2, $R_\iota^{04}(a^{04}, a_L^{04}, a_U^{04}) = p^{04} * 1 * 16$ and $R_\iota^{04}(a^{04}, a_L^{06}, a_U^{06}) \equiv p^{06} * \frac{65-40}{70-40} = p^{06} * 0.833 * 17$. Therefore, the potential change in revenue weighted by the number of points for CHD8 is $\Delta R_\iota = p^{04} * (16 * 1) - p^{06} * (17 * 0.83) =$

17

$(p^{04} * 16) - (p^{06} * 14.11)$[7].

We calculate this change in revenue for each indicator within disease control and we weight it by the proportion of total points available for disease control. This change in revenue will be positive and vary between zero and one. A graphical representation of this potential change in revenue aggregated for all four Disease Control indicators is provided in Figure 5. This is different from the marginal revenue function depicted in Figure 4 because Figure 5 represents the *potential* change in revenue that *would* arise if doctors performed as well in 2006 as they did in 2004, it does not represent what the *actual* change in revenue was. On the horizontal axis of Figure 5, we report the proportion of patients to whom the indicator applies - 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 potential change in revenue between 2006 and 2004, ΔR. This potential change in revenue is zero for practices with achievement below the 2004 lower threshold because they will not be remunerated for this level of performance either in 2004 or in 2006[8]. If the achievement rate is between 2004 and 2006 lower thresholds, the potential change in revenue 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 potential change in revenue then becomes progressively smaller for rates of achievement between 2006 lower threshold and 2004 upper threshold. This potential change in revenue 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 the 2006 upper threshold because practices achieve the maximum points both in 2004 and in 2006.

The potential change in revenue reported in Table 3 is on average positive as practices could avoid a loss in revenue in 2006 compared to 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 the 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 B.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

---

[7]In monetary terms, with the price per point for a practice with average case-mix changing from p^{04}=£75 in 2004 to p^{06}=£126 in 2006, a practice with average case-mix would lose £578 on indicator CHD8. As the increase in price per point is common across all indicators, we do not monetise the potential revenue in the regression models and we express the incentive as indicated in Table 2 and the example above, but without p^t.

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

threshold, ii) between the lower and the upper threshold, and iii) above the upper threshold in 2004 and in 2006, respectively. The proportion of practices linked to respondents in ELSA that fall in each threshold interval is very similar to that of England. In Table B.2 we report the points weighted average achievement rates and the potential change in revenue 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 potential change in revenue in the last row of Table 3 do not differ much from those in Table B.2.

4 Empirical Strategy and Results

4.1 OLS 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 Eq. (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 $t = 2004; 2006$:

$$y_{ijt} = \beta_0 x_{ijt} + \gamma_0 a_{jt} + c_{ij} + \epsilon_{ijt} \tag{22}$$

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:

$$\Delta y_{ij} = \beta_0 \Delta x_{ij} + \gamma_0 \Delta a_j + \Delta \epsilon_{ij} \tag{23}$$

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[9]. As a result, patient i registered in practice j time-invariant unobserved

---

[9]We account for the uncertainty of having 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. That is, practices with larger patient populations (i.e. practice j's "list size") are weighted more than practices with a smaller population. For instance, if patient i is matched to 5 different practices (one with 12,791 patients, one with 8,186 patients, one with 4,498 patients, one with 3,236 patients and one with 2,968 patients), then this patient will be observed 10 times. The duplication will account for the practice population and we will observe patient i in the practice with the highest population four out of the 10 times ($weight = [12,791/(2,968 + 4,498 + 12,791 + 3,236 + 8,186)] = 0.40$).

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 doctor's effort by increasing (reducing) her effort and $\gamma_0 > 0$ ($\gamma_0 < 0$).

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 between changes in cigarette and alcohol consumption, and physical activity between 2004 and 2006 to be below one percent. 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.

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. However, this differential effect might be due to the direction of the bias induced by doctor's expectation of patient's effort.

There are two reasons why estimates of γ_0 might be biased. First, patient's response to doctor's effort is generated by the doctor's anticipation of the patient behaviour. Although we use an aggregate proxy for doctor's effort (i.e. the practice achievement rate), on average there might be unobserved (to us, but not to the average doctor) patient characteristics (e.g. patient's lack of self-control towards unhealthier lifestyles) that imply a reverse causal relation between doctor and patient efforts. This bias could be severe in large practices serving a diverse population where (unobserved) attitudes towards health behaviours may vary within the same practice. A second issue is that doctor's effort is essentially unobservable. Instead, we proxy for it using the points weighted average achievement rates for disease control, that is, the practice production function. As this is determined by a variety of inputs, including doctor and patient efforts, and other practice characteristics, we essentially face a form of measurement error that biases estimates of γ_0 when we use points weighted average achievement rates for disease control as a proxy for doctor's effort.

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. We note that our results are robust to the exclusion of patients matched to multiple practices. We can make these results available on request.

4.2 Instrumental Variable models

The issues of reverse causality, omitted factors and measurement error motivate our instrumental variable approach. In other words, our aim is to isolate variations induced by doctor's effort from all other potential determinants of the practice disease control achievement. To do so, we use the potential change in revenue induced by modifications of the QOF incentive as an instrument for the practice disease control achievement. 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 potential change in revenue (Eq. (24)). We then include this predicted effort by the doctor in the second stage equation of the change in lifestyle behaviours (Eq. (25)).

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:

$$\Delta a_j = \zeta \Delta R_j + \lambda \Delta x_{ij} + \Delta \eta_{ij} \tag{24}$$

where ΔR_j is the instrument and measures the potential change in revenue 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 potential change in revenue as defined in Eq. (9). We argue that this instrument is valid since, even if patients were aware of the price and threshold changes (which is unlikely) they would not be aware of how these changes affected specific practices or health care indicators.

In addition to the socio-economic characteristics described above, Δx_{ij} includes the Low Income Scheme Index (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. (23), but with the inclusion of the first-stage predicted doctor effort:

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \gamma_1 \Delta \hat{a}_j + \Delta u_{ij} \tag{25}$$

Patient's causal response to doctor's effort in Eq. (3) 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 potential change in revenue. 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.

As an initial step in our 2SLS approach, we examine the power of the instrument and find that the potential change in revenue is a strong predictor of changes in the average doctor's disease control effort[10] as reported in Table 5. A one standard deviation change in the revenue shock results in a 0.69 standard deviation increase in disease control[11]. This means that the higher the avoided loss for practice j, i.e. the more powered the incentive that the practice is facing, the more it would do to manage its patients' disease. This positive relation between the average doctor's effort and the potential change in revenue is the one we expected from Figure 3. 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 potential change in revenue 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 as reported in Table 5.

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 0.05 points of a category and a reduction in cigarette 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). The magnitude of this effect, corresponding to about 12% of the mean, is smaller than previous research by Fichera et al. (2016) who find an 18% decline of the mean of cigarette smoking. However, the study by Fichera et al. (2016) focused on the introduction of the QOF in 2004, whose effect had already been shown to be larger than that of later years (see for example, Doran et al. (2011)) and did not investigate any specific intervention.

Finally, we examine the comparative statics derived in the theory section in Eq. (10) and estimate a reduced form model with patient's lifestyle response as a function of the potential change in revenue induced

---

[10]We run the first stage regressions separately on each sample of lifestyle behaviours for which data on lifestyle behaviours are available.

[11]This is calculated as the ratio of the standard deviations of disease control and the potential change in revenue, respectively, times ζ , that is, $\zeta^{'} = (0.077/0.062) * 0.56$.

[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. $(83 - 76) * 0.007$) and smoking (i.e. $(83 - 76) * 0.24$).

by the QOF price:

$$\Delta y_{ij} = \vartheta \Delta R_j + \lambda \Delta x_{ij} + \Delta \eta_{ij} \tag{26}$$

given that we find $\vartheta > 0$, indicating complementarity between patient and doctor efforts ($\varphi_{ya}(y^*, a) > 0$), we also expect that $\gamma_1 > 0$ because a change in patient's lifestyle behaviour can only occur through doctor's effort. In other words, we expect that patient's health behaviour responds to a potential change in revenue induced by changes in the QOF prices, i.e. a version of Eq. (10). We report ϑ estimates 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 potential change in revenue of as much as 5%, there would be a decrease of drinking equal to 0.02 points of a category and a decrease of 0.7 cigarettes per day[13].

4.2.1 Identification

In order for γ_1 in Eq. (25) to isolate only variations in doctor's effort and to represent a Local Average Treatment Effect (LATE), we assume internal validity of our instrument[14]. This is essentially untestable, but we make the following arguments in support to our identification strategy.

Assumption 1: (Conditional) Independence

$$\Delta R_j \perp [\Delta y_{ij}(\Delta R_j, \Delta a_j), \Delta a_j(\Delta R_j)] \,|\, \Delta x_{ij}$$

This assumes that the potential change in revenue is as good as randomly assigned. It implies that the potential change in revenue is orthogonal to changes in all *individual* patient determinants of lifestyle behaviours and *practice* disease control determinants, *conditional* on some (vector of) covariate(s) Δx_{ij}.

There are at least three scenarios that would violate this assumption. First, compositional patient selection could occur if there is evidence that individual health behaviours and characteristics differ between high and low achieving practices. Evidence of imbalance in the patient mix across best and worse practices might indicate selection on unobservables. We attempt to mitigate this concern by controlling for a rich set of patient characteristics and for the Low Income Scheme Index (LISI), a deprivation measure of practice

---

[13]These are calculated as the average potential change in revenue of practice j times the average change in patient i's alcohol drinking (i.e. $(5) * 0.004$) and smoking (i.e. $(5) * 0.13$).

[14]We will return to the issue of external validity in the discussion.

j's list of patients. In other words, the independence assumption needs only to hold *conditional* on these characteristics. However, we also test whether patient mix (in terms of behaviours and other characteristics) is balanced across low and high achieving practices. To do so, we define best practices as those with 2004 points weighted average achievement rates above the median of that year and estimate the following model:

$$z_{ij} = \xi I_j + \beta_1 x_j + \epsilon_{ij} \tag{27}$$

where z_{ij} represents separate outcomes: either patient health behaviours (i.e. physical activity, drinking and smoking), y_{ij}, or patient employment and wealth, w_{ij}; I_j is a dummy variable equal to one if practice j points weighted average achievement rates in 2004 are on or above the 2004 median and zero otherwise; x_j represents the LISI of practice j; ϵ_{ij} is an error term. Eq. (27) has been estimated in 2004 for the sample of individuals diagnosed with CVD. Our coefficient of interest is ξ. Our hypothesis is that there should not be any statistically significant evidence of differences between best and worse practices in patient health behaviours and characteristics amongst those who have been diagnosed with CVD in 2004 (i.e. $\xi = 0$). Indeed, we do not find any statistically significant evidence of patient selection as individual health behaviours and characteristics do not differ between high and low achieving practices (Table 8). We only find a very weakly statistical significant evidence that in high achieving practices patients are more likely to be employed.

A second indication that conditional independence might be violated arises from heterogeneity in the marginal product of doctor's effort on patient health behaviours across low and high achieving practices. Thus, even if doctor's effort is the same across best and worse practices, patient health behaviours might respond to the change in price because of the differential effect of doctor's effort. We first interact the indicator of best practices, I_j, separately with the change in disease control achievement and with our instrument. We then test whether there is a statistically significant response of patient's effort on doctor's effort in high versus low achieving practices. In the top panel of Table 9 there is evidence that reduction in cigarette smoking is stronger in practices with achievement rate below the 2004 median. However, this is not surprising because the incentive payment was such that those who were already performing well in 2004 had to do less in order to get a higher payment.

Third, practices may vary in the way they respond to the potential change in revenue 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, we modify Eq. (25) and include the initial performance of the practice in 2004:

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \gamma_1 \Delta \hat{a}_j + \gamma_2 a_j^{04} + \Delta u_{ij}. \tag{28}$$

In the bottom panel of Table 9 we show 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 0.02 points of a category and a reduction of one cigarette per day per smoker[15]. Baseline performance is statistically significant in the first stage, but it is not displayed on the table. Given that we checked whether patient characteristics and behaviours were balanced between best and worse practices, the smaller size of our coefficient might reflect the fact that best practices had to do less to avoid a larger loss. Still, our results show complementarity between drinking and smoking and doctor effort. Nevertheless, we take this result to be the low bound estimate equivalent to about 7% of the mean number of cigarettes in our sample and almost 1% of the mean frequency of alcohol drinking.

Assumption 2: Exclusion Restriction

$$[\Delta y_{ij}(\Delta R_j, \Delta a_j) = \Delta y_{ij}(\Delta a_j)] \mid \Delta x_{ij}$$

This assumes that doctor's effort is the only channel through which the potential change in revenue affects patient health behaviours. Just like for the independence assumption, it can hold *conditional* on some vector of covariates Δx_{ij}.

Checking this assumption implies thinking about the underlying mechanisms (other than doctor's effort) that could drive the association between practice achievement rate and patient health behaviours. To do

---

[15]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. $(83 - 76) * 0.003$) and smoking (i.e. $(83 - 76) * 0.14$).

so, we need to show that these other potential mechanisms do not explain our results, *conditional* on the measure for practice deprivation and other patient characteristics. A first simple check on this is to look for an association between the potential change in revenue and patient characteristics that should not be affected by doctor's effort. To do so, we estimate the same 2SLS models as in Eq. (25) and replace the outcome variables with the change in non single marital status and in household size since these are clearly unrelated to doctor effort. Our null hypothesis is that $\gamma_1 = 0$. Indeed, in Table 10 we report no evidence that the change in disease control induced by the QOF has affected patient characteristics (i.e. change in marital status, and household size).

Another useful check is to look for an association between the potential change in revenue and outcomes in samples where we would not expect doctor's effort to affect them. Our 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 change in doctor's effort induced by the potential change in revenue for CVD indicators. We check this in three ways. First, we estimate the same 2SLS models as in Eq. (25) where $\Delta \hat{a}_j$ represents the practice points weighted achievement rates for cervical screening[16]. 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. We use this change in threshold to construct the new instrument, ΔR_j, and re-estimate the 2SLS in Eqs. (25 and 24) but on the sample of women. If the change in patient health behaviours is solely attributed to the change in doctor's effort induced by the change in CVD incentives and not to a general practice improvement, then we should expect $\gamma_1 = 0$. Test A in Table 11 shows that cervical screening, whose payment thresholds changed in the 2006 QOF, has not affected women's lifestyle behaviours[17]. Second, we re-estimate the 2SLS models in Eqs. (25 and 24) on the sample of individuals without CVD. Again, our null hypothesis is that $\gamma_1 = 0$. Test B in Table 11 cannot reject this hypothesis indicating no statistical significant effect of doctor effort on patient effort. The size and sign of the coefficients are also very different from our main models in Table 6. Finally, 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 lifestyles. So we re-estimate the same 2SLS models as in Eq. (25) where Δy_{ij} represents other behavioural outcomes, namely changes in savings and in the value of the pension fund which again are unrelated to

---

[16]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.

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

doctor effort. As reported in Test C of Table 11, there is no statistical significant evidence that doctor effort impacts on behavioural outcomes other than lifestyles. All in all, these falsification tests provide stronger evidence that there are no channels other than doctor's effort that have driven our results.

Assumption 3: Monotonicity

$$\text{either } \zeta \geq 0 \text{ or } \zeta \leq 0$$

This means that while doctor response to the incentive may depend on the size of the potential change in revenue, all those who are affected by this potential change in revenue, respond in the same way. In our case where both the endogenous variable and the instrument are non-binary, it requires that doctor's effort is weakly increasing in the potential change in revenue, that is, doctor's effort should be larger, the higher the avoided loss for the practice.

We partition the instrument to resemble the structure of the incentive represented in Figure (1): (i) when $a_L^{04} \leq a^t < a_L^{06}$; (ii) when $a_L^{06} \leq a^t < a_U^{04}$; (iii) when $a_U^{04} \leq a^t < a_U^{06}$; and (iv) when $a^t \geq a_U^{06}$[18]. We then modify the first-stage regression in Eq. (24) as follows:

$$\Delta a_j = \zeta_1 \Delta R_j^1 + \zeta_2 \Delta R_j^2 + \zeta_3 \Delta R_j^3 + \lambda \Delta x_{ij} + \Delta \eta_{ij} \tag{29}$$

where $\Delta R_j^1, \Delta R_j^2, \Delta R_j^3$ are the values of the instrument when a^t lies within each of the ranges (i)-(iii) above. We expect that $(\zeta_1 > \zeta_2 > \zeta_3) > 0$ if complementarity holds and doctor's effort increases with the averted loss for the practice. The second stage regression is the same as Eq.(25).

The first stage regressions displayed in Table 12 highlight that doctor response to the potential change in revenue is higher, the steeper and stronger the incentive. For practices whose achievement was between the 2004 and 2006 lower thresholds, we would expect the potential change in revenue to be much higher as they would stand to lose much more than other practices (e.g. they would get no reward if their achievement in 2006 was below 2006 lower threshold). In the third column of Table 12, a one standard deviation change in the potential revenue shock ζ_1 results in a 0.30 standard deviation increase in disease control which is 30 times larger than the 0.01 standard deviation change for ζ_3 [19]. The second stage regressions displayed in Table 13 remain largely unchanged because we have only partitioned the previous instrument.

---

[18]We do not have practice with $a^t \leq a_L^{04}$.

[19]These are calculated as the ratio of the standard deviations of disease control and the potential change in revenue, respectively, times ζ_1, that is, $\zeta_1' = (0.028/0.062) * 0.67$. And for the price change ζ_3, it is $\zeta_3' = (0.000098/0.062) * 0.38$.

Only given the first-stage regression, the conditional independence assumption, the exclusion restriction and monotonicity, can the coefficient γ_1 in Eq. (25) be interpreted as the effect of doctor's effort on those whose treatment status can be changed by the new incentive. Under these assumptions, γ_1 in Eq. (25) is the LATE.

5   Robustness checks

We consider a number of robustness checks entailing the instrument, the choice of indicators, lifestyle behaviours and the specification strategy, and patient heterogeneity. Firstly, we use a more flexible specification of the instrument with the second order polynomial of the potential change in revenue in the first stage regression. When we allow for a quadratic specification of the instrument (first stage not reported on the top panel of Table 14), we find the complementarity effect in cigarette smoking to be of similar size to our main results.

Secondly, in the main models we 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 potential change in revenue. 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[20], there could be effort diversion issues as worse practices might put more effort on indicators for which no change in the incentive occurred. To do so, we modify Eq. (25) and consider achievement rates for the full set of 33 indicators. In the bottom panel of Table 14, we 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 cigarette consumption is slightly larger than 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.

Thirdly, we examine different definitions of health behaviours. We consider reduction in cigarette smoking for all those diagnosed with CVD in 2004, including former smokers. We also investigate whether comple-

---

[20]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.

mentarity between doctor and patient efforts occurs at the extensive margin (e.g. patients quit smoking) rather than intensive margin. We explore whether problem drinkers have responded to changes in doctor's effort by defining a variable "no daily drinking" as equal to one if patient i drinks frequently, rarely or never drinks and zero if she drinks daily. We modify Eq. (25) and consider Δy_{ij} as the change in each of these three outcomes. Although we find no statistically significant effect on quitting smoking, in the top panel of Table 15 we show that complementarity between doctor's effort and patient reduction of cigarette smoking holds even when we include former smokers. However, the size of the effect for an average practice increasing its rate of disease control from 76% to 83% is reduced to 0.4 cigarettes per day per patient. We also find that doctor's effort reduces daily drinking, although the effect is only weakly statistically significant.

Fourthly, we investigate whether patients who switch practices have significantly different lifestyle behaviours from those who do not move. To do so, we define "mover" as a dummy variable that equals one in 2004 if patient i changes practice in 2006 and re-estimate Eq. (25). If changes in patient mix are not related to observed patient characteristics then "movers" should not have statistically significant different lifestyle behaviours to non-movers. There is no statistically significant evidence of changes in health behaviours amongst those who will switch practice in 2006 (bottom panel of Table 15).

Finally, we consider whether patient response varies with her age and baseline wealth. To this purpose we modify Eq. (25) as follows:

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \gamma_0 \Delta \hat{a} + \gamma_1 \Delta \hat{a} * D_{ij} + \gamma_2 D_{ij} + \Delta u_{ij}. \tag{30}$$

We estimate Eq.(30) separately with the dummy D_{ij} indicating either age or wealth. In the first model we interact the change in disease control with D_{ij} that equals one if patient i's age is on or above 69 years. We also include this age dummy and the change in disease control in Eq.(30). We are interested in the coefficients γ_0 and γ_1 indicating the effect of doctor effort on patient lifestyle choices for younger and older patients, respectively. In the top panel of Table 16, we show that the reduction in cigarette smoking is stronger for relatively younger patients. It may be explained by the fact that these patients expect to benefit from the returns of healthier lifestyle behaviours over a relatively longer time horizon than older patients. It could also be due to the fact that relatively younger patients are more malleable and their lifestyle behaviours can be modified more easily than older patients. In the second model, we interact the change in disease control

with D_{ij} that equals one if patient i's logarithm of equivalised wealth in 2004 is on or above its median. We also include this wealth dummy and the change in disease control in Eq.(30). Again, we are interested in the coefficients γ_0 and γ_1 indicating the effect of doctor effort on patient behaviour for relatively less wealthy and wealthier patients, respectively. In the bottom panel of Table 16, we show that the reduction in cigarette smoking is stronger for relatively less wealthy patients.

6 Mechanisms

Why does doctor effort affect patient health behaviour? What are the "mechanisms" that generate complementarities between doctor and patient effort? Are doctors directly providing lifestyle change advice in addition to or *in lieu* of prescribing? Are they encouraging follow-ups or check-ups? NICE guidelines for CVD 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)). Whilst these guidelines became official after our period of interest, it is reasonable to suppose that doctors may have adopted best practice management in striving to meet QOF indicators in these conditions. Although the focus of this paper is not on the mechanisms, these are important questions for the policy implications of this study.

The initial waves of the ELSA data contain limited information on doctor effort. We have no information on whether doctors increased follow-ups or gave advice to reduce alcohol drinking and improve physical activity. We also have limited information on prescriptions and the type of drugs prescribed especially in the years where there is no nurse visit.

All the models we use are a re-estimation of Eq. (25) where the outcome variable denotes each of the "mechanisms" we consider. We investigate whether doctor prescription has changed using patient reports of the prescription of beta blockers, blood thinning medication, ACE inhibitors, or medication to reduce blood pressure. As these questions are asked by nurses during their visit, they are only available in 2004 and 2008. For those diagnosed with CVD prior to 2006 and who are observed both in 2006 and 2008, we assume their prescription did not change and we use the 2008 self-report prescription for 2006. We maintain the same assumption for the variable indicating whether they had blood pressure checks. We also consider an indicator which equals one if those who were smokers in 2004 received smoking cessation

advice. As reported in Table 17, we find that doctor effort increased her prescription behaviour and smoking cessation advice, although only the latter is (weakly) statistically significant[21]. We do not find evidence that blood pressure checks changed potentially due to the fact that these activities were undertaken by nurses not doctors. Previous studies such as the one by Doran et al. (2011) have shown that although practice performance reached a *plateau* after the QOF was introduced, in 2006-07 achievement rates for measurement and prescribing indicators were still higher than the projected pre-QOF trend. However, because of many sources of measurement error, we refrain from concluding that our results are in contrast to these previous studies.

7 Conclusions

The use of performance pay incentives is increasing in public sectors such as education and health care. Whilst numerous studies have investigated the effect of these incentives on targeted agents, not much is known about their effects for the interaction between 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 be enhanced or mitigated by complementary or substitute actions of related 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 and illustrative figures suggest that doctors respond to an increase in prices and thresholds in their payment system by increasing effort. The extent to which patients will increase or decrease their effort in response to these potential revenue changes 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

---

[21]Smoking cessation advice was directly incentivised by the QOF. However, we note that in 2004 smoking cessation advice was split across several indicators and disease areas and it was then aggregated from 2006 onwards. Because of this change in definition and because only the lower threshold (for which we have no practices) changed between 2004 and 2006, we do not consider this indicator.

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 0.05 points of a category and cigarette consumption by 1-1.7 cigarettes per day per smoker. The former corresponds to about 2% of the sample mean. The latter, corresponding to about 7-12% of the sample mean, is smaller than previous research by Fichera et al. (2016) who find an 18% decline of the mean cigarette smoking. However, this previous research had focused on the introduction of the QOF whose impact was notably larger (Doran et al. (2011)). We find no effect on physical activity. Our identification strategy relies on a 2SLS approach which instruments doctor's effort with an exogenous potential change in revenue induced by the QOF incentive scheme: a one standard deviation change in the potential revenue shock results in a 0.69 standard deviation increase in disease control. Our complementarity results contrast those by Kaestner et al. (2014) and Lakdawalla et al. (2006) who both found that medical innovation induces riskier health behaviours. However, our study focus is on the effect of a doctor incentive system which incentivised both information provision and treatment. Their analysis focuses on availability of cheaper treatments either because they are more easily available in the market (i.e. penetration of statins and new HIV treatments) or because changes in Medicaid eligibility allows for cheaper access to these treatments. According to these studies, medical innovation induces substitutability with healthier lifestyle behaviours, but it might not be so if doctor effort is incentivised.

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 which may, if anything, attenuate our results. A second possible limitation with our type of study relates to announcement versus implementation effects. In 2003/04, the "preparatory phase" of the QOF, details of the quality targets were in the public domain (but incentives were not yet available) and might have generated anticipatory effects. A study by Doran et al. (2011) used data on a sample of primary care practices from the General Practice Research Database (GPRD). They showed that there was already an

upward quality trend in primary care practices since 2002/03, but the introduction of the QOF improved performance over and above this trend at least in 2004, its first year of introduction. This was stronger for measurement indicators than for prescribing indicators. As the QOF was rolled out nationally, they recognise the absence of control practices makes it hard to establish that this jump in performance was entirely due to the QOF. Although Doran et al. (2011) could not investigate practice heterogeneities, we might expect that better practices were more able to make adjustments prior to the introduction of the QOF and, therefore, perform better thereafter. Any anticipation effects, or heterogeneity in such effects, would be an issue in our study if we were comparing pre and post-QOF outcomes but are less worrying since this is not the method we use. Instead, our identification is based on reactions to changes in financial incentives within the QOF between 2004 and 2006, which could have not been anticipated in the preparatory phase. Nevertheless, this is one of the reasons why we have controlled for baseline performance in one of our robustness checks and why we have examined practice heterogeneities. Our results are robust to these alternative specifications suggesting that potential anticipatory effects would not qualitatively change our results. A third potential limitation relates to the fact that our results are only relevant for all those practices that were not always under or over performing throughout the sample period. We note that this is actually the case for most of our sample of practices (over 80% of the 2006 practices). A fourth 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 cigarette 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. A final limitation is the lack of data on the mechanisms that might have brought about these complementarities. With some strong assumptions about data measurement, we have found that practice disease control is positively associated with doctor prescription and smoking cessation advice, though only the latter is (weakly) statistically significant. These results might suggest that doctors responded to the financial incentive by adopting best practice management for patients with CVD. This change in effort was then complemented by healthier patient lifestyles.

Nevertheless, our results suggest that by rewarding achievement on specific indicators this incentive scheme has also impacted on patient health behaviours because of the interaction between doctors and patients. 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 U.K. and elsewhere, and we would expect this development to lead to other empirical insights on health behaviours and their interaction with health care.

Figure 1: *Equilibrium performance*

Performance a

$a^*(\alpha)$

$a^{**}(\alpha)$

Upper
Threshold a_U

Lower
Threshold a_U

$\underline{\alpha}$ $\underline{\alpha}_U$ $\bar{\alpha}_U$ $\bar{\alpha}$

Altruism α
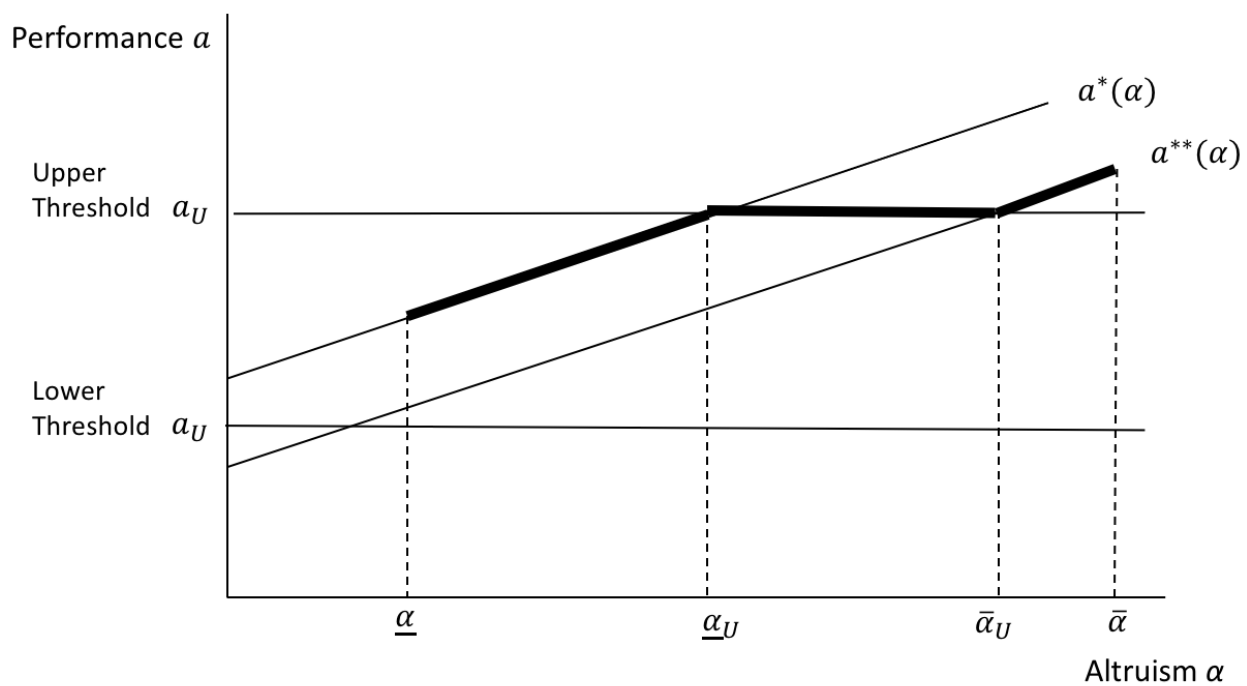
Figure 2: *Increase in upper threshold a_U*

Performance a

$a^*(\alpha)$

$a^{**}(\alpha)$

Upper
Threshold a_U

Altruism α
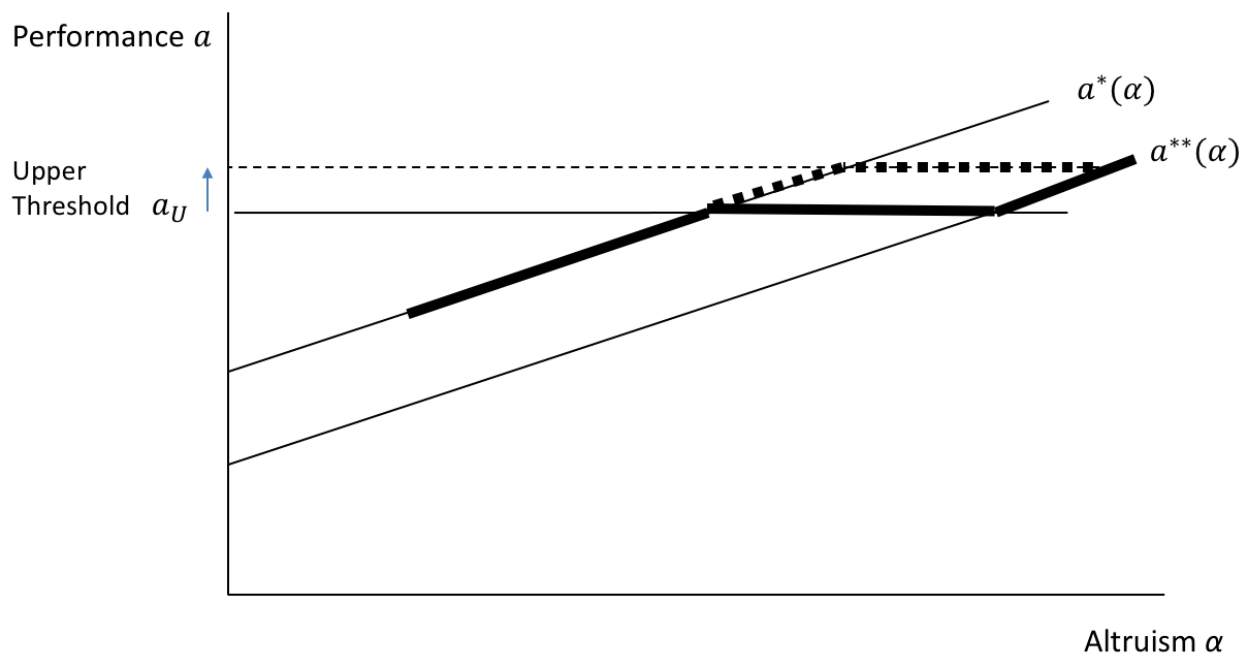
Figure 3: *Increase in upper threshold a_U and price p*

Performance

Upper
Threshold a_U

$a^*(\alpha)$

$a^{**}(\alpha)$

Altruism α
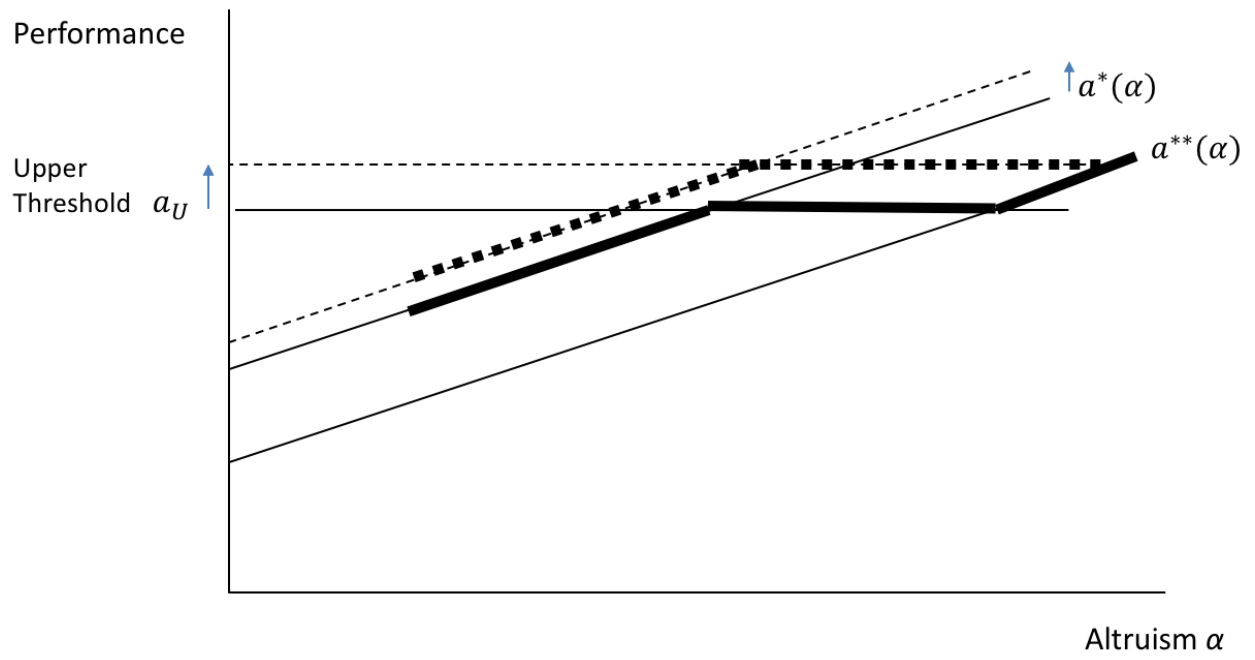
Figure 4: *Marginal revenue function: increase in price and thresholds*

p

Lower
Threshold a_L

Upper
Threshold a_U
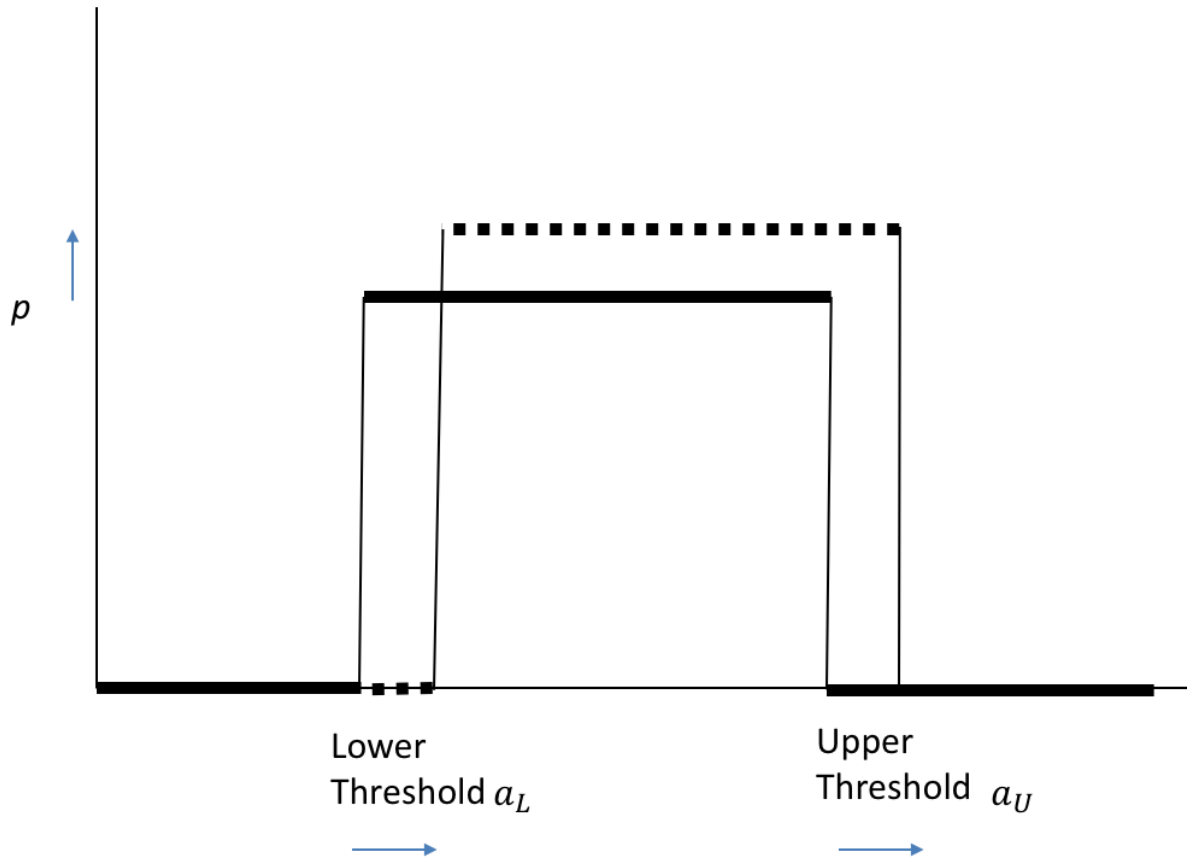
Figure 5: *The potential revenue change function*
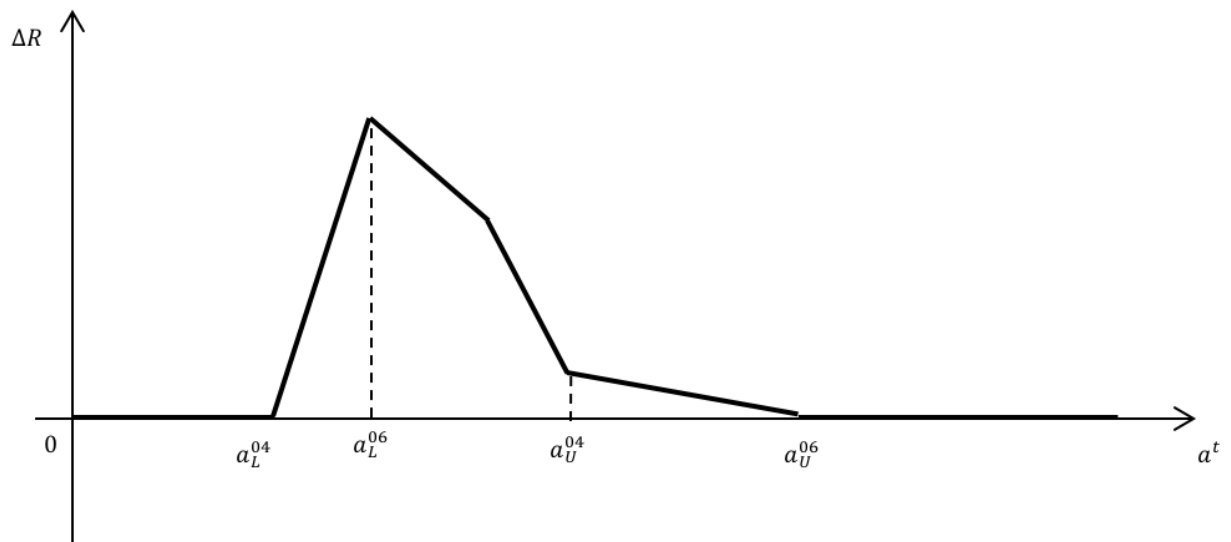
Table 1: *Definition of the variables used in the models by year*

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	11.87	1.6	11.87	1.6
	wealth in £	(2,249)		(2,249)	
Employed	Whether individual is employed vs. unemployed,	26.85%		24.41%	
	disabled, retired, home work	(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	-13.94	8.67	-12.24	9.34
	(negatively coded)	(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	11.06	2.18	11.13	2.06
	wealth in £	(238)		(238)	
Employed	Whether individual is employed vs. unemployed,	26.05%		22.27%	
	disabled, retired, home work	(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				2006			
		LT	UT	Points	Price per point	LT	UT	Points	Price per point
CHD8	The percentage of patients with coronary heart disease whose last measured 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 potential change in revenue*

	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
Potential Revenue change $[\Delta R]$	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)	0.30	-0.08	8.30
	(0.24)	(0.23)	(7.31)
Change in non single status	0.01	0.12	1.86
	(0.11)	(0.10)	(1.86)
Change in employed status	0.02	0.08*	1.34
	(0.06)	(0.04)	(2.05)
Change in household size	-0.01	0.04	-1.09
	(0.03)	(0.03)	(0.92)
Change in Ln(equivalised wealth)	0.02	0.02	-0.16
	(0.02)	(0.02)	(0.57)
Change in LISI	-0.004	-0.006	0.03
	(0.007)	(0.007)	(0.24)
Constant	-0.07***	0.07***	1.12
	(0.02)	(0.02)	(0.76)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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 potential change in revenue*

	Change in Disease Control on physical activity sample	Change in Disease Control on alcohol sample	Change in Disease Control on no. cigarettes sample
Potential revenue change (ζ)	0.52***	0.51***	0.56***
	(0.01)	(0.01)	(0.01)
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	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248
Kleibergen-Paap F-statistic	409.1	384.9	171.3

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)	0.16	0.71*	23.74**
	(0.39)	(0.40)	(10.25)
Change in non single status	0.01	0.12	1.97
	(0.11)	(0.10)	(2.02)
Change in employed status	0.02	0.08*	1.26
	(0.06)	(0.04)	(1.99)
Change in household size	-0.006	0.04	-1.10
	(0.03)	(0.03)	(1.02)
Change in Ln(equivalised wealth)	0.02	0.02	-0.15
	(0.02)	(0.02)	(0.59)
Change in LISI	0.004	0.01	0.06
	(0.007)	(0.01)	(0.24)
Constant	-0.05*	0.01	-0.14
	(0.03)	(0.03)	(0.94)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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 potential change in revenue*

	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Potential revenue change (ϑ)	0.09	0.37*	13.14**
	(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	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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 compositional patient selection*

Test A. Health behaviours	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
High achieving practices in 2004 (ξ)	0.01	-0.02	-0.23
	(0.01)	(0.03)	(0.96)
Constant	-0.07**	0.06	0.49
	(0.03)	(0.03)	(1.15)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

Test B. Patient characteristics	Model I: Employed	Model II: Ln(equivalised wealth)
High achieving practices in 2004 (ξ)	0.01*	-0.03
	(0.01)	(0.02)
Constant	0.26	12.5
	(0.01)	(0.03)
No. observations	29,028	29,028
No. individuals	2,904	2,904
No. practices	1,824	1,824

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. All models include the LISI. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 9: *Tests of practice heterogeneity using second stage models*

	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Test A. Heterogeneity of treatment and IV:			
Change in Disease Control	0.47	1.20*	42.49**
	(0.68)	(0.72)	(18.49)
Change in Disease Control for best practices	2.25	-1.21	-30.75
	(4.01)	(3.77)	(57.2)
High achieving practices in 2004 (ξ)	-0.06	0.11	3.84
	(0.19)	(0.18)	(3.34)
Constant	-0.10	-0.05	-2.83
	(0.08)	(0.08)	(2.32)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248
Test B. Including Disease Control in 2004:			
Change in Disease Control (γ_1)	0.13	0.31***	13.60***
	(0.31)	(0.10)	(3.38)
Constant	-0.05**	0.04***	0.69**
	(0.03)	(0.01)	(0.30)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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, wealth and LISI. $^*p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 10: *Tests using second stage models with patient characteristics as outcomes*

	Model I: Change in non single status	Model II: Change in household size
Change in Disease Control	-0.04	0.24
	(0.08)	(0.23)
Change in non single status	-	0.87
		(0.14)
Change in employed status	-0.0004	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)
Change in LISI	-0.001	0.0001
	(0.001)	(0.01)
Constant	-0.01	-0.05***
	(0.01)	(0.02)
No. observations	22,471	22,471
No. individuals	2,249	2,249
No. practices	1,505	1,505

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 11: *Tests using second stage models with other outcomes and samples*

Test A. Alternative indicator	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Change in Cervical Screening	0.74	-0.51	20.70
	(0.98)	(1.00)	(34.04)
Constant	-0.05***	0.10***	1.30**
	(0.02)	(0.02)	(0.57)
No. observations	12,231	11,295	1,399
No. individuals	1,224	1,130	140
No. practices	1,047	965	150

Test B. Non-CVD patients	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Change in Disease Control	0.02	0.25	-12.33
	(0.40)	(0.33)	(15.24)
Constant	-0.04	0.02	2.73**
	(0.03)	(0.03)	(1.16)
No. observations	20,182	18,007	2,350
No. individuals	1,313	1,802	235
No. practices	2,020	1,193	233

Test C. Other behaviours	Model I: Change in Savings	Model II: Change in Pension
Change in Disease Control	-8.03	-229.74
	(19.92)	(213.44)
Constant	0.45	15.08
	(2.03)	(13.38)
No. observations	26,215	18,497
No. individuals	2,624	1,852
No. practices	1,674	1,369

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, wealth and LISI. Savings and pension values divided by £1,000. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 12: *First stage coefficients of first differenced linear models of practice achievement on the partitioned potential change in revenue*

	Change in Disease Control on physical activity sample	Change in Disease Control on alcohol sample	Change in Disease Control on no. cigarettes sample
Potential revenue change (ζ_1)	0.86***	0.82***	0.67***
	(0.04)	(0.04)	(0.04)
Potential revenue change (ζ_2)	0.48***	0.47***	0.54***
	(0.01)	(0.01)	(0.01)
Potential revenue change (ζ_3)	0.23***	0.23***	0.38***
	(0.01)	(0.01)	(0.05)
Constant	0.05***	0.05***	0.05***
	(0.0004)	(0.004)	(0.001)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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 include marital status, whether employed household size, wealth and LISI. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 13: *Second stage coefficients of first differenced linear models of health behaviours (from partitioned IV models)*

	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Change in Disease Control (γ_1)	0.17	0.75*	23.69**
	(0.38)	(0.41)	(10.58)
Change in non single status	0.01	0.12	1.97
	(0.11)	(0.10)	(2.02)
Change in employed status	0.02	0.08*	1.26
	(0.06)	(0.04)	(1.99)
Change in household size	-0.01	0.04	-1.10
	(0.03)	(0.03)	(1.02)
Change in Ln(equivalised wealth)	0.02	0.02	-0.15
	(0.02)	(0.02)	(0.59)
Change in LISI	0.004	0.01	0.06
	(0.01)	(0.01)	(0.24)
Constant	-0.06	0.01	-0.13
	(0.03)	(0.03)	(0.96)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

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 14: *Tests of robustness using second stage models with different specifications*

	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Test A. Allowing for non-linearity:			
Change in Disease Control (γ_1)	0.14	0.65*	23.46**
	(0.39)	(0.39)	(10.15)
Constant	-0.05*	0.02	-0.11
	(0.03)	(0.03)	(0.93)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248
Test B. Including All CVD indicators:			
Change in Disease Control (γ_1)	-0.27	0.40	28.00**
	(0.47)	(0.45)	(13.20)
Constant	-0.04	0.04	0.13
	(0.03)	(0.03)	(0.87)
No. observations	21,244	19,632	2,262
No. individuals	2,135	1,973	227
No. practices	1,410	1,305	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, wealth and LISI. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 15: *Tests of robustness using second stage models with different health behaviours*

Test A. Alternative health behaviours	Model I: Reduction in no. cigarettes	Model II: Quit smoking	Model III: No daily drinking
Change in Disease Control (γ_1)	5.03**	0.33	0.55*
	(2.35)	(0.29)	(0.32)
Constant	-0.19	-0.004	-0.01
	(0.13)	(0.02)	(0.02)
No. observations	24,163	2,929	17,393
No. individuals	2,428	293	1,747
No. practices	1,538	293	1,198

Test B. Movers	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Change in Disease Control (γ_1)	0.17	0.72*	23.71**
	(0.39)	(0.40)	(10.17)
Mover	0.07	0.08	-2.05
	(0.07)	(0.07)	(2.83)
Constant	-0.13	-0.07	1.81
	(0.08)	(0.08)	(2.74)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248

Statistics weighted by the constructed sampling weights and clustered std. errors in (). Test A: Model I includes former; smokers Model III excludes never drinkers. Second stage regressions displayed. All models include marital status, whether employed, household size, wealth and LISI. $^*p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 16: *Tests of robustness using second stage models with patient heterogeneity*

	Model I: Increase in physical activity	Model II: Reduction in alcohol drinking	Model III: Reduction in no. cigarettes
Test A. Age:			
Change in Disease Control	0.20	0.76	34.79**
	(0.55)	(0.58)	(15.18)
Change in Disease Control for older patients	-0.06	-0.09	-27.80
	(0.78)	(0.80)	(18.26)
Older patients	-0.09	0.03	1.26
	(0.06)	(0.06)	(1.80)
Constant	-0.01	-0.003	-0.64
	(0.04)	(0.05)	(1.42)
No. observations	22,467	20,782	2,379
No. individuals	2,249	2,080	238
No. practices	1,505	1,400	248
Test B.Wealth:			
Change in Disease Control	0.05	1.14*	20.95**
	(0.60)	(0.62)	(10.66)
Change in Disease Control for wealthier patients	0.27	-0.81	14.62
	(0.78)	(0.80)	(27.90)
Wealthier patients	0.02	0.08	-1.89
	(0.02)	(0.06)	(0.59)
Constant	-0.05	-0.03	0.29
	(0.05)	(0.05)	(1.03)
No. observations	22,436	20,753	2,369
No. individuals	2,249	2,080	238
No. practices	1,497	1,391	246

Statistics weighted by the constructed sampling weights and clustered std. errors in (). Second stage regressions displayed; All models include marital status, whether employed, household size, wealth and LISI. $^*p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 17: *Second stage coefficients of first differenced linear models of mechanisms*

	Model I: Doctor prescription	Model II: Smoking cessation advice	Model III: Blood pressure check
Change in Disease Control (γ_1)	0.44	0.21*	-0.03
	(0.28)	(0.12)	(0.20)
Change in non single status	-0.03	-0.01	0.004
	(0.07)	(0.01)	(0.08)
Change in employed status	0.03	-0.01	0.03
	(0.03)	(0.01)	(0.03)
Change in household size	-0.02	-0.004	0.05
	(0.03)	(0.003)	(0.05)
Change in Ln(equivalised wealth)	-0.01	-0.004	0.01
	(0.01)	(0.003)	(0.01)
Change in LISI	0.003	0.003	-0.01**
	(0.005)	(0.002)	(0.003)
Constant	0.21***	-0.03*	0.01
	(0.02)	(0.02)	(0.02)
No. observations	26,215	2,379	14,006
No. individuals	2,624	238	1,402
No. practices	1,674	248	1,025

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 Proof of Proposition 1

If provider altruism is below $\underline{\alpha}_U$, then the provider is always better off by providing $a^*(\alpha)$ rather than $a^{**}(\alpha) = a^*(\alpha, p = 0)$ since, by the envelop theorem, the provider is always better off with a positive price than with a zero price, $\partial EW(y^*, a^*)/\partial p = (a^* - a_L) > 0$.

If provider altruism is strictly above $\underline{\alpha}_U$, then effort $a^*(\alpha)$ is no longer optimal since the price p drops to zero for any effort above $a^*(\underline{\alpha}_U) = a_U$. The provider may have an incentive to provide the upper threshold a_U. However, this will be optimal only if provider utility from providing the upper threshold a_U is higher than provider utility when providing $a^{**}(\alpha)$, i.e. the optimal effort in the absence of a positive price. To compare the two solutions define $\Delta(\alpha) = EW(y_U, a_U) - EW(y^{**}, a^{**}(\alpha))$ as the difference in provider utility in the two scenarios. Using the envelop theorem, we obtain

$$\frac{\Delta(\alpha)}{\partial \alpha} = [\varphi(y_U, a_U) - \varphi(y^{**}(\alpha), a^{**}(\alpha))] (H_1 - H_0).$$

This suggests that $\Delta(\alpha) > 0$ if $a_U > a^{**}(\alpha)$, $\Delta(\alpha) < 0$ if $a_U < a^{**}(\alpha)$ and $\Delta(\alpha) = 0$ if $a_U = a^{**}(\alpha)$. Recall that we have defined $\overline{\alpha}_U$ as $a^{**}(\overline{\alpha}_U) = a_U$. It therefore follows that a_U is the optimal solution if $\alpha \in [\underline{\alpha}_U, \overline{\alpha}_U]$ and $a^{**}(\alpha)$ is the optimal solution if $\alpha \in (\overline{\alpha}_U, \overline{\alpha}]$. In words, a_U is optimal if the optimal effort in the absence of the price is below the upper threshold.

B Additional results

Table B.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	2,675	5,609	8,287	1	666	1,978	2,645
	(0.04%)	(32.28%)	(67.68%)		(0.04%)	(25.18%)	(74.82%)	
2006/07	1	2,077	6,279	8,357	0	429	2,093	2,522
	(0.01%)	(24.85%)	(75.13%)		(0.00%)	(17.00%)	(83.01%)	

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 B.2: *Summary statistics on the points weighted average achievement rates and potential revenue 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
Potential Revenue change $[\Delta R]$	0.06 (8,055)	0	0.40	0.08

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

C Data Linkage

We use the list of practices from the National Health Service Quality Management and Analysis System (QMAS) database available at http://www.gpcontract.co.uk/download for which we have information on the Quality and Outcomes Framework (QOF). There are 8,287 practice in 2004, 8,357 in 2006 and 8,229 in 2008. The original sample of ELSA respondents contained 9,432 and 9,805 respondents in the second and third wave, respectively, and over 11,000 in the fourth wave (see Table C.1). An anonymised list of ELSA respondents with the practice name, postcode and addresses was obtained by NatCen. From this data we have 7,332 respondents (out of the original 7,666 respondents) in wave 2 and 8,138 (out of the original 8,643 respondents) in wave 4 (the waves with nurse visits).

Table C.1: *ELSA-GP practices linkage samples*

	Wave 2 (2004/05)	Wave 3 (2006/07)	Wave 4 (2008/09)
Total no. practices in QMAS data	8,287	8,357	8,229
Original ELSA sample:			
No. respondents	9,432	9,805	11,050
Respondents at nurse visit	7,666	N/A	8,643
Respondents in ELSA-practices data	7,332	N/A	8,138
Respondents after imputation	9,432	9,168	11,050
Respondents with matched GP address	7,694	6,680	8,828
Respondents with uniquely matched GP address	7,232	6,313	8,356
Respondents with multiple matched GP addresses	462	367	472
Respondents with un-matched GP address	1,738	2,488	2,222
ELSA sample of CVD patients:			
Practices matched to ELSA respondents	2,766	2,584	3,131
Practices uniquely matched to ELSA respondents	2,620	2,439	2,967
Multiple practices matched to ELSA respondents	146	145	164
Respondents per practice:			
Mean	1.12	1.08	1.10
Std. Dev.	0.77	0.57	0.71
Min.	1	1	1
Max.	13	13	16

Note: weighted statistics.

We then define the following mapping programme. First, we adopt exact matching of practices by their postcode. When the latter is missing, we use a fuzzy matching code that matches the strings of practice addresses and names in the ELSA data to the ones in the QMAS database. Imputation of practices that are

missing for ELSA respondents is performed in two stages. At the first stage, we impute missing practices for ELSA respondent i by using the practice the partner of i's respondent is registered with. At the second stage, we impute missing practices from the ones available for respondent i in the previous or subsequent waves. As a result, out of the original sample of respondents from waves 2 and 4, we identify 9,168 of respondents in wave 3 who were present both in wave 2 and 4 and did not change practice. We need to impute this sample in wave 3 from the nurse visits in waves 2 and 4, because it was in 2006 that the change in the incentive scheme occurred. The combined two-stages approach allows us to match more individuals than the ones originally given us by NatCen. We match the practice address for 7,694 respondents in wave 2 (against the original 7,332 sample), 8,828 respondents in wave 4 (against the original 8,643 sample), and 6,680 respondents in wave 3 (against the original missing sample because it was not in the nurse visit). Multiple practices are matched to individuals with incomplete postcode or address: 462 and 472 in waves 2 and 4, respectively, and 367 in wave 3.

There are 4,124 ELSA respondents with CVD in the second wave, 3,071 in the third wave and 3,028 in the fourth wave. Using the matched ELSA-practice sample, we have 2,766 and 2,584 matched ELSA respondents with CVD in wave 2 and 3, respectively, and 3,131 in wave 4. Of these samples about 146 were matched to multiple practices in wave 2, 145 in wave 3 and 164 in wave 4.

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.

Bennett D. Does clean water make you dirty? Journal of Human Resources, 2012: 47 (1): 146-173.

Bhattacharya J., Bundorf M.K., Pace N., Sood N. Does health insurance make you fat? in NBER "Economic aspects of obesity", Eds M. Grossman and N. H. Mocan, 2011: 35-64.

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.

Ehrlich I., Becker G.S. Market insurance, self-insurance, and self-protection. The Journal of Political Economy, 1972; 80 (4): 623-648

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.

Guthrie B., McLean G. and Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. British Journal of General Practice, 2006; 56(532): 836-841.

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.

Kaestner R., Darden M., Lakdawalla D. Are investments in disease prevention complements? The case of statins and health behaviors. Journal of Health Economics, 2014: 36: 151-163.

Lakdawalla D., Sood N. and Goldman D. HIV Breakthroughs and risky sexual behavior. The Quarterly Journal of Economics, 2006; 121 (3): 1063-1102.

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.

Peltzman S. The effects of automobile safety regulation. The Journal of Political Economy, 1975; 83 (4): 677-726.

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.

Wilson N.L., Xiong W. and Mattson C.L. Is sex like driving? HIV prevention and risk compensation. Journal of Development Economics, 2014; 106: 78-91.