# Health Care Spending in the US vs UK: The Roles of Medical Education Costs, Malpractice Risk and Defensive Medicine

Michael Keane[*], Barry McCormick[†]and Gosia Popławska[‡]

February 8, 2017

## Abstract

We analyze the extent to which three important factors can explain the high costs of health care in the US relative to the UK: (i) the high cost of medical education, (ii) high risks/costs associated with malpractice litigation, and (iii) excessive diagnostic testing arising from the practice of defensive medicine. To do this we formulate, calibrate and simulate a model of the demand and supply of physicians, the labor supply and treatment decisions of practicing physicians, and the market for malpractice insurance. Consistent with prior work, we find that direct costs arising from malpractice risk and malpractice insurance can explain little of the cost differential between the two countries. However, we find that the much higher cost of medical education in the US interacts with malpractice risk in an interesting way: It leads doctors to optimally choose to engage in a high level of diagnostic testing in order to mitigate risk (where by "optimal" we mean from a private rather than social perspective). This defensive behavior leads to an increase in costs far beyond what can be accounted for by malpractice insurance costs alone. Our simulation results imply that more generous medical education subsidies (like those in the UK) would lead to both improved patient welfare and reduced overall health care costs in the US system (a Pareto improvement).

Keywords: Health care expenditure, Physician Supply and Demand, Medical education costs, Malpractice insurance, Defensive medicine, Graduate medical education, Diagnostic testing

JEL codes: I13, D82, C34, C35

---

[*]University of Oxford, Department of Economics
[†]University of Oxford, Centre for Health Service Economics and Organisation
[‡]University of Oxford, Centre for Health Service Economics and Organisation

# 1    Introduction

Health care spending in the United States is higher than in any other country, both in absolute terms and as a percentage of GDP. For example, in 2015 the OECD estimates that the US spent 16.9% of GDP on health care. This is a clear outlier, as spending in other highly developed countries falls within a narrow range of 11.5% for Switzerland (2nd) to 8.9% for Portugal (20th). The United Kingdom is ranked 13th at 9.8%, making it rather typical of developed countries outside the US. Given the relative wealth of the US, its outlier status is magnified if we look at per capita spending. Using PPP exchange rates the OECD estimates per capita spending of \$9451 for the US in 2015, compared to only \$4003 in the UK (17th).

As is well known, the high level of spending in the US does not appear to translate into better outcomes or better quality of care. For example, according to OECD figues, life expectancy at birth in the US in 2015 was estimated at 78.8 years, which places it 27th in the OECD, just below the Czech Republic. The United Kingdom sits 19th in the OECD at 81.4 years. The top 4 countries (Japan, Spain, Switzerland and Italy) have achieved life expectancies exceeding 83 years. Similarly, for 2013, the OECD reports an infant mortality rate of 6.0 per 1000 live births in the US vs. only 3.9 in the UK.

Given this high level of spending, and its failure to be reflected in outcomes, a substantial literature has emerged that attempts to explain health care costs in the US. An obvious potential explanation is that, as the US is neither a purely competitive or socialized system, neither consumers or providers face the full cost of health services, and there is no one entity with the power to regulate/negotiate prices or ration services.[1] As a result, prices are "out of control" and the system is inefficient (Anderson et al. (2003)).

For instance, Berwick and Hackbarth (2012) estimate that health care waste, which includes unnecessary treatments, overpriced drugs, etc., makes up roughly a third of U.S. health care spending.[2] Research has shown that Americans do not utilize more basic medical services such as doctor visits or hospital bed days, but that they pay more for these services. They also receive more "high-tech" services such as advanced imaging, presumably because neither they nor doctors internalize the full cost (and possibly also due to defensive medicine), see Anderson et al. (2007), Laugesen and Glied (2011).

Another common argument is that technological innovations (like advanced imaging) drive up the cost of health care in the US (Smith et al. (2009)). But technology can only explain higher costs relative to other advanced economies (which have access to the same advanced

---

[1]Medicare is the largest single purchaser of health care services in the US, but it has only made sporadic attempts to use its market power to hold down prices. Furthermore, according to Farrell et al. (2008), when medicare reimbursements fall private insurers reimbursements rise to fill the gap. They attribute this to the substantial market power of care providers over private payers.

[2]Cite other papers that refer to high prices of services in the US

technologies) if either (i) consumers do not face the full cost of these innovations and/or (ii) there is no authority that can ration their use. So again it is the hybrid nature of the US system (neither competitive nor centrally planned) that seems to drive high costs. Yet another factor is that the complexity of the US system, with its array of different payers, leads to very high administrative costs, see Woolhandler et al. (2003), Cutler and Ly (2011).

We do not doubt the important roles of excessive prices, new medical technologies and high administrative costs as cost drivers. But in this paper we focus on another set of explanations for the high cost of the US health care system. These are (i) the high cost of US medical education,(ii) high risks/costs associated with malpractice litigation, and (iii) excessive diagnostic testing arising from the practice of defensive medicine. [**Note: Say something here about the size of subsides in the US vs. UK.**] To our knowledge, prior work has not attempted to assess the extent to which high tuition costs in the US (or, conversely, large tuition subsidies in the UK) can explain the differing costs of the health care systems in the two countries.

Just as the cost of becoming a doctor is much higher in the US than the UK, the cost of practicing medicine is also higher due to high malpractice insurance costs. Arguably, this is largely due to the high level of litigiousness in the US, encouraged by both the legal environment and the structure of the medical system (i.e., doctors act as private agents). In the UK, in contrast, the National Health Service bears the cost of malpractice insurance and acts as both defender/investigator of doctors in the event of malpractice claims. Thus, most of the risk of litigation is spread across the health care system as a whole rather than being borne by individual physicians.

Conservative commentators have argued that high malpractice expenses are a key driver of costs in the US system. However, academic research has often failed to support this argument. For example, Mello et al. (2010) conclude that "Overall annual medical liability system costs, including defensive medicine, are estimated to be $55.6 billion in 2008 dollars, or 2.4% of total health care spending", and "This is less than some imaginative estimates put forward in the health reform debate, and it represents a small fraction of total health care spending." Obviously, such figures can do little to explain cost differences between the US and other OECD countries like the UK.

We do not challenge this basic calculation regarding direct costs of the medical litigation system. However, we will use a simple model to show that the high cost of medical education, malpractice risk and diagnostic spending interact in an interesting way to increase medical costs quite dramatically - and far more than any one of these factors viewed in isolation.

Specifically, we formulate, calibrate and simulate a model of the demand and supply of physicians, the labor supply and treatment decisions of practicing physicians, and the market for malpractice insurance. We use the model to analyze the extent to which three

important factors can explain the high costs of health care in the US relative to the UK: (i) the high cost of medical education, (ii) high risks/costs associated with malpractice litigation, and (iii) excessive diagnostic testing arising from the practice of defensive medicine. Consistent with prior work, we find that *direct* costs arising from malpractice risk and malpractice insurance can explain only a modest fraction of the cost differential between the two countries. However, we find that the much higher cost of medical education in the US interacts with malpractice risk in an interesting way: It increases the reservation wage form becoming a doctor substantially, and it causes practicing doctors to optimally choose to engage in a high level of diagnostic testing in order to mitigate risk (where by "optimal" we mean from a private rather than social perspective). This combination of high wages and defensive behavior leads to an increase in costs far beyond what can be accounted for by malpractice insurance costs alone. Our simulation results imply that more generous medical education subsidies (like those in the UK), combined with reductions in malpractice risks, would lead to both improved patient welfare and reduced overall health care costs in the US system (a Pareto improvement).

Our model is calibrated to a number of key statistics about the US economy and health care system, of which several are particularly notable. *Diagnostic Testing*: It is very difficult to determine the amount of health care spending in the US that goes to various services, because the national accounts record spending based on source of payment rather than service provided. However, based on several sources we estimate that 11.6% of health expenditure goes to diagnostic testing, which is about 2% of GDP. In contrast, the UK figure is only 0.5%. *Medical Training Costs*: In the US it takes at least seven years, including medical school and residency to become a doctor. The average debt of a medical graduate is $180k (AAMC 2014). We assume this leads to $14k per year in interest payments for a practicing physician. Of course, another major cost of becoming a doctor is forgone earnings in the alternative occupation during training. *Malpractice Risk and Insurance*: By combining data from the National Practitioner Data Bank and the Physician Workforce report, we estimate that the annual probability of a practicing physician losing his/her license due to malpractice is 0.2%. We calibrate the probability of a malpractice fine at 1.4%, which is consistent with the estimates from NEJM 2011 and the NPDB report on payments in 2012. Finally, we estimate an average malpractice insurance premium of $10k per year. *Aggregate Spending*: We estimate that the component of medical costs due to physician salaries and diagnostic testing is 4.08% of GDP (which compares to only XXX% in the UK).

In our counterfactual experiments we analyze the consequences of subsidizing medical education and lowering malpractice risk. In one experiment we eliminate medical training debt entirely (by letting it be born by the government, as in the UK) and lower malpractice risk to the level of the least litigious US state (Minnesota, where the risk of losing a license is only 0.05% per year). Our model implies that this would reduce the share of doctor wages and diagnostic costs from 4.08% to 2.66% of GDP, a saving of 1.42% of GDP. This corresponds to a saving of $802 per person per year. This cost savings accounts for the taxes needed to finance the tuition subsidies. Our model also implies that, in the new

equilibrium, welfare of doctors and health of patients is unchanged, while the utility of patients is increased. Thus, the experiment is Pareto improving.

We obtain this outcome for two reasons: First, physician wages must be high under the baseline to compensate for high training costs. The policy change, which makes it less expensive become a doctor, leads to lower physician wages in equilibrium. This lowers the gap between the physician wage and the alternative wage. Interestingly, this change in relative wages reduces the practice of defensive medicine (i.e., the level of diagnostic testing), even holding malpractice risk constant, simply because the economic loss that would result from losing ones license is less severe. Second, if we also reduce malpractice risk, the level of defensive medicine is reduced further (via the direct reduction in risk) and the physician wage is reduced further (as the compensating differential for risk is reduced). So we obtain a "virtuous cycle" whereby the cost reductions from lower tuition and lower malpractice risk reinforce each other. Finally, note that the Pareto improvement arises because the level of diagnostic testing is inefficiently high under the baseline, because high tuition costs and malpractice risk distort physician behavior.

The outline of the paper is as follows. In Sections 2 - 4 we present the model. We first present a basic model in which malpractice risk is exogenous. Then in Section 3 we present a model in which malpractice risk depends on the choice of the level of diagnostic testing. And in Section 4 we endogenize the labor supply of doctors. Section 5 describes the calibration of the model, including a description of our data sources. Section 6 presents our counterfactual experiments, and Section 7 concludes.

# 2    Model Part 1: Medical Education Decisions

For the sake of clarity we present our equilibrium model in three parts. This Section discusses the supply of physicians, taking malpractice risk, the price of malpractice insurance and physician behavior as given. Consider an individual who has graduated from university and is deciding whether to pursue a career as a doctor or to pursue a generic alternative. Work in the alternative may commence immediately, but to become a doctor requires seven years of medical school and residency. Thus, there is both a direct cost (tuition) and an opportunity cost (seven years of low wages) to becoming a physician. Furthermore, we assume that the wage stream in the alternative occupation is risk free, while a practicing physician faces two types of risk: having to pay a malpractice fine, or losing their license and having to revert to the alternative occupation.[3] Doctors can buy insurance against malpractice fines, but they cannot insure against possible job loss due to malpractice.

--------

[3]In our model, malpractice is not the result of incompetence or negligence. Rather it is a risk that all doctors face due to the risk of adverse treatment outcomes (and adverse court case outcomes) that may arise stochastically. In the next section we will allow doctors to control this risk through their behavior.

To proceed, the agent is characterized by a time-separable utility function:

$$V(x_1, x_2, \ldots) = \sum_{t=1}^{\infty} \beta^{t-1} u(x_t)$$

where $\beta$ denotes the discount factor and $x_t$ is the consumption level in period $t$. We assume $u' > 0$, $u'' < 0$, and we ignore savings (i.e., the agent consumes his/her current income in each period). To simplify the analysis we assume that career horizon (both in the medical and alternative occupations) is infinite.

At the beginning of the first period, the agent chooses between becoming a doctor and following an alternative career. To become a doctor, he/she has to pay a tuition fee for medical education at the beginning of the first period. We also assume (realistically) that medical training is subsidized. Let us denote the total cost of medical education by $C$. Medical students need to pay only $\gamma C$ as tuition fee, while $(1 - \gamma)C$ is the subsidy from the state, with $0 \leq \gamma \leq 1$. In our analysis we will treat the true cost of medical training $C$ as a technological given, but we treat the share of this cost that is borne by medical students $\gamma$ as a policy variable that can be chosen by government.

We assume that in the first period the agent does not have any accumulated wealth to pay tuition. Therefore he/she must finance it using a bank loan. The loan is paid back in instalments $\gamma f$ for the rest of the infinite life, and the size of instalments is such that:

$$\frac{\gamma f}{1 - \beta} \geq \gamma C$$

so the present value of the stream of instalments is at least as great as the value of the loan.

We assume that wages in both occupations are constant over time. If the agent decides to invest in medical education, he/she will become a doctor and receive wage $w$ for the rest of his/her career *unless* at some point the doctor is found guilty of malpractice. Malpractice happens with probability $\rho$ per period and results in the doctor paying the fine $M$ in the current period and receiving the lower wage $b$ instead of wage $w$ from the next period on. The wage $b$ is not to be confused with the alternative occupation wage. Indeed, it may be higher or lower than that[4], and we treat $b$ as an unknown in our calibration. Finally, after committing malpractice, the doctor has to continue paying back the tuition loan.

While doctors cannot insure themselves against the risk of losing their career due to malpractice, they can buy insurance against the risk of paying a malpractice fine. The insurance premium is a function of the risk of malpractice and the size of malpractice fine, and it includes a proportional loading $\alpha$ over the actuarially fair price:

$$p = (1 + \alpha)\rho M$$

---

[4]We assume $b - \gamma f < w_o$, meaning that consumption after committing malpractice is lower than in the alternative occupation that the agent could have chosen if he/she had not pursued a medical career.

Thus, the level of consumption in period $t$ of a practicing doctor is given by

$$x_t = w - \gamma f - p = w - \gamma f - (1+\alpha)\rho M$$

where $w$, is the wage, $\gamma f$ the loan instalment payment and $p$ is the insurance premium.

In each period the doctor faces an infinite horizon, so the value of a medical career and the value of a career after malpractice are stationary. The value of being a practicing doctor, which we denote by $D$, is:

$$D = u(w - \gamma f - (1+\alpha)\rho M) + \beta \left[\rho U + (1-\rho)D\right]$$

where $U$ denotes the the value of a career after malpractice. This equals:

$$U = \frac{u(b - \gamma f)}{1 - \beta}$$

Hence:

$$D = \frac{1}{1 - \beta(1-\rho)} \left[u(w - \gamma f - (1+\alpha)\rho M) + \frac{\rho\beta}{1-\beta}u(b - \gamma f)\right]$$

Rather than choosing a medical career, the agent can decide to choose the alternative occupation, which pays a constant opportunity wage $w_o$ in each period. As no investment beyond college is needed to choose this career, its lifetime value is simply given by $\frac{u(w_o)}{1-\beta}$.

Thus, an individual will choose to become a doctor if and only if:

$$\frac{1}{1 - \beta(1-\rho)} \left[u(w - \gamma f - (1+\alpha)\rho M) + \frac{\rho\beta}{1-\beta}u(b - \gamma f)\right] \geq \frac{u(w_o)}{1 - \beta} \tag{1}$$

This yields reservation wage for individuals to pursue a medical career:

$$w \geq w^* = \gamma f + (1+\alpha)\rho M + u^{-1}\left\{u(w_o) + \frac{\rho\beta}{1-\beta}\left[u(w_o) - u(b - \gamma f)\right]\right\} \tag{2}$$

It is important to note that the demand for physicians in the US is not competitively determined. Medical residency slots are rationed, and the number of available slots is determined by a complex set of actors, including medical licensing boards, teaching hospital administrators, the federal government (which, through the Medicare program, provides substantial subsidies for residency slots), State governments and universities (which finance the building of teaching hospitals) and so on. Rather than attempting to model this complex set of actors, we simply assume that a social planner determines the desired number of doctors. The doctor wage must then adjust to clear the market at this level.[5] The following proposition summarizes some interesting properties of the physician wage.

---

[5]We assume that individual doctors have no bargaining power. Hence, they are paid the minimum wage that makes them indifferent between the medical and alternative occupations. Of course, the bargaining power of the medical profession is reflected in licensing boards that influence the number of residency slots.

**Proposition 1** *The reservation wage for individuals to pursue a medical career is increasing in the size of the opportunity wage in the alternative occupation, the malpractice fine, the insurance premium loading, the risk of malpractice, and medical tuition fees. Moreover, the greater is the tuition cost, the greater is the effect of malpractice risk on the reservation wage.*

Proof: By differentiation of the reservation wage equation we have:

$$\frac{\partial w^*}{\partial w_o} = \frac{\rho\beta}{1-\beta} \cdot \frac{u'(w_o)}{u'(w^* - \gamma f - (1+\alpha)\rho M)} > 0$$

$$\frac{\partial w^*}{\partial M} = (1+\alpha)\rho > 0$$

$$\frac{\partial w^*}{\partial \alpha} = \rho M > 0$$

$$\frac{\partial w^*}{\partial \rho} = (1+\alpha)M + \frac{1}{u'(w^* - \gamma f - (1+\alpha)\rho M)} \cdot \frac{\beta}{1-\beta}[u(w_o) - u(b - \gamma f)] > 0$$

$$\frac{\partial w^*}{\partial \gamma} = f + \frac{\rho\beta f}{1-\beta} \cdot \frac{u'(b - \gamma f)}{u'(w^* - \gamma f - (1+\alpha)\rho M)} > 0$$

$$\frac{\partial^2 w^*}{\partial\gamma\partial\rho} = \frac{\beta f u'(b-\gamma f)}{1-\beta} \cdot \frac{u'(w^* - \gamma f - (1+\alpha)\rho M) - \rho u''(w^* - \gamma f - (1+\alpha)\rho M)\left[\frac{\partial w^*}{\partial\rho} - (1+\alpha)M\right]}{u'(w^* - \gamma f - (1+\alpha)\rho M)} > 0$$

The cross derivative is crucial to understanding the functioning of our model. It captures the complementarity between tuition costs and malpractice risk in driving up physician wages (and hence overall medical costs) that we emphasized in the Introduction.

It is interesting that this complementarity does not depend on risk aversion. If $u'' = 0$ the cross partial between the share of tuition borne by the student $\gamma$ and malpractice risk $\rho$ simplifies to:

$$\frac{\partial^2 w^*}{\partial\gamma\partial\rho} = \frac{\beta f u'(.)}{1-\beta} > 0$$

Thus, the greater is tuition, the greater is the increase in the doctor wage needed to compensate doctors for any given increase in malpractice risk. Risk aversion magnifies this effect. Finally, it is worth recalling that we treat the true cost of medical training, and hence $f$, as a given, but we treat the share of this cost that is borne by medical students ($\gamma$) as a policy variable that can be chosen by government. We also treat malpractice risk $\rho$ as a parameter that can be influenced by government (by manipulation of aspects of malpractice law such as liability caps). In the next section we also let the malpractice risk be influenced by physician behavior, specifically the level of diagnostic testing.

# 3 Model Part 2: Diagnostic Tests and Defensive Medicine

In this Section we consider in more detail the behavior of practicing physicians. In our model we will assume that doctor's time and diagnostic testing can both lead to improved patient outcomes. However, for the sake of clarity, in this Section we focus on the role of diagnostics (assume doctors provide a fixed time input per patient). In Section 4 we will extend the model to include the intensive margin of labor supply of physicians.

## 3.1 Diagnostic Tests and Defensive Medicine Holding the Physician Wage Fixed

In our model diagnostic tests serve two functions. They improve patient outcomes by improving diagnosis, and they reduce the doctor's risk of malpractice. Specifically, we assume that the per period probability of committing malpractice $\rho$ is a decreasing and convex function of the quantity of diagnostics prescribed to each patient, which we denote by $d$. That is, we have $\rho(d)$, with $\rho'(d) < 0$ and $\rho''(d) > 0$.

For the sake of clarity, in this subsection we consider the behavior of practicing physicians in this environment, holding the physician wage fixed. In the next subsection we will extend the model to bring in equilibrium considerations (i.e., to allow for how physician behavior affects the equilibrium wage).

To proceed, we assume the utility of the representative patient, $K$, is additively separable in consumption and health, $H$. We assume that each patient is matched with one doctor, who decides to prescribe $d$ diagnostic tests. More tests increase health of the patient, but they also lower patient's consumption:

$$K(d) = u(I - p_d d) + H(d)$$

Here $I$ denotes a patient's income, $d$ the number of diagnostics prescribed and $p_d$ the price per diagnostics. Patient's utility of consumption is modeled using the same function as doctor's utility of consumption. Moreover, we assume that patient's health and total utility are twice differentiable functions such that $H' > 0$, $H'' < 0$ and $K'(d) > 0$ for $d < d_P^*$, $K'(d) < 0$ and $K''(d) < 0$ for $d > d_P^*$. The function $K$ obtains a maximum at $d_P^*$, which we call the 'patient's optimum'.

$$d_P^* = \underset{d}{\operatorname{argmax}} \ K(d)$$

As we shall see, the doctor may chose a level of diagnostics higher than the patient's optimum due to the incentives created by malpractice risk.

Next, we return to a doctor's decision problem. Let the doctor derive a non-pecuniary benefit $N(d)$ from the utility of the patient as follows:

$$N(d) = \mu\, K(d)$$

Here $\mu$ is a parameter describing doctor's altruism. If $\mu > 0$ then doctors care about the utility of their patients.

We assume a doctor's per period utility is additively separable in the utility of own consumption $u(w - \gamma f - p)$ and the non-pecuniary reward from treating patients, as in $u(w - \gamma f - p) + N(d)$. A practicing doctor chooses the level of diagnostic procedures $d = d^*$ that maximizes his expected lifetime utility:

$$
\begin{aligned}
D(d) &= u(w - \gamma f - p) + \mu K(d) + \beta\left[\rho(d)U + (1 - \rho(d))D(d)\right] \\
&= \frac{1}{1 - \beta(1 - \rho(d))}\left[u(w - \gamma f - p) + \mu K(d) + \frac{\beta\rho(d)}{1 - \beta}u(b - \gamma f)\right] \quad (3)
\end{aligned}
$$

Here we emphasize the fact that doctors face no direct cost of prescribing additional diagnostics.[6] Costs only arise indirectly because diagnostics are expensive to patients (to the extent they face co-pays), and doctors care about patient welfare. Note also that we assume that doctors are price takers on the malpractice insurance market, so they do not internalize the fact that reducing their malpractice risk might reduce the insurance premium $p$.

The following proposition characterizes the level of diagnostic testing chosen by doctors, and how it depends on medical school tuition and other key factors:

**Proposition 2** *Assume $\rho''(d) > 0$ (diminishing returns to prescribing additional diagnostics), $K''(d) < 0$ for $d > d_P^*$ (patients' costs of excess diagnostics over their optimum are convex), and the following two conditions are satisfied:*

$$
\begin{aligned}
u(b - \gamma f) &< u(w - \gamma f - p) + \mu K(d) \quad \text{for each } d < d_P^* & (4) \\
\mu K'(d_{max}) &< \frac{-\beta(-\rho'(d_{max}))}{1 - \beta(1 - \rho(d_{max}))}\left[u(w - \gamma f - p) + \mu K(d_{max}) - u(b - \gamma f)\right] & (5)
\end{aligned}
$$

*where $d_{max}$ is a technologically determined maximum level of diagnostics.*

*Then there exists a unique level of diagnostics $d^*$ that maximizes the doctor's expected lifetime utility and that satisfies the following first order condition:*

$$
\frac{K'(d^*)}{\rho'(d^*)} = \frac{\beta}{\mu(1 - \beta(1 - \rho(d^*)))}\left[u(w - \gamma f - p) - u(b - \gamma f) + \mu K(d^*)\right] \quad (6)
$$

---

[6]This is a fairly realistic assumption under the US health care system. A possible exception is that doctors who work for managed care organization may come under some pressure to lower costs if they are identified as using excessive diagnostics. But our understanding is that this is rare.

*The optimal level of diagnostics chosen by the doctors will exceed the patient's optimum. Moreover, the optimal level of diagnostics is increasing in tuition fees and in doctors' wages, but decreasing in insurance premia, the post-malpractice wage and doctors' altruism towards patients.*

Short Proof:

Consider first the derivative of $D(d)$:

$$D'(d) = \frac{\beta(-\rho'(d))}{[1 - \beta(1 - \rho(d))]^2}\left[u(w - \gamma f - p) + \mu K(d) - u(b - \gamma f)\right] + \frac{\mu}{1 - \beta(1 - \rho(d))}K'(d)$$

Condition (4) implies that $D'(d) > 0$ for $d \leq d_P^*$. Thus, the doctor will always choose a level of diagnostics higher than the patient's optimum. Moreover, condition (5) implies that $D'(d_{max}) < 0$.[7]

As $D'(d)$ is continuous, $D'(d) > 0$ for $d \leq d_P^*$, and $D'(d_{max}) < 0$, there exists an interior solution $d^*$ to the doctor's optimization problem, which satisfies the first order condition $D'(d^*) = 0$. Moreover, the doctor's optimum is unique: As $\rho$ is decreasing and convex and $K$ is decreasing and concave for $d > d_P^*$, we have that $D(d)$ is concave at all points where $D'(d) = 0$. Hence, there exists only one level of diagnostics $d^*$ which maximizes the utility of the doctor (See Appendix for a detailed proof).

Finally, we present a comparative statics analysis. Denote $\bar{w} = w - \gamma f - p$ and $\bar{b} = b - \gamma f$, and $z(d) = \frac{K'(d)}{\rho'(d)}$. Note that:

$$z'(d^*) = \frac{-K''(d^*) - K'(d^*)\frac{\rho''(d^*)}{-\rho'(d^*)}}{-\rho'(d^*)} > 0$$

By total differentiation, we obtain:

$$
\begin{aligned}
\frac{dd^*}{d\gamma} &= A_1\, f\, \left[u'(\bar{b}) - u'(\bar{w})\right] > 0 \\
\frac{dd^*}{dw} &= A_1\, u'(\bar{w}) > 0 \\
\frac{dd^*}{dp} &= A_1\, \left[-u'(\bar{w})\right] < 0 \\
\frac{dd^*}{db} &= A_1\, \left[-u'(\bar{b})\right] < 0 \\
\frac{dd^*}{d\mu} &= A_1\, \frac{-\left[u(\bar{w}) - u(\bar{b})\right]}{\mu} < 0
\end{aligned}
$$

---

[7]A sufficient condition for (5) to hold is that $K(d_{max})$ is low enough, in which case $u(w - \gamma f - p) + \mu K(d_{max}) - u(b - \gamma f) < 0$. Alternatively, $\rho'(d_{max})$ must be sufficiently close to 0.

where $A_1 = \frac{\beta}{z'(d^*)\mu\,[1-\beta(1-\rho(d^*))]}$, $A_1 > 0$.

Thus, as claimed, the level of diagnostics chosen by the doctors is increasing in tuition fees $\gamma$, increasing in wages $w$, but decreasing in the size of insurance premia $p$, the wage rate obtained after malpractice $b$ and the degree of altruism $\mu$. QED

Comment: By prescribing diagnostics in excess of the patient's optimum, doctors reduce their malpractice risk. But this also increases costs to patients. These costs are only partly internalized by doctors (to the degree they are altruistic), causing the level of diagnostics to be inefficiently high from a social perspective.

## 3.2  Equilibrium: Wages under Defensive Medicine

In Section 3.1 we analyzed the behavior of practicing physicians, taking the physician wage as given. In this subsection we refine the analysis to account for the fact that the behavior of practicing physicians (i.e., their choice of the level of diagnostics) may affect the physician wage. First, we revise the decision rule (1) for choosing to become a doctor to accommodate the new utility function presented in Section 3.1 (where doctors choose diagnostics and are altruistic with respect to patient welfare).

The modified decision rule for becoming a doctor is given by $D(d^*) \geq \frac{u(w_o)}{1-\beta}$, which will hold with equality in the equilibrium if doctors have no bargaining power, giving:

$$u(w^* - \gamma f - p) - u(b - \gamma f) + \mu K(d^*) = \frac{1 - \beta(1 - \rho(d^*))}{1 - \beta}\,[u(w_o) - u(b - \gamma f)] \qquad (7)$$

where $w^*$ denotes the equilibrium wage.

The level of wages and the number of diagnostics prescribed in equilibrium is jointly determined by condition (7) along with (6) and the equilibrium condition $p = (1+\alpha)\rho M$:

$$\begin{cases} u(w^* - \gamma f - p) - u(b - \gamma f) + \mu K(d^*) = & \frac{1-\beta(1-\rho(d^*))}{1-\beta} \cdot [u(w_o) - u(b - \gamma f)] \\ u(w^* - \gamma f - p) - u(b - \gamma f) + \mu K(d^*) = & \frac{\mu[1-\beta(1-\rho(d^*))]}{\beta} \cdot \frac{K'(d^*)}{\rho'(d^*)} \\ p = & (1+\alpha)\rho(d^*)M \end{cases}$$

The following proposition summarizes how the equilibrium level of diagnostics is influenced by exogenous factors in the model:

**Proposition 3** *Assume (as above) there are diminishing returns to prescribing additional diagnostics and the patients' costs of diagnostics are increasing and convex. Then, the equilibrium level of diagnostics is increasing in tuition fees and opportunity wages, and decreasing in the post-malpractice wage and doctors' altruism towards patients. Moreover, the equilibrium level of diagnostics does not depend on the size of malpractice fines or the insurance loading.*

11

Short proof:

Note that the left hand sides of the first two equilibrium equations are equal. Hence, we can write the optimal level of diagnostics as a function of $w_o, b - \gamma f$ and $\mu$ as follows:

$$\frac{K'(d^*)}{\rho'(d^*)} = \frac{\beta}{\mu(1-\beta)} [u(w_o) - u(b - \gamma f)] \qquad (8)$$

Denote $z(d) = \frac{K'(d)}{\rho'(d)}$. From our assumptions we have $z'(d) > 0$ for $d > d_P^*$. Moreover, $z(d_P^*) = 0$. To show that the equilibrium exists, the following condition must be satisfied:

$$z(d_{max}) > \frac{\beta}{\mu(1-\beta)} [u(w_o) - u(b - \gamma f)] \qquad (9)$$

Under this condition, equation (8) gives a solution for the level of diagnostics prescribed in equilibrium. Total differentiation of (8) yields:

$$\frac{dd^*}{d\gamma} = A_2 \, f \, u'(b - \gamma f) > 0$$

$$\frac{dd^*}{dw_o} = A_2 \, u'(w_o) > 0$$

$$\frac{dd^*}{db} = A_2 \, [-u'(b - \gamma f)] < 0$$

$$\frac{dd^*}{d\mu} = A_2 \, \frac{-[u(w_o) - u(b - \gamma f)]}{\mu} < 0$$

where $A_2 = \frac{\beta}{z'(d^*)\mu(1-\beta)}$, $A_2 > 0$.

Thus, in equilibrium, the level of diagnostics is increasing in the tuition fee and the opportunity wage, and decreasing in the post-malpractice wage and doctors' degree of altruism for patients. Furthermore, we also have:

$$\frac{dd^*}{dM} = 0$$

$$\frac{dd^*}{d\alpha} = 0$$

so the optimal level of diagnostics is not affected by the size of malpractice fine $M$ or the insurance loading $\alpha$. *QED*

Comment: The reason that diagnostics are not affected by the malpractice fine or the insurance loading is that any change in these variables is fully compensated by changes in the equilibrium wage.

The following proposition summarizes how the equilibrium level of wages is influenced by exogenous factors in the model:

**Proposition 4** *Assume (as above) there are diminishing returns to prescribing additional diagnostics and the patients' costs of diagnostics are increasing and convex. Then the equilibrium physician wage is increasing in the level of malpractice fines and the insurance loading. Moreover, if malpractice fines are not too high (in a sense made precise below), then the physician wage is increasing in tuition fees and the opportunity wage, and decreasing in the post-malpractice wage.*

Short proof:

The equilibrium wage is given by:

$$u(w^* - \gamma f - (1+\alpha)\rho(d^*)M) = u(w_o) + \frac{\beta\rho(d^*)}{1-\beta}\left[u(w_o) - u(b - \gamma f)\right] - \mu K(d^*) \quad (10)$$

By total differentiation of (10) we have that:

$$\frac{dw^*}{dM} = (1+\alpha)\rho(d^*) > 0$$
$$\frac{dw^*}{d\alpha} = \rho(d^*)M > 0$$

Thus, higher levels of malpractice fine $M$ or the insurance loading $\alpha$ translate into higher equilibrium wages. Total differentiation of (10), also allows us to obtain the derivative of the equilibrium wage with respect to the tuition fee:

$$u'(\bar{w})(\frac{\partial w^*}{\partial \gamma} - f - (1+\alpha)M\rho'(d^*)\frac{\partial d^*}{\partial \gamma}) = \frac{\partial d^*}{\partial \gamma}\left\{-\mu K'(d^*) + \frac{\beta}{1-\beta}\rho'(d^*)\left[u(w_o) - u(\bar{b})\right]\right\} +$$
$$+ \frac{\beta}{1-\beta}f\rho(d^*)u'(\bar{b})$$

where we denote $\bar{w} = w^* - \gamma f - (1+\alpha)\rho M$ and $\bar{b} = b - \gamma f$. Note that because doctors optimize the number of diagnostics, condition (8) implies the term in curly brackets is zero. So we have:

$$\frac{dw^*}{d\gamma} = f\left[1 + \frac{\beta}{1-\beta}\frac{u'(\bar{b})}{u'(\bar{w})}\rho(d^*)\right] - (1+\alpha)M\left[-\rho'(d^*)\right]\frac{dd^*}{d\gamma} \quad (11)$$

By using the expression for $\frac{\partial d^*}{\partial \gamma}$ from the proof of proposition 3, we obtain:

$$\frac{dw^*}{d\gamma} = fu'(\bar{b})\frac{\beta}{1-\beta}\left[\frac{1-\beta}{\beta}\frac{1}{u'(\bar{b})} + \frac{\rho(d^*)}{u'(\bar{w})} - \frac{1+\alpha}{\mu}\frac{-\rho'(d^*)}{z'(d^*)}M\right]$$

Thus, the physician wage in equilibrium is increasing in the tuition fee iff:

$$\frac{dw^*}{d\gamma} > 0 \iff M < \bar{M}_\gamma = \frac{z'(d^*)}{-\rho'(d^*)}\frac{\mu}{(1+\alpha)}\frac{1}{u'(\bar{w})}\left[\frac{1-\beta}{\beta}\frac{\mu u'(\bar{w})}{u'(\bar{b})} + \rho(d^*)\right] \quad (12)$$

13

The right hand side of (12) is a threshold for level of the malpractice fine, below which the effect of tuition fees on the physician wage is positive.[8]

Finally, we consider how the equilibrium wage depends on the opportunity wage $w_o$, the post-malpractice wage $b$ and the degree of altruism $\mu$:

$$\frac{dw^*}{dw_o} = u'(w_o)\frac{\beta}{1-\beta}\left[\frac{1-\beta(1-\rho(d^*))}{\beta u'(\bar{w})} - \frac{1+\alpha}{\mu}\frac{-\rho'(d^*)}{z'(d^*)}M\right]$$

$$\frac{dw^*}{dw_o} > 0 \iff M < \bar{M}_{w_o} = \frac{z'(d^*)}{-\rho'(d^*)}\frac{\mu}{(1+\alpha)}\frac{1}{u'(\bar{w})}\left[\frac{1-\beta}{\beta} + \rho(d^*)\right]$$

$$\frac{dw^*}{db} = -u'(b-\gamma f)\frac{\beta}{1-\beta}\left[\frac{\rho(d^*)}{u'(\bar{w})} - \frac{1+\alpha}{\mu}\frac{-\rho'(d^*)}{z'(d^*)}M\right]$$

$$\frac{dw^*}{db} < 0 \iff M < \bar{M}_b = \frac{z'(d^*)}{-\rho'(d^*)}\frac{\mu}{(1+\alpha)}\frac{1}{u'(\bar{w})}\rho(d^*)$$

$$\frac{dw^*}{d\mu} = \frac{-K(d^*)}{u'(\bar{w})} + \frac{1+\alpha}{\mu}\frac{-K'(d^*)}{z'(d^*)}M$$

$$\frac{dw^*}{d\mu} < 0 \iff M < \bar{M}_\mu = \frac{z'(d^*)}{-K'(d^*)}\frac{\mu}{(1+\alpha)}\frac{K(d^*)}{u'(\bar{w})} \iff$$

$$M < \bar{M}_\mu = \frac{f'(d^*)}{-\rho'(d^*)}\frac{\mu}{(1+\alpha)}\frac{1}{u'(\bar{w})}\left[\rho(d^*) - \frac{1-\beta}{\beta}\frac{u(\bar{w})-u(w_o)}{u(w_o)-u(\bar{b})}\right]$$

These conditions give thresholds on the values of malpractice fines $M$, under which wages in the equilibrium will be increasing in the opportunity wage $w_o$, decreasing in the post-malpractice wage, and decreasing in the degree of altruism $\mu$. *QED*

Comment 1: If malpractice fines are higher than the threshold level $\bar{M}_\gamma$, then the effect of higher tuition on the physician wage will be *negative*. The reason is that, due to higher tuition, doctors will behave as if they are more averse to losing a medical career. They will prescribe additional diagnostics, which will reduce the observed probability of malpractice. As a result, in equilibrium, doctors will benefit from lower malpractice insurance premiums. If malpractice fines are high enough, the benefit from lower premiums will outweigh the direct cost of higher tuition. As a result, a lower physician wage is needed in equilibrium to induce individuals to choose a medical career. While a theoretical possibility, we would regard this as a fairly unrealistic case.

---

[8]Note that the right hand side of (12) does not depend on $M$. First, the optimal level of diagnostics does not depend on the size of malpractice fines. Moreover, the value of $\bar{w} = w^* - \gamma f - (1+\alpha)\rho(d^*)M$ is given by the equilibrium condition (18), hence it also does not depend on $M$ (because wages will always adjust to compensate for higher $M$). Therefore, it is legitimate to say that the right hand side of (12) forms a threshold for the values of malpractice fines, under which the effect of tuition fees on doctor's wages in the equilibrium is positive.

Comment 2: For sufficiently high values of $M$, the physician wage in equilibrium will be decreasing in the opportunity wage $w_o$, or increasing in the post-malpractice wage $b$ or the level of altruism $\mu$. These surprising results are due to the same equilibrium mechanism whereby more diagnostic testing reduces the malpractice probability and the insurance premium.

Comment 3: Note that

$$\bar{M}_\mu < \bar{M}_b < \bar{M}_\gamma < \bar{M}_{w_o}$$

Hence, if $M < \bar{M}_\mu$, the equilibrium wage is increasing in tuition and the opportunity wage but decreasing in the post-malpractice wage and the doctor's altruism. However, if $M > \bar{M}_{w_o}$ then all these effects are reversed. In between, there exists a range of values of malpractice fines for which the signs of the effects of altruism, the post-malpractice wage, tuition and the opportunity wage are reversed, in that order.

Comment 4: The direct positive effect of tuition on the physician wage is captured by the first term in (11). The indirect equilibrium effect that arises because higher tuition leads to more diagnostic testing, which in turn lowers malpractice risk and malpractice premiums, is captured by the second term in (11). This equilibrium effect will be stronger, the higher is the malpractice fine. This is the only channel through which increased diagnostics (due to higher tuition) affects wages in the equilibrium. This effect is present, even though doctors optimize number of diagnostics, because they do not internalize the fact that the level of diagnostics they choose will affect the insurance premiums on the market.

To summarize, our model shows that the level of diagnostics that physicians prescribe in equilibrium is increasing in the tuition they have to pay at the beginning of their careers. Moreover, under the condition that malpractice fines are not too large, the wages that are offered to doctors in equilibrium will also be increasing in the tuition fees. Hence, subsidizing physician fees will not only lower doctor's wages, but also have the benefit of reducing the inefficiently high level of diagnostics that doctors prescribe, bringing it closer to the social optimum $d_P^*$.

## 3.3   Graphical Description of the Equilibrium

This section provides a graphical depiction of equilibrium, holding the malpractice insurance premium $p$ fixed to simplify the analysis. This will facilitate intuition about how tuition fees affect physician wages and the level of defensive medicine.

First, taking the level of diagnostics $d$ as a given, the wage is set as $w(d)$ to satisfy the following equation:

$$u(w(d) - \gamma f - p) = u(w_o) + \frac{\beta \rho(d^*)}{1 - \beta} \cdot [u(w_o) - u(b - \gamma f)] - \mu K(d^*)$$

15

The slope of the function $w(d)$ is given by:

$$\frac{dw}{dd} = \frac{-\rho'(d)}{u'(w - \gamma f - p)}\left[\mu \frac{K'(d)}{\rho'(d)} - \frac{\beta}{1-\beta}\left[u(w_o) - u(b - \gamma f)\right]\right]$$

Because $u(x)$ is increasing, it is trivial to show that for each $d$ there exists only one $w$ satisfying this condition. Note that the expression in the bracket takes a value 0 when $d = d^*$ (see (16)), is negative for $d < d^*$, and positive for $d > d^*$. Hence, the function $w(d)$ is decreasing for $d < d^*$, reaches a minimum at $d = d^*$ and is afterwards increasing.

Next, taking wages as given, doctors will choose $d$, to satisfy the following condition:

$$u(w - \gamma f - p) = u(b - \gamma f) - \mu K(d) + \frac{\mu[1 - \beta(1 - \rho(d))]}{\beta} \cdot \frac{K'(d)}{\rho'(d)}$$

The right hand side is increasing in $d$, so we can write $d$ as a function of $w$. The slope of $d(w)$ is given by:

$$\frac{dd}{dw} = \frac{u'(w - \gamma f - p)}{\mu \frac{1 - \beta(1 - \rho(d))}{\beta} z'(d)} > 0$$

Hence, $d$ is an increasing function of $w$.

Now, consider the effect of increased tuition fees on $w(d)$ and $d(w)$:

$$\frac{dw}{d\gamma f} = 1 + \frac{\beta}{1-\beta}\rho(d)\frac{u'(b - \gamma f)}{u'(w - \gamma f - p)} > 1$$

$$\frac{dd}{d\gamma f} = \frac{\beta}{z'(d)\mu\left[1 - \beta(1 - \rho(d))\right]}(u'(\bar{b}) - u'(\bar{w})) \geq 0$$

Hence, if tuition fees go up, then for each $d$, the curve $w(d)$ is shifted upwards, by a value that is higher than the change in $\gamma f$. Moreover, if tuition fees go up, the curve $d(w)$ is shifted upwards at every level of $w$. In the new equilibrium doctors earn higher wages and prescribe more diagnostics. This is illustrated in Figure 1.

Figure 2 illustrates the special case where doctors are risk neutral. Then changes in tuition fees have no effect on the $d(w)$ curve, as $u'(\bar{b}) - u'(\bar{w}) = 0$. However, due to the shift in the $w(d)$ curve, in the new equilibrium doctors will earn higher wages and also prescribe more diagnostics. Thus, we see that the effect of tuition fees on diagnostics does not derive fundamentally from risk aversion. Rather it derives from the fact that higher tuition raises the gap between the physician and non-physician wages, creating a situation where doctors have more to lose from a malpractice charge. Risk aversion only magnifies this effect.
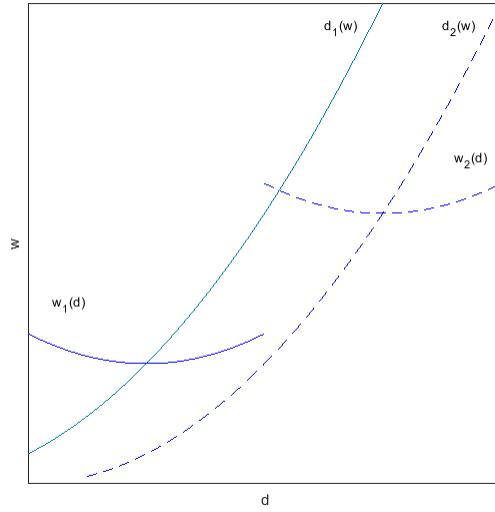
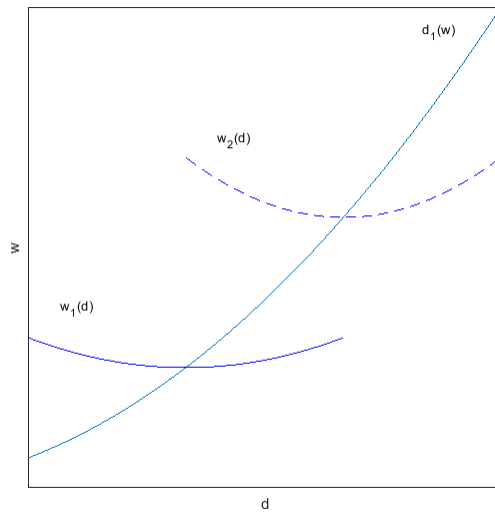Figure 1: The effect of increased tuition fee on diagnostics and wages if doctors are risk averse



Figure 2: The effect of increased tuition fee on diagnostics and wages if doctors are risk neutral

# 4  Model with Physician Time and Training Costs

There is clear evidence that physicians in the US not only use more diagnostics but also spend less time with patients than physicians in the UK. In order to be able to address this phenomenon, in this section we extend the model to allow the time that doctors spend with patients to enter the health production function (in addition to diagnostics). Clearly, the wage rate of doctors may be expected to alter the way they practice medicine; a higher wage may induce them to substitute diagnostic tests for consultation time in the diagnosis/treatment process. An implication is that the high level of diagnostic testing in the US may not be due primarily to defensive medicine. It may instead be an optimal response to high wage rates induced by high tuition costs.

## 4.1  Specifications of utility and health production functions

### 4.1.1  The Patient's Utility Function when Health Depends on Physician Time

To proceed, we assume a patient's utility depends on his/her own consumption and leisure, as well as health $H$, which is produced by diagnostics and consultation time with doctors:

$$K(d, \mathbf{nt}) = U^P(c_P, t_P) + H(d, \mathbf{nt})$$

Here $c_P$ is the patient's (non-medical) consumption and $t_P$ is the patient's leisure. We assume for simplicity that the health component of utility is additively separable from the consumption/leisure component. As before, $d$ denotes diagnostics. But now we also let health depend on the level of consultation time $t$ chosen by doctors. Let $n$ denote the number of (identical) doctors per patient, a variable that, as we noted earlier, is chosen by a social planner. Given there are $n$ doctors per patient, the total amount of consultation time per patient is given by $nt$.[9]

We choose this formulation with $n$ (identical) doctors per patient (rather than having one representative doctor) because below we will be interested in scenarios where the social planner alters the number of physicians $n$.[10] For instance, consider an experiment where we alter tuition or malpractice risk. Unlike in Sections 2-3, where physician time was fixed,

---

[9]This does not imply that each doctor spends exactly $t$ units of time with each patient. If we define $k = 1/n$ as the number of patients per doctor, then each doctor spends an average of $t/k$ units of time with each patient, but a doctor may spend more time with some patients than others. In fact, we will need to assume that each patient has one primary doctor (see footnote 10) to rationalize how we model diagnostics.

[10]We assume the level of diagnostics $d$ does not depend on the number of doctors. We justify this by assuming that, regardless of $n$, each representative patient is assigned one representative doctor who determines his/her level of diagnostics. Thus, while the time patients can spend consulting with doctors is constrained by $n$ (in the sense that doctors' labor supply is not perfectly elastic), the level of $d$ is not constrained by $n$ because the supply of diagnostic testing is perfectly elastic, consistent with our fixed price $p_d$ assumption.

such policy changes may now alter physician labor supply $t$. The social planner may want to compensate by adjusting $n$ so as to hold total physician consultative services $nt$ fixed.[11] For this purpose, it will also be useful to define $S$ as the marginal cost of training new doctors. The yearly cost of subsidies per doctor will equal the yearly cost of 'unpaid debt':

$$S = (1 - \gamma)f$$

The budget constraint of the representative consumer/patient is given by:

$$c_P = I - p_d d - n(zwt + \psi S)$$

where $I$ is the patient's income, $p_d d$ is the cost of prescribed diagnostics, $nwt$ is the cost of doctor consultation time, and $nS$ is the cost of doctor training subsidies. The parameter $z > 1$ captures the gross social cost of a unit of physician time (which includes not just the wage but also fringe benefits), while the parameter $\psi > 1$ is the marginal social cost of funds (which captures inefficiencies generated by allocating funds to training subsidies). We assume patient's utility $U$ is linearly separable in the utility of consumption $u^P$ and utility of leisure $v$:

$$U^P(c_P, t_P) = u^P(c_P) + v(t_P)$$

Finally, we will assume that doctor's actions do not significantly influence patient's leisure. Therefore we can drop patients' leisure from the analysis, setting $v(t_P) = 0$.


### 4.1.2 The Doctor's Utility Function and Budget Constraint extended to include consultation time

First, consider the budget constraint. We assume that the doctor's working time consists entirely of the consultation time with patients $t$, and that the doctor is paid (net) wage $w$ per unit of time (with $t = 1$ denoting full-time work). This gives the following budget condition for the doctor:

$$\bar{w} = wt - \gamma f - p$$

Next, we assume a doctor's utility is additively separable between a component that depends on his/her own consumption and leisure, and and a non-pecuniary benefit from treating patients, as follows:

$$U = U(c, t) + \mu K(d, \mathbf{nt})$$

Here $c$ denotes doctor's consumption and $t$ time spent on consulting patients. Obviously, the doctor's leisure is decreasing in t. The term $K(d, \mathbf{nt})$ is the utility of a representative patient, and $\mu$ is the altruism parameter.

---

[11]More generally, the present formulation allows us to examine how changing $n$ affects physician consultation time $t$, as well as wages and diagnostics, in the new equilibrium.

We will further assume the utility of consumption and leisure are additively separable:

$$U(c,t) = u(c) + v(t)$$

with $u'(c) > 0, u''(c) < 0$ and $v(0) = 0, v'(t) < 0, v''(t) < 0$.

Finally, we assume that time spent with patients $t$ does not have an effect on the malpractice risk $\rho$. This is a justifiable assumption, given that, in the event of a malpractice claim, the time spent with the patient cannot be used as an argument in court (**Can we give a citation for this?**). Diagnostics are easier to record and give an easily verifiable measure of the doctor's effort. Therefore, they have a clear effect on the risk of being charged with malpractice.

### 4.1.3 A Set of Simplifying Assumptions

The addition of physician labor supply greatly complicates analytic solution of the model. Thus, we make a number of assumptions to simplify the analysis.

First, it is convenient to make the following approximation: We assume that, when evaluating the patient's utility of consumption, the doctor linearly approximates $u(c)$ around patient's income $I$, as in $u^P(I, x) = u(I) - u'(I)x$, so that:

$$u^P(I - p_d d - nzwt - n\psi S) = u(I) - u'(I)p_d d - u'(I)nzwt - u'(I)n\psi S$$

This linear approximation allows us to avoid cross derivatives of $t, w$ and $d$, which makes the model analytically much more tractable.[12]

Second, we assume the health production function is additively separable, so that time spent with patients can be perfectly substituted by prescribing diagnostic tests [13]:

$$H(d, nt) = h(d) + s(nt)$$

with $h' > 0, h'' < 0, s' > 0, s'' > 0$.

Given these assumptions, the utility of the patient, as evaluated by the doctor, is given by:

$$K(d,t) = u(I) - u'(I)p_d d - u'(I)nzwt - u'(I)n\psi(1-\gamma)f + h(d) + s(nt)$$

---

[12] We justify using this local linear approximation by the fact that the share of spending on diagnostics and doctors' wages is a small fraction of patient's total consumption (on average around 4%).

[13] This condition allows us to separate the FOC for $d$ and $t$

### 4.1.4 The Doctor's Optimization Problem

Under these assumptions, the doctor's problem is to choose the level of diagnostics $d^*$ and the amount of time spent consulting with patients $t^*$ so as to maximize expected lifetime utility $D(d,t)$:

$$
\begin{aligned}
D(d,t) &= \frac{U(wt - \gamma f - p, t) + \mu K(d,t) - U(bt_b^* - \gamma f, t_b^*)}{1 - \beta(1 - \rho(d))} + \frac{U(bt_b^* - \gamma f, t_b^*)}{1 - \beta} \\
&= \frac{1}{1 - \beta(1 - \rho(d))}\left\{ u(wt - \gamma f - p) + v(t) - u(bt_b^* - \gamma f) - v(t_b^*) + \mu\Big[u(I) \right. \\
&\quad \left. - u'(I)\left(p_d d + nzwt + n\psi(1 - \gamma)f\right) + h(d) + s(nt)\Big]\right\} + \frac{u(bt_b^* - \gamma f) + v(t_b^*)}{1 - \beta}
\end{aligned}
$$

where $t_b^*$ denotes the optimal use of time after committing malpractice. $t_b^*$ is set to maximize lifetime utility after committing malpractice, which equals $\frac{u(bt - \gamma f, t)}{1 - \beta}$, and therefore does not depend on $d^*, t^*$ and $w$. Hence, $d^*$ and $t^*$ will satisfy the first order conditions of the doctor's problem, which are given by:

$$
\frac{\partial D}{\partial d} = 0 \iff -\beta \rho'(d^*)\left\{ D(d^*, t^*) - \frac{u(bt_b^* - \gamma f) + v(t_b^*)}{1 - \beta} \right\} = \mu\left[p_d u'(I) - h'(d^*)\right] \quad (13)
$$

$$
\frac{\partial D}{\partial t} = 0 \iff u'(\bar{w}^*)w + v'(t^*) + \mu n s'(nt^*) - \mu nzw u'(I) = 0 \quad (14)
$$

where we define $\bar{w}^* = wt^* - p - \gamma f$.

In Section A.1 we prove that $(d^*, t^*)$ satisfying the above conditions will maximize doctor's utility.

## 4.2 Comparative statics in partial equilibrium

Given the complexity of the full equilibrium model, it is not feasible to derive analytical comparative static results. Nevertheless, in order to gain intuition into the consequence of introducing physician labor supply, we consider the partial equilibrium effects of physician wages and tuition costs on diagnostics and consultation time.

### 4.2.1 Choice of diagnostics and consultation hours

**Proposition 5** *The level of diagnostics prescribed by doctors is increasing in the physician wage, the tuition cost and the income level of patients. It is decreasing in the malpractice insurance premium, the post-malpractice wage and the degree of altruism. It does not depend on the wage in the alternative occupation.*

Short proof:
In equation (13), the first order condition for diagnostics, the optimal level of diagnostics is expressed as a function of optimal physician consultation time $t^*$. However, $t^*$ appears

in this equation only inside the function that is optimized, $D(d,t)$. We can therefore apply the envelope theorem, according to which:

$$\frac{dD(d^*,t^*)}{d\eta} = \frac{\partial D(d^*,t^*)}{\partial \eta}$$

Similarly, because $t_b^*$ is chosen to maximize lifetime utility after malpractice: [14]

$$\frac{d[u(bt_b^* - \gamma f) + v(t_b)]}{d\eta} = \frac{\partial[u(bt_b^* - \gamma f) + v(t_b)]}{\partial \eta}$$

Hence, we do not have to use the total derivatives of $t^*$ and $t_b^*$ with respect to all parameters. Thus, the total derivatives of $d^*$ are given by:

$$\frac{dd^*}{dw} = B_1 V_1^{-1} t^* u'(\bar{w}^*) > 0$$

$$\frac{dd^*}{d\gamma} = B_1 V_1^{-1} f \left[ u'(\bar{b}) - u'(\bar{w}^*) + \mu n \psi u'(I) \right] > 0$$

$$\frac{dd^*}{dp} = B_1 V_1^{-1} \left[ -u'(\bar{w}) \right] < 0$$

$$\frac{dd^*}{db} = B_1 V_1^{-1} \left[ -u'(\bar{b}) \right] < 0$$

$$\frac{dd^*}{d\mu} = -B_1 V_1^{-1} \mu^{-1} \left[ u(\bar{w}) + v(t^*) - u(\bar{b}) - v(t_b^*) \right] < 0$$

$$\frac{dd^*}{dw_o} = 0$$

$$\frac{dd^*}{dp_d} = V_1^{-1} \mu u'(I) \left[ -B_1 d^* + 1 \right] < 0$$

$$\frac{dd^*}{dI} = V_1^{-1} \left\{ B_1 \mu \left[ u'(I) - u''(I) (p_d d^* + nzwt^* + n\psi(1-\gamma)f) \right] + \mu p_d u''(I) \right\} > 0$$

$$\frac{dd^*}{dn} = B_1 V_1^{-1} \mu \left[ t^*(s'(nt^*) - zwu'(I)) - u'(I)\psi(1-\gamma)f \right] \overset{s'(nt) > u'(I)\left[zw + \frac{\psi}{t^*}(1-\gamma)f\right]}{>} 0$$

where $B_1 = \frac{\beta \rho'(d^*)}{1 - \beta(1 - \rho(d^*))}$, $B_1 < 0$, and $V_1 = -\beta \rho''(d^*) \left\{ D(d^*,t^*) - \frac{u(bt_b^* - \gamma f)}{1-\beta} \right\} + \mu h''(d^*)$, $V_1 < 0$. QED

**Proposition 6** *The level of physician consultation hours is decreasing in the physician wage if income effects are sufficiently strong (in a sense made precise below). Physician hours are increasing in the tuition cost, the malpractice premium, the income level of patients and the level of altruism. They do not depend on the post-malpractice wage or the wage in the alternative occupation.*

---

[14]for all $\eta$ other than $\beta$

Short proof:

First, consider the effect of wages on the amount of time doctors decide to work:

$$\frac{dt^*}{dw} = \frac{u'(\bar{w}^*) + wt^*u''(\bar{w}^*) - \mu nzu'(I)}{-[w^2u''(\bar{w}^*) + v''(t^*) + \mu n^2 s''(nt^*)]}$$

Because $v, u$ and $s$ are concave, the denominator is negative. This derivative will be negative, i.e. doctors with high wages will spend less time working, iff:

$$u'(\bar{w}^*) + wt^*u''(\bar{w}^*) < \mu nzu'(I) \tag{15}$$

This condition requires that the income effect is sufficiently strong relative to the substitution effect.

The remaining results are obtained as follows:

$$\frac{dt^*}{d\gamma} = V_2^{-1}\left[-fwu''(\bar{w}^*)\right] > 0$$

$$\frac{dt^*}{dp} = V_2^{-1}\left[-wu''(\bar{w}^*)\right] > 0$$

$$\frac{dt^*}{dI} = V_2^{-1}\left[-nzwu''(I)\right] > 0$$

$$\frac{dt^*}{d\mu} = V_2^{-1}\left[ns'(nt^*) - nzwu'(I)\right] \overset{s'(nt)>zwu'(I)}{>} 0$$

$$\frac{dt^*}{dn} = V_2^{-1}\left[\mu\left[s'(nt^*) - zwu'(I) + nt^*s''(nt^*)\right]\right] \overset{s'(nt)>u'(I)\left[zw+\frac{\psi}{t^*}(1-\gamma)f\right]}{>} 0$$

$$\frac{dt^*}{db} = 0$$

$$\frac{dt^*}{dw_o} = 0$$

$$\frac{dt^*}{dp_d} = 0$$

where $V_2 = -[w^2u''(\bar{w}^*) + v''(t^*) + \mu n^2 s''(nt^*)], V_2 > 0$. $QED$

Comment 1: Consider the case when the utility of consumption is given by $u(c) = \log(c)$. Then the left hand side of (15) is given by:

$$u'(\bar{w}^*) + wt^*u''(\bar{w}^*) = \frac{\bar{w} - wt}{\bar{w}^2} = \frac{-(p + \gamma f)}{\bar{w}^2} < 0$$

Therefore, doctor consultation time is decreasing in the physician wage provided that $p + \gamma f > 0$. That is, the physician's fixed costs (malpractice insurance plus student loan

interest) are positive.

Comment 2: It is notable that when the utility of consumption is given by $\log(c)$ we find that the time physicians spend with patients is decreasing in the wage. This may seem surprising, as with log utility income and substitution effects on labor supply normally cancel. However, in the present case, in order to practice, doctors need to pay off their debt and cover their malpractice insurance even if they work zero hours. Because of this, the income effect dominates the substitution effect. And the higher is the cost of practicing, the greater is the distortion of doctors' working times.

Comment 3: Combining Propositions 5 and 6, we see that higher wages cause doctors to substitute diagnostics for consultation time provided that condition (15) holds.

## 4.3 Equilibrium

### 4.3.1 The Decision to Become a Physician

With the introduction of physician labor supply, the determination of the physician wage in equilibrium becomes somewhat more complicated. So we revisit the decision to become a physician. In order to become a doctor, one has to go through training for $l$ years. During each period of training, medical students/residents consume $c_T$ and work $t_T$ hours, which brings them utility $u(c_T, t_T)$.

In every period after training, doctors receive the physician wage $w$, make the training debt repayment $\gamma f$ and pay the malpractice insurance premium $p$. However, if a doctor is found guilty of malpractice, which happens with probability $\rho(d)$ in each period, his/her wage drops to the lower level $b$, and training costs must still be repaid. Thus, the expected value of a medical career after finishing training is:

$$D(d,t) = \frac{1}{1-\beta(1-\rho(d))}\left[u(wt - \gamma f - p, t) + \mu K(d, \mathbf{nt}) + \frac{\beta\rho(d)}{1-\beta}u(bt_b - \gamma f, t_b)\right]$$

where $t$ denotes the physician choice of hours and $t_b$ denotes the choice of hours in the post-malpractice environment.

A potential medical student can instead choose the alternative occupation which does not require training, pays wage $w_o$ and brings utility $u(w_o t_0, t_0)$ in each period, for an infinite number of periods. Here $t_0$ denotes the choice of hours in the alternative occupation. Note that the solution for optimal levels of $t_b$ and $t_0$ are standard so we do not describe then here.

Thus, a potential doctor will pursue medical training iff:

$$\sum_{k=0}^{l-1} \beta^k U(c_T, t_T) + \beta^l D(d,t) \geq \frac{U(w_o t_0, t_0)}{1-\beta}$$

24

It is convenient to rewrite the reservation wage condition as:

$$D(d,t) \geq \frac{U(w_o t_0, t_0) + (\beta^{-l} - 1)(U(w_o t_0, t_0) - U(c_T, t_T))}{1 - \beta}$$

Letting $u_0(w_o, t_0, c_T, t_T, l)$ denote the numerator of this equation, we can now write that a person will persue a medical career iff:

$$D(d,t) \geq \frac{u_0(w_o, t_0, c_T, t_T, l)}{1 - \beta}$$

### 4.3.2 Equilibrium Conditions

In the equilibrium, doctors choose optimal levels of diagnostics and work time $(d^*, t^*)$, and the physician wage is set to the level $w^*$ that equates the values of the medical career and the alternative occupation. As before, the malpractice insurance premium is set so that it covers the malpractice fine $M$ times the risk of malpractice $\rho(d^*)$ along with a proportional loading $\alpha$ over the actuarially fair price. Also, to add greater realism, we now assume that the risk of a fine exceeds the risk of losing one's license by the factor $\Psi$. This gives rise to the following equilibrium conditions:

$$\frac{\partial D}{\partial d} = 0 \iff -\beta\rho'(d^*)\left\{D(d^*, t^*) - \frac{u(bt_b^* - \gamma f) + v(t_b^*)}{1 - \beta}\right\} = \mu\left[p_d u'(I) - h'(d^*)\right] \quad (16)$$

$$\frac{\partial D}{\partial t} = 0 \iff u'(\bar{w}^*)w^* + v'(t^*) + \mu n\left[s'(nt^*) - zw^* u'(I)\right] = 0 \quad (17)$$

$$D(d^*, t^*) = \frac{u_0(w_o, t_0^*, c_T, t_T, l)}{1 - \beta} \quad (18)$$

$$p = (1 + \alpha)\Psi\rho(d^*)M \quad (19)$$

where $\bar{w}^* = w^* t^* - \gamma f - p$.

Note that, given the functional form and other simplifying assumptions laid out in Section 4.1, we have that $D(d,t)$ is given by:

$$\begin{aligned} D(d,t) = \frac{1}{1 - \beta(1 - \rho(d^*))}&\left\{u(w^* t^* - \gamma f - p) + v(t^*) - u(bt_b^* - \gamma f) - v(t_b^*) + \mu\big[u(I)\right.\\ &\left. - u'(I)\left[p_d d^* + nzw^* t^* + n\psi(1 - \gamma)f\right] + h(d^*) + s(n^* t^*)\big]\right\} + \frac{u(bt_b^* - \gamma f) + v(t_b^*)}{1 - \beta} \end{aligned}$$

This model is too complex to yield analytic solutions, but we can can obtain comparative static results for the special case where the malpractice premium is fixed exogenously. This case is of independent interest because malpractice premiums are heavily regulated by State insurance regulators, so it may be reasonable to assume they do not adjust to equilibrium levels quickly. Furthermore, we can verify in the calibration/simulation exercise (see Section 5) that many key features of the model are not affected much by this assumption.

### 4.3.3 Comparative Static Results the Case of a Fixed Insurance Premium

We consider a simplified equilibrium in which we assume the insurance premium $p$ is fixed exogenously, so we do not include condition (19). In this case we can calculate comparative statics by total differentiation of the following conditions $F_d$, $F_t$ and $F_w$:

$$F_d: \quad -\frac{\beta\rho'(d^*)\left[u_0(w_o, t_0^*, c_T, t_T, l) - u(bt_b^* - \gamma f) - v(t_b^*)\right]}{1-\beta} + \mu\left[-p_d u'(I) + h'(d^*)\right] = 0$$

$$F_t: \quad u'(\bar{w}^*)w^* + v'(t^*) + \mu n\left[s'(nt^*) - zw^* u'(I)\right] = 0$$

$$F_w: \quad u(\bar{w}^*) + v(t^*) + \mu\left[u(I) - u'(I)\left(p_d d^* + nzw^* t^* + n\psi(1-\gamma)f\right) + h(d^*) + s(nt^*)\right]$$
$$-u_0(w_o, t_0^*, c_T, t_T, l) - \frac{\beta\rho(d^*)}{1-\beta}\left[u_0(w_o, t_0^*, c_T, t_T, l) - u(bt_b^* - \gamma f) - v(t_b^*)\right] = 0$$

For any quantity of interest X we obtain:

$$\frac{dd^*}{dX} = \frac{\frac{\partial F_d}{\partial X}}{\rho''(d^*)\frac{\beta}{1-\beta}(u_0 - u_b) - \mu h''(d)}$$

$$\frac{dw^*}{dX} = \frac{-\frac{\partial F_w}{\partial X}}{t^*\left[u'(\bar{w}^*) - nz\mu u'(I)\right]}$$

$$\frac{dt^*}{dX} = \frac{\frac{\partial F_t}{\partial X} + \frac{dw^*}{dX}\left[u''(\bar{w}^*)w^* t^* + u'(\bar{w}^*) - nz\mu u'(I)\right]}{-\left[u''(\bar{w}^*)w^{*2} + v''(t^*) + \mu n^2 s''(nt^*)\right]}$$

The following three propositions characterize the equilibrium behavior of diagnostics, the physician wage, and physician consultation hours, respectively, with respect to the exogenous variables in the model:

**Proposition 7** *The equilibrium level of* diagnostics *is increasing in medical tuition fees, the opportunity wage, hours worked during medical training, the length of medical training and patient's income. The level of diagnostics is decreasing in the post-malpractice wage, consumption during medical training, doctors' altruism towards patients, and the price of diagnostics. Moreover, the equilibrium level of diagnostics does not depend on the malpractice insurance premium.*

Proof: See Appendix.

**Proposition 8** *Assume* [15]

$$u'(\bar{w}^*) - n\mu u'(I) > 0 \tag{20}$$

---

[15]With log utility: $u'(\bar{w}^*) - n\mu u'(I) > 0 \iff I > \mu nw^*$. If $n = 0.0032$ and $\mu < 1$, it suffices for $I > \mu n\bar{w}^*$ that the income of the patient is at least 0.32% of the net income of the doctor.

*Then the equilibrium* physician wage *is increasing in medical tuition fees, the opportunity wage, hours worked during medical training, the length of medical training, the price of diagnostics and the malpractice insurance premium. The physician wage is decreasing in the post-malpractice wage, consumption during medical training, doctors' altruism towards patients and patients' income.*

Proof: See Appendix.

Comment: Intuitively, doctors derive more utility from treating higher income patients, as such patients demand higher quality care, and altruistic doctors derive utility from delivering high quality care.

**Proposition 9** *Assume*[16]

$$-u''(\bar{w}^*)w^*t^* > u'(\bar{w}^*) - n\mu u'(I) > 0 \tag{21}$$

*Then the equilibrium level of physician* consultation hours *is increasing in the post-malpractice wage, doctors' altruism towards patients, the malpractice premium, and patients' income. Consultation hours are decreasing in the opportunity wage, the price of diagnostics and both the length of medical training and hours worked during medical training. The sign of the relationship between consultation hours and the level medical tuition fees is ambiguous.*

Proof: See Appendix.

Comment: It is interesting that the effect of medical tuition on the level of consultation time is ambiguous, while the effects of opportunity costs of become a doctor (as reflected in the length of medical training, hours worked during medical training and the outside opportunity wage) are negative. Both tuition and opportunity costs raise the physician wage, creating a negative income effect on consultation hours. The asymmetry arises because the tuition cost is borne *after* training is complete (when the loan is repaid) and so it shifts downward the budget constraint of practicing physicians. This budget constraint creates a *positive* income effect that tends to *increase* consultation hours.

Table 1 summarizes the comparative static results for diagnostics, consultation time and the physician wage. A key result is that higher tuition fees (as well well as higher opportunity costs of becoming a doctor) increase both diagnostic spending and physician wages. In contrast to the simple model of Section 2, here we cannot look directly at the effect of malpractice risk, which is now endogenous. However, the bottom four rows of the table refer to parameters of the malpractice risk and other functions that we will assume for the calibration exercise in the next section. Suffice it to say we will assume that malpractice risk is determined by the function $\rho(d) = \theta \exp(-a_\rho d)$, so that parameter $\theta$ increases malpractice risk for any given level of diagnostics. We can see in table 1 that an increase in $\theta$

---

[16]satisfied with log utility

will increase diagnostics, increase the physician wage and reduce consultation time. Thus we have that greater malpractice risk shifts the practice of medicine away from consultation time towards diagnostics (at least for this functional form).

Table 1: Comparative statics in the equilibrium with fixed premium $p$

| $x$ | $\frac{dd^*}{dx}$ | $\frac{dw^*}{dx}$ | $\frac{dt^*}{dx}$ |
|---|---|---|---|
| $\gamma$ | + | + | |
| $b$ | - | - | + |
| $w_o$ | + | + | - |
| $\mu$ | - | - | + |
| $p_d$ | - | + | - |
| $I$ | + | - | + |
| $p$ | 0 | + | + |
| $c_T$ | - | - | + |
| $t_T$ | + | + | - |
| $l$ | + | + | - |
| $n$ | 0 | | |
| $\theta$ | + | + | - |
| $\nu_t$ | 0 | + | - |
| $a_d$ | + | - | + |
| $a_t$ | 0 | - | + |

# 5    Calibration

In this section we analyze the quantitative behavior of the model using calibration and simulation. This forms the heart of the analysis, as we learn about the quantitative behavior of the model given realistic parameter values. Our strategy is to calibrate the model to features of the US economy. We will then investigate if changes in medical tuition fees and malpractice risk that bring them more into line with the lower levels that prevail in the UK can help explain much of the difference between US and UK health costs.

## 5.1    Functional Form Assumptions

In order to calibrate the model of Section 4 we need to specify functional forms for preferences and technology. These are listed in Table 1. The first set of rows present the doctor's utility function. We assume it depends on log consumption, log leisure and patient utility, and is additively separable between the three. Note that the doctor's total time is normalized to 1, so that leisure is given by 1-t, where t is the fraction of time devoted

to work. The parameters $\nu$ and $\mu$ determine how the doctor weights leisure and patient utility relative to own consumption.

The next set of rows describe the patient utility function $K$, which depends on consumption and health.[17] The health production function is specified as additively separable between diagnostics and the doctor time input, and to be logarithmic in each. The parameter $a_d$ determines the productivity of diagnostics, and the parameter $a_t$ determines the productivity of physician consultation time. Given the log specification, there are diminishing return to both inputs. We included a quadratic term in doctor hours to accentuate this concavity, as otherwise we found that patient demand for doctor consultation time was too great at plausible wage rates.

Next we have the malpractice probability function, which we assume is an exponential function of diagnostics with parameters $\theta$ and $a_\rho$. Thus the complete set of parameters to be calibrated is $\nu$, $\mu$, $a_d$, $a_t$, $\theta$ and $a_\rho$. We choose these to match as closely as possible a set of calibration targets that we discuss in the next section.

Finally, the last three rows of the table describe the doctor's utility after malpractice and in the alternative occupation, and the equation for the malpractice insurance premium. The next section describes how we use various data sources to construct the opportunity wage, post-malpractice wage and the insurance premium loading. The model contains several additional parameters that we also take directly from the data, such as doctors per capita $(n)$, student loan payments $(\gamma f)$, and so on (see below).

## 5.2   Data and Calibration

In this section we describe the data and procedures we use to calibrate the model parameters. The complete set of parameters and their values is summarized in Table 3. We express all monetary values in US$1000  and we attempt to match US figures for 2013 (or as close to that year as possible).

### 5.2.1   Parameters Set Directly from Data

GDP per capita is set to $53k$, while the income of the patient, $I$, is set to $37k$. This is the difference between average income in the US, $45k$ (US statistics for 2013), and the part of health spending that is not explained by our model, which is around 15% of GDP per capita.[18]

---

[17]Recall that we do not specify how patient utility depends on leisure as this is assumed invariant to physician behavior, just as we do not have to specify how doctor utility depends on health.

[18]Health spending per capita is expressed as the sum of spending on doctors' gross wages, expenditure on diagnostics, nurses' wages, pharmaceuticals and administration in healthcare.

Table 2: Functions used in the calibration

| functional form | explanation |
| --- | --- |
| $U = u(wt - \gamma f - p) + v(t) + \mu K(d,t)$ | doctor's per period utility |
| $u(c) = \log(c)$ | utility of income |
| $v(t) = \nu_t \log(1 - t)$ | utility of leisure |
| $wt - \gamma f - p$ | doctor's net income (consumption) |
| | |
| $K(d,t) = u^P(I - p_d d - n(zwt + \psi S)) + h(d) + s(nt)$ | patient's utility |
| $h(d) = a_d \log(1 + d)$ | productivity of diagnostics |
| $s(nt) = a_t \log(1 + nt - (nt)^2)$ | patient's utility of time spent with doctors |
| $u^P(I - c) \approx u(I) - u'(I)c$ | patient's utility from consumption |
| $I - p_d d - n(zwt + \psi S)$ | patient's consumption |
| | |
| $\rho(d) = \theta \rho_0(d)$ | malpractice probability |
| $\rho_0(d) = \exp(-a_\rho d)$ | 'technology' of reducing malp. risk |
| | |
| $u(bt_b^* - \gamma f, t_b^*) = \log(bt_b^* - \gamma f) + \nu_t \log(1 - t_b^*)$ | per period doctor's utility after malpractice |
| $u(w_o t_0^*, t_0^*) = \log(w_o t_0^*) + \nu_t \log(1 - t_0^*)$ | per period utility at alternative occupation |
| | |
| $p = (1 + \alpha)\Psi\rho(d)M$ | malpractice insurance premium |

To calibrate the number of doctors per capita we pool together physicians and dentists. There are 3.2 doctors per 1000 patients (2.6 physicians and 0.6 dentists, see OECD data for ****). Thus, we set $n = 0.0032$.

The opportunity income of doctors is set equal to the average salary of college graduates in the US, which was to $60k$ in 2013 (census data, 2014). We assume that the training time required to become a doctor is 7 years, which includes 4 years of medical school and 3 tears of residency. We assume that during medical school individuals do not receive any income, while during the residency they receive $49k$ per annum (**citation**). Discounting this amount with a 7% annual interest rate gives annual consumption of $20k$ during training.

During medical school and residency, individuals have to work very long hours. We set the required hours during raining at 90 per week. Moreover, in the calibrated model we assume that there is a 5% probability of dropping out of the medical training (**citation**). This was suppressed in the theoretical model of Section 4 for notational simplicity. Those who drop out of medical training earn the opportunity wage, but still have to pay back

their educational debt.

After finishing the training, doctors are expected to repay 14 per year as for their training debt. This is an estimate based on the size of the average debt of a medical graduate, which equals 180k (AAMC 2014). In order to take into account that part of the borrowed funds do not come from banks, we round up the total debt to $200k$, and we assume it is repaid over an infinite horizon, with an interest rate of 7% (interest rates range between 6 and 8%) (**citation**).

We also need to calibrate the level of medical education subsidies. According to XXX, 15bn $ is spent yearly on subsidies to residency programs. However, literature YYY suggests that this transfer of money should not be treated to subsidy because ... [19]. Therefore, our model we assume that doctors pay the full costs of medical training.[20] In order to calibrate the cost to society of increasing the supply of doctors, we assume that the marginal cost of funds (or deadweight loss coefficient) is qual to [21].

Finally we assume that $b$, the wage earned after committing malpractice, is simply equal to the opportunity wage in the alternative occupation (i.e., $b = w_o$). We consider different values of $b$ in the robustness checks.

### 5.2.2 Data Targeted in the Calibration

The model parameters are calibrated so as to fit, as closely as possible, the following targets: (i) the physician wage, (ii) the share of health spending on diagnostics, (iii) physician work hours, (iii) the share of health spending in GDP, (iv) the malpractice insurance premium, and (vi) the probability of a physician losing his/her license due to malpractice.

The physician wage is set to $230k$ per year (statistics range from $195k$ to $284k$) (**citation**). We assume that the gross wage including fringe benefits is 50% higher (i.e, we set $z = 1.5$). Moreover, given that there are 3.2 doctors per 1000 inhabitants, per capita expenditure on doctors is $1104.

Calculating the level of spending on diagnostics in the US is very difficult. Given the way the National Accounts are constructed, medical spending is categorized either by the source of funds (e.g., Medicare, Medicaid, private insurers, the VA etc.) or by very broad categories of service (e.g., hospital care, physician and clinical services, home health care, etc.). It is very difficult to find spending on specific medical services such as diagnostics, chemotherapy, anesthesia, deliveries, etc. Thus, for purposes of this study, we made a sub-

---

[19]add arguments here

[20]In a robustness check we consider a case where the subsidy is the full $15bn, which is equivalent to 800k per resident. Assuming interest rate of 7% this is equivalent to a subsidy of 56k per year of medical training

[21]add number here

stantial effort to develop reasonable estimates of diagnostic spending in both the US and the UK. The details of our calculations of diagnostic spending are discussed in **Appenidix ****.

We estimate that diagnostics account for approximately 11.6% of total health expenditure in the US. Given that 17.6% of GDP was spent on healthcare in ****, this means that a substantial 2% of GDP is spent on diagnostics. In contrast, we estimate that in the UK only 0.5% of GDP is devoted to diagnostic spending. This is a substantial difference between the two countries that our model seeks to explain.

We set the price of diagnostics in the US equal to $1k$ per procedure. This is consistent with the data suggesting that the average price of an MRI scan in 2013 in the US was \$1145 (see IFHP Price Report).[22]

Another target of calibration is the yearly probability of a practicing physician losing his/her license due to malpractice. We set this probability based on the number of licenses revoked, suspended or surrendered by physicians in 2012, as reported in the National Practitioner Data Bank. By combining this information with the number of physicians (see Physician Workforce report), we obtain an estimated malpractice probability of 0.2%. We set the probability of a malpractice fine at 1.4%, which is a compromise between estimates in the literature (NEJM (2011) reports 1.60% using data on payments in years 1995-2005, while NPDB (****) reports and 1.05% using data on payments in 2012). This means we set $\Psi$, the ratio of the probability of a fine to losing ones license, equal to 7. The average malpractice fine $M$ is set to 300k (NPDB data).

We also target the price of malpractice insurance. Recall that in the model the annual price of malpractice insurance is $p = (1 + \alpha)M\Psi\rho(d_D^*)$. Using NAIC data, we estimate that the average annual malpractice insurance premium is roughly $10k$ [23]. And we set the loading $\alpha$ to 1.38, based on NAIC data (**citation**) indicating the ratio of insurance payouts to revenues equals 0.42.

Finally, we target physician hours of work. We express hours worked as the share of available time spent on market work. Following Prescott (2004) and Osuna and Rios-Rull (2003), we set the available hours per week equal to 100 hours. The values of hours worked by practicing doctors and in the opportunity occupation are both targets. Erosa, Fuster

---

[22]Other components of health expenditure that are assumed constant include spending on pharmaceuticals - we set it equal to 1.034 (based on an OECD estimate), and spending on administration in healthcare, which we set equal to 0.667 (see Health Affairs 2014). Moreover, we add a component equal to the per capita spending on nurses. Wages of nurses are assumed equal to 70 (lower bound estimate based on BLS data). Given that there are 11 nurses per 1000 inhabitants, this means that the amount spent per capita on nurses is set equal to 0.770.

[23]This is a lower bound estimate, based on the total value of premiums paid divided by the number of physicians. If some doctors do not have insurance, or pay lower insurance if they work part-time, this figure should be higher

and Kamourov (2015) estimate that college graduates in the US spend on average 44 hours per week on work. Hence, we would like to set the hours worked at opportunity occupation to 44%. According to OECD estimates, GPs, who represent between 35 and 48% of doctors in the US, work 51 hours per week, while specialists work 54. We assume therefore that practicing physicians work 53 hours per week [24].

## 5.3 Identification of the Calibrated Parameters

In this section we describe how we calibeate the m,odel parameters to match the data targets. We first focus on setting spending on diagnostics to 2% of GDP in the equilibrium, and 0.50% of GDP at the social optimum. Then, we set the resulting risk of malpractice equal to 0.05%. These two steps combined allow us to calibrate parameters $p_d, a_d, \theta, a_\theta$. Finally, we set wages equal to 230, which allows us to calibrate $\mu$.

### 5.3.1 Spending on diagnostics

We first focus on spending on diagnostics. In the model, we distinguish between spending on diagnostics at the social optimum $(p_d d_P^*)$ or in the equilibrium (at doctor's optimum $p_d d_D^*$).

First, we set the share of spending on diagnostics in GDP at the social optimum equal to this share in the UK, where spending is not distorted by high risk of malpractice. According to our estimates, based on diagnostic activities performed in the NHS and by public healthcare providers, in the UK 6.1% of health spending is spent on diagnostics, which is equal to 0.54% of GDP. With the current choice of $K(d, t)$, the social optimum of diagnostics is given by

$$d_P^* = \frac{a_d I - p_d}{p_d}$$

Then the share of spending on diagnostics at the social optimum equals:

$$\frac{p_D d_P^*}{GDP} = \frac{a_d I - p_d}{GDP}$$

To set this value equal to $\epsilon_{GDP} = 0.5\%$ and the price of diagnostics equal to $p_d = 1$, we set $a_d$ to

$$a_d = \frac{p_d + \epsilon_{GDP} \, GDP}{I} = 0.0341$$

Second, there are two targets related to the level of diagnostics in the equilibrium $d_D^*$: spending on diagnostics and the risk of malpractice. Setting spending on diagnostics equal to 2% of GDP, allows us to pin down the level of $d_D^*$ : $d_D^* = \frac{2\% \, GDP}{p_d} = 1.06$.

---

[24] Staiger, Auerbach 2010: doctors work on average 51 hours per week

### 5.3.2 The risk of malpractice

The probability of malpractice is modeled by $\rho(d) = \theta e^{-a_\rho d}$. Setting it equal to $\epsilon_\theta = 0.2\%$ requires:

$$\theta(a_\rho) = \epsilon_\theta e^{a_\rho d_D^*}$$

This allows us to define $\theta$ as a function of $a_\rho$.

Third, $a_\rho$ is calibrated so that $d_D^*$, the solution of the model, is approximately equal to 1.06. $a_\rho$ is calibrated simultaneously with $\mu$ and $a_t$. This gives $a_\rho = 0.0079$ and $\theta = 0.2017\%$.

### 5.3.3 Time worked

First, hours worked in the alternative occupation satisfy the following equation:

$$v'(t_0) = -\frac{1}{t_0}$$

Using $v(t) = \nu_t \log(1-t)$ allows us to calibrate the parameter $\nu_t$:

$$\nu_t = \frac{1 - t_0}{t_0} \stackrel{t_0 = 0.44}{=} 1.2727$$

In order to adjust hours worked in the equilibrium, we can adjust the remaining free parameters $a_t$ or $\mu$ - doctors' hours increase in both $a_t$ and $\mu$. $\mu$ is set to adjust doctors' wages. We end up with $a_t = 3877.6$. The value of this parameter seems high, relative to the weight of utility of consumption, which is equal to 1, and utility from diagnostics, which is equal to 0.0341 Note however that the units of measurement of doctor's time in the utility function are very small, and time is additionally multiplied by $n = 0.0032$.

### 5.3.4 Wages

Finally, we adjust $\mu$ to set yearly wages of doctors to around 230k. This gives $\mu = 0.0487$.

### 5.3.5 Calibration of $\mu, a_t, a_\rho$

If $s(t) = a_t \log(1 + nt)$, then it is possible to calibrate $a_t$ as a function of $\mu$ using first order condition for $t^*$ (??):

$$u'(\bar{w}^*)w^* + v'(t^*) + \mu n s'(nt^*) = 0 \iff a_t = \frac{1 + nt^* - (nt^*)^2}{n(1 - 2nt^*)\mu} [-u'(\bar{w}^*)w - v'(t^*)]$$

Then:

$$\mu s(nt) = \mu \, a_t(\mu) \, \log(1 + nt^*) = [-u'(\bar{w}^*)w^* - v'(t^*)] \frac{1 + nt^* - (nt^*)^2}{n(1 - 2nt^*)} \log(1 + nt^* - (nt^*)^2)$$

In order to calibrate $\mu$, we can plug in this expression to (??):

$$U(\bar{w}^*, t^*) + \mu [u^P(c_P) + h(d^*)] + \mu s(nt^*) = u_0 + \frac{\beta \rho(d^*)}{1 - \beta} [u_0 - U(bt_b - \gamma f, t_b)] \iff$$

34

$$\mu \;=\; \frac{u_0 - U(\bar w^*, t^*) + \frac{\beta \rho(d^*)}{1-\beta}\left[u_0 - U(bt_b - \gamma f, t_b)\right] - \left[-u'(\bar w^*)w^* - v'(t^*)\right]\frac{1 + nt^* - (nt^*)^2}{n(1-2nt^*)}\log(1 + nt^* - (nt^*)^2)}{u^P(c_P) + h(d^*)}$$

where $c_P = I - p_d d^* - n z w^* t^* - n\psi(1-\gamma)f$.

We calculate $\mu$ using target values of $d^*, w^*, t^*, \rho(d^*)$ and parameters $\nu_t, a_d$ that have already been calibrated. This allows us to calibrate $a_t$ in the next step, using the calibrated value of $\mu$.

If order to calibrate $a_\rho$ we use the equilibrium condition for diagnostics (**??**)

$$-\frac{\beta \rho'(d^*)}{1-\beta}\left[u_0 - U(bt_b^* - \gamma f, t_b^*)\right] = \mu\left[p_d u'(I) - h'(d^*)\right]$$

By plugging in $\rho'(d) = -a_\rho \rho(d)$ we obtain:

$$a_\rho = \frac{1-\beta}{\beta}\,\frac{\mu\left[p_d u'(I) - h'(d^*)\right]}{\rho(d^*)\left[u_0 - U(bt_b^* - \gamma f, t_b^*)\right]}$$

We calculate $a_\rho$ using target values of $d^*, \rho(d^*)$ and parameters $\nu_t, a_d, \mu$ that have been already calibrated.

# 6  Counterfactual experiments

## 6.1  Welfare analysis and the effects on healthcare spending

Calibrating the model allows us to quantify the effects of counterfactual experiments on the share of explained health spending in the GDP, on welfare and on health of the patients. We define health spending per capita explained by our model as the sum of spending on diagnostcs, doctors' wages and subsidies to medical training:

$$HS = p_d d^* + n(z w^* t^* + \psi(1-\gamma)f)$$

where $\psi(1-\gamma)f$ denotes per doctor cost of subsidies to medical education. Note that we assume that gross wage of a doctor equals 1.5 times his net wage.

The utility of the patient is a sum of the utility from consumption, which is a function of health spending, and patient's health produced by diagnostics and time spent with the doctors:

$$K = u(I) - u'(I)\left[p_d d^* + n(z w^* t^* + \psi(1-\gamma)f)\right] + h(d^*) + s(nt^*) = u(I) - u'(I)\cdot HS + H(d^*, nt^*)$$

First, we approximate analytically the effects of experiments on health spending, health and welfare. The change in patient's utility equals the sum of the change in health spending, multiplied by $-u'(I)$, and the change in patient's health.

$$dK = -u'(I)dHS + dH$$

35

Table 3: Parameters used in the calibration

| parameter | value | explanation |
|---|---|---|
| GDP | 53 | GDP per capita |
| I | 37.05 | patient's income - equal to average wage minus unexplained HS |
| n | 0.0032 | number of doctors per capita |
| z | 1.5 | total cost of doctor's wage/ doctor's wage |
| $\gamma f$ | 14.0 | repayment of tuition fee loan, per year |
| $S = (1-\gamma)f$ | 0 | total cost of training subsidy per doctor, per year |
| $\psi$ | 1.3 | coefficient of deadweight loss |
| $\Psi$ | 7.003 | ratio of prob. of being fined for malpractice to prob. of losing the license |
| $\alpha$ | 1.380 | malpractice insurance premium loading (proportional) |
| $M$ | 300 | size of malpractice fine |
| $p_d$ | 1.000 | price per diagnostic |
| $l$ | 7 | length of the training (in years) |
| $c_T$ | 20 | consumption during the training |
| $t_T$ | 90 | number of hours worked per week during the training |
| $p_{noT}$ | 5% | probability of not finishing the training |
| $\beta$ | 0.950 | discount factor |
| | | |
| Targets | | |
| $t_D^*$ | 53 | weekly number of hours worked by practicing doctors |
| $w^* t_D^*$ | 230 | yearly income of a doctor |
| $t_0$ | 44 | weekly number of hours worked per week in the opportunity occupation |
| $w_o t_0$ | 60 | opportunity income |
| $\frac{p_d d_D^*}{GDP}$ | 2% | share of diagnostic spending in GDP in the equilibrium |
| $\frac{p_d d_P^*}{GDP}$ | 0.5% | share of diagnostic spending in GDP at social optimum |
| $\rho(d_D^*)$ | 0.2% | yearly probability of losing the license due to malpractice |
| | | |
| Calibrated | | |
| $\nu_t$ | 1.2727 | weight of leisure in doctor's utility |
| $w_o$ | 136.4 | opportunity wage (per full time) |
| $a_d$ | 0.0341 | patient's weight of the utility of own health relative to consumption |
| $\theta$ | 0.2017% | risk of malpractice with no diagnostics |
| $a_\rho$ | 0.0079 | scaling factor in malpractice probability function |
| $a_t$ | 3877.6 | weight of time spent with the doctor in patient's utility |
| $\mu$ | 0.0487 | how much a doctor cares about the patient's utility |
| | | |
| Assumed | | |
| $b$ | $w_o$ | doctor's wage after committing malpractice |

Consider an experiment that changes the cost of per doctor subsidy by $\Delta S$. If we assume that the supply of doctors is constant, then the effect of of this experiment on health spending, patients' health and their utility can be approximated by

$$
\begin{aligned}
dHS_1 &\approx \frac{dd^*}{d^*}p_d d^* + \frac{dt^*}{t^*}nzw^*t^* + \frac{dw^*}{w^*}nzw^*t^* + n\Delta S \\
dH_1 &\approx \frac{dd^*}{d^*}h'(d^*)d^* + \frac{dt^*}{t}s'(nt^*)nt^* \\
dK_1 &\approx -u'(I)\left[\frac{dd^*}{d^*}p_d d^* + \frac{dt^*}{t^*}nzw^*t^* + \frac{dw^*}{w^*}nzw^*t^* + n\Delta S\right] + \frac{dd^*}{d^*}h'(d^*)d^* + \frac{dt^*}{t}s'(nt^*)nt^*
\end{aligned}
$$

If we allow the supply of doctors to change, then the effect of an experiment on health spending and the utility of patients can be approximated by:

$$
\begin{aligned}
dHS_2 &\approx dHS_1 + \frac{dn}{n}n(zw^*t^* + \psi(1-\gamma)f) \\
dH_2 &\approx dH_1 + \frac{dn}{n}s'(nt^*)nt^* \\
dK_2 &\approx -u'(I)\,dHS_2 + dH_2
\end{aligned}
$$

We consider following experiments: removing doctors' debt (lowering it from 14 to 0), reducing malpractice risk to the level observed in a State with the lowest probability of losing the license - Minnesota, and introducing co-payments for diagnostics for doctors. The results of these experiments are presented in Table 4 and 5. Panel (1) refers to the baseline version of the calibrated model. When doctors are expected to repay 14 per year, the risk of malpractice is 0.2% and prescribing diagnostics is not costly to doctors, then the share of health spending in GDP that our model explains equals 4.08%, with 1060$ spent per capita on diagnostics and 1103 on doctors' wages. The utility of a patient equals 10.17 (from 10.167 to 10.170).

## 6.2   Removing doctor's debt

We first consider removing the medical training debt, by lowering debt repayment in each year from 14 to 0. We consider a long run steady state result on health spending, which requires a whole new generation of doctors to be trained, so it might take about 20 or 30 years for all that saving to materialize. The results of this experiment are presented in panel (2) of Table 4. Removing the training debt lowers not only wages per doctor (from 230k to 205k), but also per capita spending on diagnostics (from 1060 to 886). Taking into account that individuals have to cover costs of doctors' tuition, spending on diagnostics, doctors' gross wages and tuition fees drops from 4.08% to 3.64%, which is a reduction by 0.44 pp GDP. This experiment leads however to a reduction in patients' utility: from 10.17 to 10.01, which is equivalent to lowering patients' consumption by by 5928$ per year. This reduction in utility is caused by a high reduction in patient's health due to shorter time that doctors spend on consulting patients. In order to keep patients' health constant

Table 4: Results of experiments

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Values of parameters used in experiments: | | | | | |
| debt per year in 1000$ | 14 | 0 | 14 | 0 | 28 |
| $\theta$ | 0.00202 | 0.00202 | 0.0005 | 0.0005 | 0.00202 |
| doctor's cost per diagnostic | 0 | 0 | 0 | 0 | 0 |
| Results of experiments with constant number of doctors: | | | | | |
| $n$ (per 1000) | 3.20 | 3.20 | 3.20 | 3.20 | 3.20 |
| d | 1.060 | 0.886 | 0.400 | 0.378 | 1.337 |
| w*t, in 1000$ | 230.0 | 204.7 | 210.0 | 186.4 | 255.6 |
| t | 53.0 | 51.7 | 52.4 | 50.9 | 54.1 |
| malp risk in eq | 0.200% | 0.200% | 0.050% | 0.050% | 0.200% |
| diag pc in $ | 1060 | 886 | 400 | 378 | 1337 |
| 1.5*wages pc in $ | 1103 | 983 | 1008 | 895 | 1227 |
| spending on tuitions pc in $ | 0 | 45 | 0 | 45 | -45 |
| diag+1.5*wages+paid tuitions/GDP | 4.08% | 3.64% | 2.66% | 2.51% | 4.73% |
| patients' utility | 10.17 | 10.01 | 10.10 | 9.91 | 10.30 |
| Results of experiments with adjusted $n$ to keep H constant: | | | | | |
| $n$ (per 1000) to keep H constant | 3.20 | 3.27 | 3.24 | 3.32 | 3.14 |
| diag pc in $ | 1060 | 886 | 400 | 378 | 1337 |
| 1.5*wages pc in $ | 1103 | 1001 | 1018 | 923 | 1208 |
| spending on tuitions pc in $ | 0 | 46 | 0 | 47 | -44 |
| HS/GDP | 4.08% | 3.67% | 2.68% | 2.57% | 4.69% |
| change in HS per capita (in $) | 0 | -216 | -745 | -802 | +324 |
| patients' utility | 10.17 | 10.17 | 10.19 | 10.18 | 10.16 |

indifferent, the supply of doctors would have to be increased from 3.2 to to 3.27 per 1000 patients. This experiment would then result in a net decrease in per capita health spending by 216$ per year and an increase in patients utility up to 10.17.

Increasing the yearly costs of doctors' debt from 14 to 30, results in an increase of spending on diagnostics and doctors' wages to 4.73% of GDP (this result includes the effect of lower costs of tuition fees, see panel (5)).

## 6.3 Lowering malpractice risk

Part of the reason malpractice risk is so high is that Americans are so litigious. But malpractice risk differs a lot by State, either because the people in some States are less litigious than in other, or because the legal institutions differ by State. We consider an experiment "What if malpractice rates could be reduced from the National rate to that in the lowest malpractice rate States?".

We consider a measure of malpractice risk in each State which equals the total number of doctors with licenses revoked, surrendered or suspended during years 2008-2012, per 1000 practicing doctors in a given State. This measure shows a lot of variation between the States, varying between 2.4 (MN) and 3.6 (MA) , up to around 11-11.5 (NC, FL, OH, PA, VA). Hence, the lowest level of malpractice risk is observed in Minnesota, where between years 2008 and 2012, on average 2.4 per 1000 practicing physicians lost their license, which translates to a yearly probability of approximately 0.05%.

We consider an experiment in which malpractice risk in the US is reduced to a level observed in Minnesota. Because malpractice risk is an endogenous measure in our model, we lower it by lowering the parameter $\theta$, which represents risk of malpractice without any diagnostics, from 0.002017 to 0.0005. The results of this experiment are presented in panel (3) of Table 4. The experiment results in a big decrease on the level of diagnostics chosen by doctors (from 1060 to 400 per capita) as well as a reduction in wages (from 230k to 210k). Without increasing the supply of doctors, lower malpractice risk reduces the share of GDP explained by diagnostics and doctor's wages from 4.08% to 2.66%, which is a reduction of 1.42 pp. However, this experiment lowers the utility of patients from 10.17 to 10.10, which is equivalent to a reduction in their consumption by 2594$ per year. In order to keep patients' health constant, the supply of doctors would have to be increased from 3.2 to to 3.24 per 1000 patients. This experiment would then result in a net decrease in per capita health spending by 745$ per year and increase patient's utility to 10.19.

## 6.4 Combining lower malpractice risk with no training debt

Panel (4) of Table 4 presents consequences of combining lower malpractice risk with removing the training debt of doctors. This experiment lowers the share of diagnostics and doctors wages, including extra costs of tuition fees, in GDP to 2.66%, which is a reduction of 1.42 pp. By increasing the supply of doctors to 3.32 per 1000 patients, one can reduce yearly health spending per capita by 802$, keeping health of patients' unchanged and increasing patients' utility to 10.18

## 6.5 Doctors' co-payments for diagnostics

A major difference between US and other countries is the fact that American physicians are paid based on fee for service. Due to the fee for service, doctors do not internalize the costs of treatments and diagnostics they prescribe. We consider an experiment in which doctors are charged a small, constant price $q$ per each diagnostic prescribed. Then, if a doctor prescribes $d$ diagnostics per average patient, then the amount spent by him on diagnostics prescribed for all $\frac{1}{n}$ patients equals $d\frac{q}{n}$. Doctor's decision problem then becomes:

$$\max_{d,t} D(d,t) = \frac{U(wt - \gamma f - p - \frac{q}{n}d, t) - U(bt_b^* - \gamma f, t_b^*) + \mu K(d,t)}{1 - \beta(1 - \rho(d))} + \frac{U(bt_b^* - \gamma f, t_b^*)}{1 - \beta}$$

If $U(c,t) = u(c) + v(t)$, solving this model gives the following first order conditions:

$$-\beta\rho'(d^*)\left\{D(d^*,t^*) - \frac{u(b-\gamma f) + v(t_b^*)}{1-\beta}\right\} = \mu\left[p_d u'(I) - h'(d^*)\right] + \frac{q}{n}u'(\bar{w}^*)$$

$$u'(\bar{w}^*)w + v'(t^*) + \mu n\left[s'(nt^*) - zw^*u'(I)\right] = 0$$

where $\bar{w}^* = w^*t - \gamma f - p - \frac{q}{n}d^*$. In the equilibrium, these conditions are combined with the equilibrium conditions for doctor's wage and for the price of insurance:

$$u(\bar{w}^*) + v(t^*) - u(b-\gamma f) - v(t_b^*) \quad + \quad \mu K(d^*, t^*) =$$
$$= \frac{1 - \beta(1 - \rho(d^*))}{1-\beta}\left[u_0 - u(b-\gamma f) - v(t_b^*)\right]$$
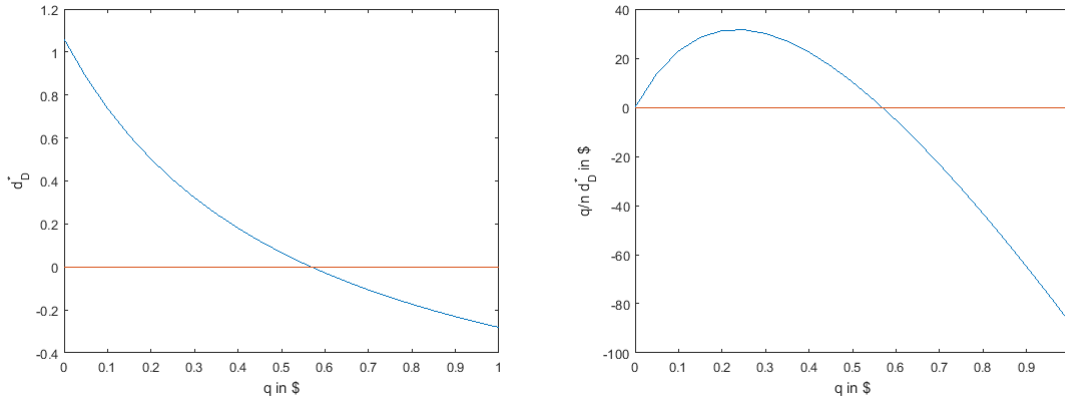$$p = (1+\alpha)\rho(d^*)M$$



Figure 3: Number of diagnostics prescribed per patient and the cost of prescribing diagnostics per doctor as a function of doctor's price per diagnostic

Figure 3 shows that even a small price per diagnostic strongly reduces the amount of diagnostics prescribed in the equilibrium. According to Table 5, setting the doctor's cost per diagnostic to 0.2\$ reduces the amount of diagnostics prescribed per patient from 1.06 to 0.558, which decreases the share of explained health spending in GDP from 4.08 to 3.03%. The total cost of diagnostics for doctors equals then 37\$.

Table 5: Results of experiments introducing the price of diagnostics to the doctor, keeping health of the patient constant

|  | (1) | (2) | (3) |
|---|---|---|---|
| doctor's cost per diagnostic in $ | 0 | 0.2 | 0.5 |
| d | 1.060 | 0.558 | 0.066 |
| w*t, in 1000$ | 229.7 | 229.8 | 229.8 |
| t | 53.0 | 53.0 | 53.0 |
| malp risk in eq | 0.200% | 0.201% | 0.202% |
| $n$ (per 1000) to keep H constant | 3.200 | 3.204 | 3.209 |
| diag+1.5*wages-paid diagnostics/GDP | 4.08% | 3.03% | 2.21% |
| reduction in spending |  | -558 | -991 |
| patients utility | 10.17 | 10.18 | 10.19 |

## 6.6 Robustness - results with higher spending at patient's optimum

Tables 6 and 7 present the results of experiments when with modified values of the key parameters of the model. As before, Table 6 presents results of reducing tuition fees to 0 (panel (2)), reducing malpractice risk to a level in Minnesota (panel (3)), combining reduction of tuition fees with a reduction of malpractice risk (4), and doubling tuition fees (5). Table 7 show the effect of introducing co-payment for diagnostics for doctors, with prices 0.2 (panel (2)) and 0.5 (3).

First, based on the average debt of medical graduates, we have so far assumed that yearly cost of tuition fee repayment $\gamma f$ equals 14. In order to capture costs that are not included in the debt of medical graduates, we calibrate the model allowing $\gamma f$ to be over twice as high, taking value 30. Moreover, following the discussion in the literature XXX, we allow the per doctor size of subsides to medical training to be equal to 800 000$ in total, which is equivalent to paying around 56 per year (4 times the cost of graduate's debt at 200 000$). We recalibrate the model using each combination of the costs of tuition fee and subsidies. Second, we analyze the results of the model if the social optimum of spending diagnostics is at 1% instead of 0.5% of GDP , which captures the fact that prices in the UK are higher than in the US, so the optimal volume of diagnostics should be 50 instead of 25% of what is observed now in the US.

Finally, although our model assumes that malpractice insurance premiums are endogenous and determined by the level of observed risk of malpractice, we could only calculate comparative statics for the case with fixed $p$. Therefore, we calibrate the model in which malpractice insurance premium is fixed at 10.

The results show that the cost subsidies does not have a large effect on the reduction in health spending due to each experiment. For instance, when we assume current subsidies are 0, then reducing tuition fees to zero results in a decrease in health spending by 216$ per capita per year. If we assume the yearly cost per doctor of subsidies is 56, then the reduction is slightly lower, at 211$. Similarly, reducing the risk of malpractice results in a reduction of health spending by 745 or 742$. Finally, the results of introducing co-payment for diagnostics for doctors also do not change much if we increase the costs of subsidies. Assuming that the cost of tuition fees is 30 instead of 14 does not have any effect on our estimate of the effect of reducing malpractice risk. However, because now the level of tuition fees is higher, reducing them to 0 also results in higher drop in health spending: by 370 instead of 216, and by 357 if we assume that the cost of subsidies equals 56. Finally, assuming higher cost of tuition fees decreases the size of the effect of introducing co-payment of 20 cents for diagnostics: from -558 to -320.

Increasing the level of spending on diagnostics in the social optimum reduces the effects of our experiments. This is because higher spending at the social optimum reduces the scope for savings on diagnostics.[25] As a result, cutting tuition fees to 0 results in smaller

---

[25]The reductions in the size of the effects are quite high, so it would be good to add some discussion here why we believe that comparing prices does not make much sense and we should compare spending instead

decrease in health spending: by 142 instead of 216$. The effect of reducing malpractice risk on health spending is also lower: -510 instead of -745$. Finally, co-payments for diagnostics also result in a smaller decrease in health spending : by 484 instead of 558$.

Finally, assuming that insurance premiums are fixed has a negligible effect on the results of each experiment.

Table 6: Results of experiments keeping patient's health constant, depending on parameter values

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| debt per year in 1000$ | $\gamma f$ | 0 | $\gamma f$ | 0 | $2\gamma f$ |
| $\theta$ | 0.00202 | 0.00202 | 0.0005 | 0.0005 | 0.00202 |
| doctor's cost per diagnostic | 0 | 0 | 0 | 0 | 0 |
| **$\gamma \mathbf{f} = 14, \mathbf{S} = 0$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.27 | 3.24 | 3.32 | 3.14 |
| HS/GDP | 4.08% | 3.67% | 2.68% | 2.57% | 4.69% |
| change in HS per capita (in $) | | -216 | -745 | -802 | +324 |
| patients' utility | 10.17 | 10.17 | 10.19 | 10.19 | 10.16 |
| **$\gamma \mathbf{f} = 14, \mathbf{S} = 56$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.27 | 3.24 | 3.32 | 3.14 |
| HS/GDP | 4.52% | 4.12% | 3.12% | 3.03% | 5.13% |
| change in HS per capita (in $) | | -211 | -742 | -793 | +320 |
| patients' utility | 10.15 | 10.15 | 10.17 | 10.17 | 10.14 |
| **$\gamma \mathbf{f} = 30, \mathbf{S} = 0$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.38 | 3.25 | 3.46 | 3.08 |
| HS/GDP | 4.08% | 3.38% | 2.67% | 2.48% | 6.08% |
| change in HS per capita (in $) | | -370 | -745 | -850 | +1061 |
| patients' utility | 5.69 | 5.70 | 5.71 | 5.71 | 5.67 |
| **$\gamma \mathbf{f} = 30, \mathbf{S} = 56$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.38 | 3.25 | 3.46 | 3.08 |
| HS/GDP | 4.52% | 3.85% | 3.12% | 2.95% | 6.50% |
| change in HS per capita (in $) | | -357 | -742 | -831 | +1052 |
| patients' utility | 5.68 | 5.69 | 5.70 | 5.70 | 5.66 |
| **$\mathbf{p_d d_P^*}$/GDP = 0.1, $\gamma \mathbf{f} = 14, \mathbf{S} = 0$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.27 | 3.24 | 3.32 | 3.14 |
| HS/GDP | 4.08% | 3.81% | 3.12% | 3.02% | 4.45% |
| change in HS per capita (in $) | | -142 | -510 | -561 | +196 |
| patients' utility | 10.18 | 10.18 | 10.19 | 10.19 | 10.18 |
| **fixed premium $\mathbf{p} = 10$** | | | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.27 | 3.20 | 3.27 | 3.14 |
| HS/GDP | 4.08% | 3.67% | 2.75% | 2.65% | 4.69% |
| change in HS per capita (in $) | | -216 | -705 | -761 | +325 |
| patients' utility | 10.17 | 10.17 | 10.19 | 10.18 | 10.16 |

Table 7: Results of experiments introducing the price of diagnostics to the doctor, keeping health of the patient constant, depending on parameter values

|  | (1) | (2) | (3) |
|---|---|---|---|
| doctor's cost per diagnostic in $ | 0 | 0.2 | 0.5 |
| $\gamma \mathbf{f} = \mathbf{14}, \mathbf{S} = \mathbf{0}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.204 | 3.209 |
| diag+1.5*wages-paid diagnostics/GDP | 4.08% | 3.03% | 2.21% |
| reduction in spending | | -558 | -991 |
| patients utility | 10.17 | 10.18 | 10.19 |
| $\gamma \mathbf{f} = \mathbf{14}, \mathbf{S} = \mathbf{56}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.204 | 3.209 |
| HS/GDP | 4.52% | 3.47% | 2.65% |
| reduction in spending | | -557 | -989 |
| patients utility | 10.15 | 10.16 | 10.18 |
| $\gamma \mathbf{f} = \mathbf{30}, \mathbf{S} = \mathbf{0}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.208 | 3.217 |
| HS/GDP | 4.08% | 3.47% | 2.87% |
| reduction in spending | | -320 | -641 |
| patients utility | 5.69 | 5.70 | 5.71 |
| $\gamma \mathbf{f} = \mathbf{30}, \mathbf{S} = \mathbf{56}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.208 | 3.217 |
| HS/GDP | 4.52% | 3.92% | 3.31% |
| reduction in spending | | -319 | -639 |
| patients utility | 5.68 | 5.69 | 5.70 |
| $\mathbf{p_d d_P^*}/\mathbf{GDP} = \mathbf{0.1}, \gamma \mathbf{f} = \mathbf{14}, \mathbf{S} = \mathbf{0}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.205 | 3.210 |
| HS/GDP | 4.08% | 3.17% | 2.39% |
| reduction in spending | | -484 | -895 |
| patients utility | 10.18 | 10.20 | 10.21 |
| fixed premium $\mathbf{p} = \mathbf{10}$ | | | |
| $n$ (per 1000) to keep H constant | 3.2 | 3.205 | 3.210 |
| HS/GDP | 4.08% | 3.03% | 2.21% |
| reduction in spending | | -558 | -991 |
| patients utility | 10.17 | 10.18 | 10.19 |

# 7 Discussion and conclusions

# A  Calculation for the main model

## A.1  Proof that $d^*$ and $t^*$ maximize doctor's utility

First derivatives of the function $D(d,t)$ are as follows:

$$\frac{\partial D}{\partial d} = \frac{1}{1-\beta(1-\rho(d^*))}\left[-\beta\rho'(d^*)\left(D(d^*,t^*)-\frac{u(b-\gamma f)}{1-\beta}\right)-\mu p_d u'(I)+\mu h'(d^*)\right]$$

$$\frac{\partial D}{\partial t} = \frac{1}{1-\beta(1-\rho(d^*))}\left\{u'(\bar{w}^*)w+v'(t^*)+\mu n s'(nt^*)-\mu n w u'(I)\right\}$$

If there exists a point $(d^*, t^*)$, such that these first derivatives at $(d^*, t^*)$ are equal to zero, then, at this point, the second order derivatives are the following:

$$\frac{\partial^2 D}{\partial d^2} = \frac{-\beta\rho''(d^*)\left(D(d^*,t^*)-\frac{u(b-\gamma f)}{1-\beta}\right)+\mu h''(d^*)}{1-\beta(1-\rho(d^*))} < 0$$

$$\frac{\partial^2 D}{\partial t^2} = \frac{1}{1-\beta(1-\rho(d^*))}\left\{u''(\bar{w}^*)w^2+v''(t^*)+\mu n^2 s''(nt^*)\right\} < 0$$

$$\frac{\partial^2 D}{\partial d\partial t} = 0$$

Then $D_{dd}D_{tt} - D_{dt}^2 > 0$ and $D_{dd} < 0, D_{tt} < 0$, hence the point $(d^*, t^*)$ is a local maximum.

To prove existence of this maximum, we need to add extra assumptions on the values of functions, such that $D_d$ and $D_t$ cross 0 at some point:[26]

1. $D'(d) > 0$ for $d < d_P^*$:

$$-\beta\rho'(d)\left(D(d,t^*)-\frac{u(b-\gamma f)}{1-\beta}\right)+\mu[-p_d u'(I)+h'(d^*)] > 0$$

   The second term will be positive (by definition of $d_P^*$). The first term will be positive iff $u(b-\gamma f) < u(\bar{w})+\mu K(d,t^*)$ for each $d < d_P^*$, which means that $K$ should not be too low for $d < d_P^*$.

2. $D'(d) < 0$ for $d = d_{max}$:

$$-\beta\rho'(d)\left(D(d,t^*)-\frac{u(b-\gamma f)}{1-\beta}\right)+\mu[-p_d u'(I)+h'(d^*)] < 0$$

   The second term, $\frac{\partial K(d_{max},t^*)}{\partial d}$ is negative. Hence, this condition will be satisfied iff:

$$K'(d_{max}) < \frac{\beta\rho'(d_{max})}{\mu}\left(D(d_{max},t^*)-\frac{u(b-\gamma f)}{1-\beta}\right)$$

   Hence, $K$ must decrease fast enough for high values of $d$

---

[26]For each t, d? Think about the line of the proof

3. $D'(t) > 0$ for $t = t_{min}$:

$$u'(wt_{min} - p - \gamma f)w + \mu ns'(nt_{min}) > -v'(t_{min}) + \mu nwu'(I)$$

4. $D'(t) < 0$ for $t = t_{max}$:

$$u'(wt_{max} - p - \gamma f)w + \mu ns'(nt_{max}) < -v'(t_{max}) + \mu nwu'(I)$$

Consider the case when $t \in [0,1]$ denotes the share of individual time spent working. Then, there will exist a solution $t^* \in [\frac{p+\gamma f}{w^*}, 1]$ iff:

$$u'(0)w^* > -v'(\frac{p+\gamma f}{w^*}) - \mu ns'(n\frac{p+\gamma f}{w^*}) + \mu nwu'(I)) \text{ , and}$$
$$v'(1) < -u'(w^* - p - \gamma f)w^* - \mu ns'(n) + \mu nwu'(I)$$

These conditions will be satisfied if we let $u'(0) \to +\infty$, $s'(n\frac{p+\gamma f}{w^*})$ , $v'(\frac{p+\gamma f}{w^*})$ and $u'(I)$ be bounded, and if $v'(1) \to -\infty$, $u'(w^* - p - \gamma f)$ and $s'(n)$ be bounded. In particular, they will be satified for $u(c) = \log(c)$ and $v(t) = \theta_t \log(1 - t)$.

48

## A.2  Selfish/fully patient-oriented doctors: $\mu = 0$ or $\mu = 1$

Consider the care when doctors do not care about the utility of their patients, so when $\mu = 0$. They will set then the level of diagnostics to maximize the following function:

$$\psi(d) = \frac{1}{1 - \beta(1 - \rho(d))} \left[ u(w - \gamma f - p) - u(b - \gamma f) \right]$$

Then, $\psi'(d) > 0$, so the lifetime utility of doctors will always be increasing in the number of diagnostics prescribed, which will result in the maximum level of diagnostics, $d = d_{m}ax$, and minimum level of malpractice risk, $\rho = \rho_{min}$. In the equilibrium, doctors will be offered the wage $w^*$ satisfying the following condition:

$$u(w^* - \gamma f - \rho_{min} M) = u(w_o) + \frac{\beta \rho(d^*)}{1 - \beta} \left[ u(w_o) - u(b - \gamma f) \right]$$

Hence, all the predictions following from the condition (2) will be satisfied (wage will be increasing in the opportunity wage, malpractice fine, tuition fees and the the (minimum) level of malpractice risk).

Now assume that doctors do not care about their wages at all and they are fully patient-oriented, so that $\mu \to \infty$. In this case, if $K(d_P^*) > u(b - \gamma f)$, doctors will still prescribe more diagnostics than what would be patient's optimum, because they will prefer to make their careers longer. To see this, consider the utility function they maximize:

$$\psi(d) = \frac{1}{1 - \beta(1 - \rho(d))} \left[ K(d) - u(b - \gamma f) \right]$$

The first derivative of this function is given by:

$$\psi'(d) = \frac{\beta(-\rho'(d))}{[1 - \beta(1 - \rho(d))]^2} \left[ K(d) - u(b - \gamma f) \right] + \frac{1}{1 - \beta(1 - \rho(d))} K'(d)$$

We know that that in the patient's optimum $K'(d_P^*) = 0$, hence,

$$\psi'(d_P^*) = \frac{\beta(-\rho'(d))}{[1 - \beta(1 - \rho(d_P^*))]^2} \left[ K(d_P^*) - u(b - \gamma f) \right] \overset{K(d_P^*) > u(b - \gamma f)}{>} 0$$

Hence, at point $d = d_P^*$, if $K(d_P^*) > u(b - \gamma f)$, the doctor will benefit from increasing the number of diagnostics, hence, he will choose the level of diagnostics that will be higher than $d_P^*$. The level of diagnostics that is chosen by the doctor will satisfy (13) with $\mu = 0$ and it will increase in tuition fees (according to (??)).
Only if $K(d_P^*) = u(b - \gamma f)$ the doctor's optimum will be equal to the patient's optimum. If $K(d_P^*) < u(b - \gamma f)$ then the doctors would get higher utility from not practicing then from practicing, because $K(d_P^*)$ represents now one period utility of practicing, so we can rule this case out.
If $\mu \to \infty$, then all the utility that doctors get comes from patient's utility, so they will not care about wages at all, hence in the equilibrium $w^* = 0$ as long as

$$K(d_P^*) \geq u(w_o) + \frac{\beta \rho(d^*)}{1 - \beta} \left[ u(w_o) - u(b - \gamma f) \right]$$

49

## A.3 Comparison with optimal working time as a doctor

The nice feature of this model is that the comparative statics of $t^*$ are not affected by the amount of time worked in alternative occupations. We obtain them by total differentiation of the equation $u'(\bar{w}^*)w + v'(t^*) + \mu s'(t^*) = 0$. In this equation, the utility of other states does not show up.

However, the comparative statics of $d^*$ might be affected by adjustments in the working time in other states (in particular, in the state after committing malpractice).

The doctors' optimal working time satisfies condition 14:

$$u'(wt^* - p - \gamma f)w + v'(t^*) + \mu s'(t^*) = 0 \iff \frac{1}{t^* - \frac{p - \gamma f}{w}} + v'(t^*) + \mu n s'(nt^*) - \mu n z w u'(I) = 0$$

Define $g(t) = \frac{1}{t - \frac{p + \gamma f}{w}} + v'(t) + \mu n s'(nt) - \mu n z w u'(I)$. Then:

$$g(t_0^*) = \frac{1}{t_0^* - \frac{p + \gamma f}{w}} + v'(t_0^*) + \mu n s'(nt_0^*) - \mu n z w u'(I) = \frac{1}{t_0^* - \frac{p + \gamma f}{w}} - \frac{1}{t_0^*} + \mu n s'(nt_0^*) - \mu n z w u'(I) > 0$$

Therefore, $t_0^* > t^*$ ($g(t)$ is positive for $t < t^*$ and negative afterwards). Hence, doctors work less than what they would work in the alternative occupation. Similarly, after committing malpractice, doctors will choose to supply more hours than in the opportunity occupation:

$$\frac{1}{t_0^* - \frac{\gamma f}{b}} + v'(t_0^*) > 0$$

Hence, $t_0 < t^*$ and $t_0 < t_b^*$.

Doctors will work longer hours after committing malpractice than as practicing doctors, iff:

$$g(t_b^*) < 0 \iff \frac{1}{t_b^* - \frac{p + \gamma f}{w}} - \frac{1}{t_b^* - \frac{\gamma f}{b}} + \mu n s'(nt_b^*) - \mu n z w u'(I) < 0$$

# B    Comparison of models

**Baseline model - with minimum $w^*$ set to incentivise potential doctors to follow medical training**

$$u(w^* - \gamma f - p) - u(b - \gamma f) = \frac{1 - \beta(1 - \rho(d^*))}{1 - \beta} \left[ u(w_o) - u(b - \gamma f) \right]$$

**Model with choice of diagnostics**

$$u(w^* - \gamma f - p) - u(b - \gamma f) + \mu \left[ u(I - p_d d^*) + H(d^*) \right] = \frac{1 - \beta(1 - \rho(d^*))}{1 - \beta} \left[ u(w_o) - u(b - \gamma f) \right]$$

$$\max_d D(d) = \frac{1}{1 - \beta(1 - \rho(d))} \left\{ u(\bar{w}) - u(b - \gamma f) + \mu \left[ u(I - p_d d) + H(d) \right] \right\} + \frac{u(b - \gamma f)}{1 - \beta}$$

FOC:

$$\frac{\partial D}{\partial d} = 0 \iff -\beta \rho'(d^*) \left\{ D(d^*) - \frac{u(b - \gamma f)}{1 - \beta} \right\} = \mu \left[ p_d u'(I - p_d d^*) - H'(d^*) \right]$$

**Model with choice of diagnostics and time**

$$u(w^* t^* - \gamma f - p, t^*) - u(b - \gamma f, 0) + \mu \left[ u(I - p_d d^*, 0) + H(d^*, t^*) \right] = \frac{1 - \beta(1 - \rho(d^*))}{1 - \beta} \left[ u(w_o, 0) - u(b - \gamma f, 0) \right]$$

$$\max_{d,t} D(d, t) = \frac{1}{1 - \beta(1 - \rho(d))} \left\{ u(\bar{w}, t) - u(b - \gamma f) + \mu \left[ u(I - p_d d, 0) + H(d, t) \right] \right\} + \frac{u(b - \gamma f)}{1 - \beta}$$

FOC:

$$\frac{\partial D}{\partial d} = 0 \iff -\beta \rho'(d^*) \left\{ D(d^*, t^*) - \frac{u(b - \gamma f, 0)}{1 - \beta} \right\} = \mu \left[ p_d u'(I - p_d d^*, 0) - h'(d^*) \right]$$

$$\frac{\partial D}{\partial t} = 0 \iff u'(\bar{w}^*) w + v'(t^*) + \mu s'(t^*) = 0$$

# C  Model with $\rho(t)$, without diagnostics

Consider the case in which doctors can minimize their risk of malpractice by spending more time with patients. If we keep the number of doctors per patient fixed, we can then consider the case in which $\rho$ is a function of $t$ (time spent with all patients is equivalent to time spent with an average patient). To simplify, assume that there are no diagnostics in the model, so time is the only mean used to decrease the risk of malpractice. The lifetime utility of the doctor is then given by (we denote the opportunity utility by $u_0$ and assume for now it does not depend on $\gamma$, $w$ or $t$):

$$D(t) = \frac{1}{1 - \beta(1 - \rho(t))} \left[u(wt - \gamma f - p) - u(bt_b - \gamma f) - v(t_b) + \mu K(t)\right] + \frac{u(bt_b - \gamma f) + v(t_b)}{1 - \beta}$$

where $K(t) = u(I) - u'(I)nzwt + s(nt)$.

Doctors will choose $t^*$ to maximize $D(t)$. $t^*$ will satisfy the following first order condition:

$$F_t: -\beta\rho'(t^*)(D(t^*) - \frac{u(bt_b - \gamma f) + v(t_b)}{1 - \beta}) + u'(wt^* - \gamma f - p)w + v'(t^*) + \mu n(-u'(I)zw + s'(nt^*)) = 0$$

$t^*$ will then be the maximum of $D(t)$:

$$D''(t^*) = \frac{1}{1 - \beta(1 - \rho(t^*))} \left[-\beta\rho''(D - \frac{u_b}{1 - \beta}) + u''w^2 + v'' + \mu n^2 s''\right] < 0$$

In this partial equilibrium (with fixed wage $w$), $t^*$ will be increasing in tuition fees, however, the sign of the relationship between $t^*$ and hourly wage will be ambiguous:

$$\frac{dt^*}{d\eta} = \frac{\frac{\partial F_t}{\partial \eta}}{-(-\beta\rho''(t) + u''(wt - \gamma f - p)wt + v''(t) + \mu n^2 s''(nt))}$$

$$\frac{dt^*}{d\gamma} = f \frac{\frac{\beta\rho'}{1 - \beta(1 - \rho)}(u'(b) - u'(\bar{w})) - u''(\bar{w})w}{-(-\beta\rho''(t) + u''(wt - \gamma f - p)wt + v''(t) + \mu n^2 s''(nt))} > 0$$

$$\frac{dt^*}{dw} = \frac{\frac{-\beta\rho'(t)}{1 - \beta(1 - \rho(t))}u'(\bar{w})t + u''(\bar{w})wt + u'(\bar{w}) - \mu nzu'(I)}{-(-\beta\rho''(t) + u''(wt - \gamma f - p)wt + v''(t) + \mu n^2 s''(nt))} <> 0$$

In the full equilibrium, $w^*$ will be determined such that $D = \frac{u_0}{1 - \beta}$. Hence, $(t^*, w^*)$ will be given by:

$$F_t: \quad -\beta\rho'(t^*)\frac{u_0 - u(bt_b - \gamma f) - v(t_b)}{1 - \beta} + u'(w^*t^* - \gamma f - p)w^* + v'(t^*) + \mu n(-u'(I)zw^* + s'(nt^*)) = 0$$

$$F_w: \quad u(w^*t^* - \gamma f - p) + v(t) - u(bt_b - \gamma f) - v(t_b) - \mu\left[u(I) - u'(I)nzw^*t^* + s(nt^*)\right]$$
$$-\frac{1 - \beta(1 - \rho(t^*))}{1 - \beta}(u(bt_b - \gamma f) + v(t_b) - u_0) = 0$$

In the equilibrium, doctors will work more and receive higher wages per hour if tuition fees are higher (with log utility):

$$\frac{dw^*}{d\eta} = \frac{-\frac{\partial F_w}{\partial \eta}}{u'(\bar{w}^*)t^* - \mu u'(I)nzt^*}$$

$$\frac{dt^*}{d\eta} = \frac{\frac{\partial F_t}{\partial \eta} - \frac{\partial F_w}{\partial \eta}\left[\frac{u''w}{u' - \mu u'(I)nzt} + 1\right]}{-(-\beta \rho''(t) + u''(w^*t^* - \gamma f - p)wt + v''(t) + \mu n^2 s''(nt))}$$

$$\frac{dw^*}{d\gamma} = f\frac{u'(w) + \frac{\beta \rho}{1-\beta}u'(b)}{u'(\bar{w}^*)t^* - \mu u'(I)nzt^*} > 0$$

$$\frac{dt^*}{d\gamma} = f\frac{-\frac{\beta \rho'}{1-\beta(1-\rho)}u'(b) - u''(\bar{w})w^* - \left(u'(w) + \frac{\beta \rho}{1-\beta}u'(b)\right)\left[\frac{u''w}{u' - \mu u'(I)nzt} + 1\right]}{-(-\beta \rho''(t) + u''(wt - \gamma f - p)wt + v''(t) + \mu n^2 s''(nt))} \overset{u=logc}{>} 0$$

The effect of the change in fees on the utility of the patient can be approximated by:

$$dK \approx \frac{dw}{w}(-u'(I)nzwt) + \frac{dt}{t}(-u'(I)nzwt + nts'(nt))$$

$$dK > 0 \iff \frac{dt}{t}nts'(nt)) > u'(I)nzwt\left(\frac{dt}{t} + \frac{dw}{w}\right)$$

# D   Partial insurance

Now suppose doctors are partially insured from malpractice fines. This means that in case they are found guilty of malpractice, they have to pay the amount $m$ out of pocket, while $M - m$ is paid by the insurance.

If doctors are free to choose the level of diagnostics $d$, then they will choose $d^*$ to maximize

$$
\begin{aligned}
D \;=\; & \max_{d > d_P^*} \frac{1}{1 - \beta(1 - \rho(d))} \Big[ \mu \left[(1 - \rho(d))u(w - \gamma f - p) + \rho(d)u(w - \gamma f - p - m)\right] \\
& + (1 - \mu)(K(d_P^*) - C(d)) - u(b - \gamma f) \Big] + \frac{1}{1 - \beta} u(b - \gamma f)
\end{aligned}
$$

Equivalently, they will choose $d$ to maximize:

$$
\psi(d) = \frac{1}{1 - \beta(1 - \rho(d))} \Big[ \mu[(1 - \rho(d))u(\bar{w}) + \rho(d)u(\bar{w} - m)] + (1 - \mu)(K(d_P^*) - C(d)) - u(b - \gamma f) \Big]
$$

where $\bar{w} = w - \gamma f - p$. The first derivative of $\psi(d)$ is given by:

$$
\psi'(d) = \frac{1}{1 - \beta(1 - \rho(d))} \left[ \beta(-\rho'(d))\psi(d) + \mu(-\rho'(d))(u(\bar{w}) - u(\bar{w} - m)) - (1 - \mu)C'(d) \right]
$$

The resulting level of diagnostics $d^*$ will then satisfy the first order condition:

$$
\begin{aligned}
\psi'(d^*) \;&=\; 0 \iff \\
(1 - \mu)C'(d^*) \;&=\; \beta(-\rho'(d^*))\psi(d^*) + \mu(-\rho'(d^*))(u(\bar{w}) - u(\bar{w} - m)) \qquad (22)
\end{aligned}
$$

**Proposition 10** *If there are diminishing returns to prescribing additional diagnostics, the patients' costs of diagnostics are increasing and convex and doctors are partially insured from malpractice fines with a deductible $m$, then the optimal level of diagnostics maximizing doctors' expected lifetime utility is increasing in the size of the deductible, tuition fee and decreasing in the benefits that doctors obtain after committing malpractice. Moreover, if*

$$
u'(\bar{w}) > (1 - \beta)u'(\bar{w} - m)
$$

*then the optimal level of diagnostics is increasing in wages and decreasing the malpractice insurance premiums.*

Hence, higher level of deductibles will translate into doctors taking more actions to avoid malpractice. This means that full insurance creates incentives for the doctors to set the number of diagnostics closer to the social optimum $d_P^*$. Note however that this is the case of partial equilibrium.

Proof:

The level of diagnostics chosen by doctors will be increasing in the deductible $m$, tuition fees[27] and decreasing in the benefits $b$:

$$\frac{dd^*}{dm} = \frac{\mu\,(1-\beta)u'(\bar{w}-m)}{f'(d^*)(1-\mu)(1-\beta(1-\rho(d^*)))} > 0$$

$$\frac{dd^*}{d\gamma} = \frac{f\left[\beta u'(b-\gamma f) + \mu((1-\beta)u'(\bar{w}-m) - u'(\bar{w}))\right]}{f'(d^*)(1-\mu)(1-\beta(1-\rho(d^*)))} > 0$$

$$\frac{dd^*}{db} = \frac{-\beta u'(b-\gamma f)}{f'(d^*)(1-\mu)(1-\beta(1-\rho(d^*)))} < 0$$

The level of diagnostics might be either increasing or decreasing in wages and premiums:

$$\frac{dd^*}{dw} = \frac{\mu}{f'(d^*)(1-\mu)(1-\beta(1-\rho(d^*)))} \cdot \left[u'(\bar{w}) - (1-\beta)u'(\bar{w}-m)\right]$$

$$\frac{dd^*}{dp} = \frac{\mu}{f'(d^*)(1-\mu)(1-\beta(1-\rho(d^*)))} \cdot \left[-u'(\bar{w}) + (1-\beta)u'(\bar{w}-m)\right]$$

For small values $m$ (relative to net wages $\bar{w}$) or high values of $\beta$, for which the condition $u'(\bar{w}) > (1-\beta)u'(\bar{w}-m)$ is satisfied, doctors will be willing to prescribe more diagnostics, the higher their wages are and the lower their premiums are. However, for high values of $m$ or small values of $\beta$, the effect of wages and premium on the level of diagnostics prescribed will be opposite: doctors will be willing to prescribe less diagnostics, the higher the wages and the lower the premiums.

For example, if utility is given by CARA: $u(c) = -e^{-rc}$, then the threshold for $m$ will not depend on the level of wages, but only on the other parameters describing individual risk and time preferences, $r$ and $\beta$:

$$u'(\bar{w}) - (1-\beta)u'(\bar{w}-m) = re^{-rw}\left[1 - (1-\beta)e^{rm}\right] > 0 \iff m < -\frac{1}{r}\log(1-\beta)$$

## D.1  Equilibrium

In the equilibrium with partial malpractice insurance, the condition giving the minimum wage and the insurance premium are given by:

$$\mu\left[(1-\rho)u(w^* - \gamma f - p) + \rho u(w^* - \gamma f - p - m)\right] + (1-\mu)(K(d_P^*) - C(d^*))$$

$$= u(w_o) + \frac{\beta\rho}{1-\beta}\left[u(w_o) - u(b-\gamma f)\right] \tag{23}$$

$$p = (1+\alpha)\Psi\rho(M-m) \tag{24}$$

---

[27]The level of diagnostics will be increasing in the tuition fees because if $\bar{w} > b - f$, then:

$$\frac{\beta}{\mu}u'(b-f\gamma) + (1-\beta)u'(\bar{w}-m) > \beta u'(b-f\gamma) + (1-\beta)u'(\bar{w}-m) > \min(u'(b-f\gamma), u'(\bar{w}-m)) > u'(\bar{w})$$

Then, conditions (22), (23) and (24) yield equilibrium values of $(d^*, w^*)$ in an equilibrium where doctors have no bargaining power and are only partly insured from malpractice fines. Denote $\bar{w}^* = w^* - \gamma f - (1+\alpha)\rho(d^*)(M-m)$ and $\bar{b} = b - \gamma f$. Then, the equilibrium will be given by:

$$
\begin{cases}
\mu\left[(1-\rho)u(\bar{w}^*) + \rho u(\bar{w}^* - m)\right] + (1-\mu)(K(d_P^*) - C(d^*)) & = u(w_o) + \frac{\beta\rho}{1-\beta}\left[u(w_o) - u(\bar{b})\right] \\
\mu\left[(1-\rho)u(\bar{w}^*) + \rho u(\bar{w}^* - m)\right] + (1-\mu)(K(d_P^*) - C(d^*)) & = u(b - \gamma f) + \frac{1-\beta(1-\rho)}{\beta(-\rho')}\Big[(1-\mu)C'(d^*) \\
& \qquad - \mu(-\rho')[u(\bar{w}^*) - u(\bar{w}^* - m)]\Big]
\end{cases}
$$

Combining both conditions and rearranging, we obtain:

$$
\begin{cases}
\mu[u(\bar{w}^*) - u(\bar{w}^* - m)] = & \frac{1-\mu}{-\rho'(d^*)}C'(d^*) - \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right] \\
\mu u(\bar{w}^*) = & u(w_o) + \frac{(1-\mu)\rho(d^*)}{-\rho'(d^*)}C'(d^*) - (1-\mu)(K(d_P^*) - C(d^*))
\end{cases} \tag{25}
$$

First, we show that under two (sufficient) conditions for values of $K(d_P^*)$ and $f(d_{max}) = \frac{C'(d_{max})}{-\rho'(d_{max})}$, there will exist a unique equilibrium (see Appendix D.2).[28]

**Proposition 11** *If doctors can only partial insurance, then if [ $K(d_P^*)$ and $f(d_{max}) = \frac{C'(d_{max})}{-\rho'(d_{max})}$...] there exists a unique equilibrium value for the number of diagnostics prescribed and wages.*

Second, by total differentiation we can analyze how $\gamma, b, w_o, \mu$ and $m$ affect the values of $d^*, w^*$ in the equilibrium.

**Proposition 12** *In the equilibrium with partial insurance, the number of diagnostics is increasing in deductibles, tuition fees, the degree to which doctors care about their financial benefits relative to the utility of the patient, and decreasing in the benefits that doctors obtain after committing malpractice. The relationship between the number of diagnostics and the opportunity wage $w_o$ is ambiguous. The number of diagnostics in the equilibrium will be increasing in the opportunity wages if*

$$
u'(w^* - \gamma f - (1+\alpha)\rho(d^*)(M-m)) > (1-\beta)u'(w^* - \gamma f - (1+\alpha)\rho(d^*)(M-m) - m)
$$

**Proposition 13** *In the equilibrium with partial insurance, if malpractice fines are not too high, then the minimum wages are increasing in deductibles, tuition fees and decreasing in the benefits that doctors obtain after committing malpractice. Wages in the equilibrium will be increasing in opportunity wages if malpractice fines are not too high and if the following condition is satisfied:*

$$
u'(w^* - \gamma f - (1+\alpha)\rho(d^*)(M-m)) > (1-\beta)u'(w^* - \gamma f - (1+\alpha)\rho(d^*)(M-m) - m)
$$

---

[28]CHECK $(1+\alpha)$ in all the formulas below!!!!!

Proof diagnostics:

$$\frac{dd^*}{dm} = \frac{\mu\, u'(\bar{w}-m)}{f'(d^*)(1-\mu)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]} > 0$$

$$\frac{dd^*}{d\gamma} = \frac{\beta f\, u'(b-\gamma f)}{f'(d^*)(1-\mu)(1-\beta)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]} > 0$$

$$\frac{dd^*}{dw_o} = \frac{u'(w_o)}{f'(d^*)(1-\mu)(1-\beta)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]} \left[u'(\bar{w})-(1-\beta)u'(\bar{w}-m)\right]$$

$$\frac{dd^*}{db} = \frac{-\beta u'(b-\gamma f)}{f'(d^*)(1-\mu)(1-\beta)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]} < 0$$

$$\frac{dd^*}{d\mu} = \frac{\frac{1}{1-\mu}(u(\bar{w})-u(\bar{w}-m))+\frac{\frac{u'(\bar{w}-m)}{u'(\bar{w})}-1}{1-\mu}(u(\bar{w})-u(w_o))+\frac{\beta}{1-\beta}(u(w_o)-u(b))}{f'(d^*)(1-\mu)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]} > 0$$

Proof wages: [29]

$$\frac{dw^*}{dm} = \frac{\mu\, u'(\bar{w}-m)}{\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]}\left[\frac{\rho(d^*)}{\mu u'(\bar{w})}-\frac{-\rho'(d^*)}{(1-\mu)f'(d^*)}(M-m)\right]$$

$$> 0 \iff M < \bar{M}_m = m + \frac{f'(d^*)}{-\rho'(d^*)}\frac{1-\mu}{\mu}\frac{\rho(d^*)}{u'(\bar{w})}$$

$$\frac{dw^*}{d\gamma} = f + \frac{\beta f u'(\bar{b})}{(1-\beta)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]}\left[\frac{\rho(d^*)}{\mu u'(\bar{w})}-\frac{-\rho'(d^*)}{(1-\mu)f'(d^*)}(M-m)\right]$$

$$> 0 \iff M < \bar{M}_\gamma = m + \frac{f'(d^*)}{-\rho'(d^*)}\frac{1-\mu}{\mu}\left[\frac{\mu(1-\beta)\left[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}\right]}{\beta u'(\bar{b})}+\frac{\rho(d^*)}{u'(\bar{w})}\right]$$

$$\frac{dw^*}{dw_o} = \frac{u'(w_o)}{\mu u'(\bar{w})}+\frac{u'(w_o)[u'(\bar{w})-(1-\beta)u'(\bar{w}-m)]}{(1-\beta)u'(\bar{w})[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}]}\left[\frac{\rho(d^*)}{\mu u'(\bar{w})}-\frac{-\rho'(d^*)}{(1-\mu)f'(d^*)}(M-m)\right]$$

$$> 0 \iff M < \bar{M}_{w_o} = m + \frac{f'(d^*)}{-\rho'(d^*)}\frac{1-\mu}{\mu}\left[\frac{(1-\beta)[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}]}{[u'(\bar{w})-(1-\beta)u'(\bar{w}-m)]}+\frac{\rho(d^*)}{u'(\bar{w})}\right]$$

$$\frac{dw^*}{db} = -\frac{\beta}{1-\beta}\frac{u'(\bar{b})}{[1-\rho+\rho\frac{u'(\bar{w})}{u'(\bar{w}-m)}]}\left[\frac{\rho(d^*)}{\mu u'(\bar{w})}-\frac{-\rho'(d^*)}{(1-\mu)f'(d^*)}(M-m)\right]$$

$$< 0 \iff M < \bar{M}_b = \bar{M}_m = m + \frac{f'(d^*)}{-\rho'(d^*)}\frac{1-\mu}{\mu}\frac{\rho(d^*)}{u'(\bar{w})}$$

It is not clear if the condition $u'(\bar{w}) > (1-\beta)u'(\bar{w}-m)$ will hold in the equilibrium. If

---

[29]I am still calculating dw/dmu

preferences are given by CARA utility function, then the condition will be satisfied iff

$$m < -\frac{1}{r}\log(1-\beta)$$

where $r$ denotes the average risk aversion parameter.

## D.2 Proof of existence of equilibrium with partial insurance

We show that for each $w^*$ there will exist a unique $d^*$ satisfying the first equation of (25). Moreover, by plugging in the solution for $d$ in terms of $w^*$ from the first equation into the second one, we show that there exists a unique $w^*$ satisfying the second equation.
Denote $t(\bar{w}^*) = u(\bar{w}^*) - u(\bar{w}^* - m)$. $t$ is decreasing in $\bar{w}^*$ and its values run between $u(m) - u(0)$ and 0.
Denote as before $f(d) = \frac{C'(d)}{-\rho'(d)}$. Recall that $f'(d) > 0$. Moreover, by definition of $C$, at social optimum $d_P^*$, $C(d_P^*) = 0$ and therefore $f(d_P^*) = 0$.
First, let us fix $w^*$. For simplicity of notation, let us call $t = t(w^*)$. After plugging in $t$ and $f$, the first equation of (25) reads:

$$
\begin{aligned}
\mu t &= (1-\mu)f(d^*) - \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right] \iff \\
f(d^*) &= \frac{1}{1-\mu}\left(\mu t + \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right]\right)
\end{aligned}
\tag{26}
$$

Assume

$$f(d_{max}) > \sup_t \frac{1}{1-\mu}\left(\mu t + \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right]\right) = \frac{1}{1-\mu}\left(\mu(u(m) - u(0)) + \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right]\right)$$

Then, $f$ is continuous, increasing and:

$$
\begin{aligned}
f(d_P^*) &< \inf_{t>0}\frac{1}{1-\mu}\left(\mu t + \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right]\right) \\
f(d_{max}) &> \frac{1}{1-\mu}\left(\mu t + \frac{\beta}{1-\beta}\left[u(w_o) - u(\bar{b})\right]\right)
\end{aligned}
$$

Therefore, for every $t$ there exists a unique solution $d^* = d(t)$ satisfying condition (26). Moreover, it is straightforward to show that $d'(t) > 0$.
The second equation of (25) can be rewritten as:

$$\mu u(\bar{w}^*) - g(d^*) = u(w_o) - (1-\mu)K(d_P^*)$$

where:

$$g(d) = \frac{(1-\mu)\rho(d^*)}{-\rho'(d^*)}C'(d^*) + (1-\mu)C(d^*)$$

and:
$$g'(d) = \frac{\rho(d^*)}{-\rho'(d^*)} \left\{ C''(d^*) + \frac{C'(d^*)\rho''(d^*)}{-\rho'(d^*)} \right\} > 0$$

Consider $d^* = d(t) = d(u(\bar{w}^*) - u(\bar{w}^* - m))$. Then the optimal $w^*$ will satisfy:

$$\mu u(\bar{w}^*) - g(d(u(\bar{w}^*) - u(\bar{w}^* - m))) = u(w_o) - (1 - \mu)K(d_P^*)$$

Denote $z(\bar{w}^*) = \mu u(\bar{w}^*) - g(d(u(\bar{w}^*) - u(\bar{w}^* - m)))$. We show that this function is increasing in $\bar{w}^*$:

$$z'(\bar{w}^*) = \mu u'(\bar{w}^*) - g'(d(u(\bar{w}^*) - u(\bar{w}^* - m)))d'(u(\bar{w}^*) - u(\bar{w}^* - m))(u'(\bar{w}^*) - u'(\bar{w}^* - m))$$

Note that $\mu u' > 0$, $g' > 0$, $d' > 0$ and $u'(\bar{w}^*) - u'(\bar{w}^* - m) < 0$. Hence $z'(\bar{w}^*) > 0$.
To complete the proof, it suffices to show that the function $z$ can take values smaller or higher than $u(w_o) - (1 - \mu)K(d_P^*)$. First, note that if $\bar{w} \to \infty$, then $z(\bar{w}) \to \infty$, so for high values of $\bar{w}$, $z$ will take higher values than $u(w_o) - (1 - \mu)K(d_P^*)$. Moreover:

$$z(m) = \mu u(m) - g(d(u(m) - u(0)))$$

To show that $z(m) > u(w_o) - (1 - \mu)K(d_P^*)$, we plug into the expression for $z$ the expression for $f$ from 26 and show equivalently that

$$\mu \left[ (1 - \rho)u(m) + \rho u(0) \right] < u(w_o) + \frac{\beta \rho(d^*)}{1 - \beta} \left[ u(w_o) - u(\bar{b}) \right] - (1 - \mu)(K(d_P^*) - C(d^*))$$

where $d^* = d(u(m) - u(0))$. This condition will not always be satisfied and it leads to an assumption on $K(d_P^*)$ being not too high:[30]

$$K(d_P^*) < u(w_o) - \mu \left[ (1 - \rho)u(m) + \rho u(0) \right] + \frac{\beta \rho(d(u(m) - u(0)))}{1 - \beta} \left[ u(w_o) - u(\bar{b}) \right] + (1 - \mu)C(d(u(m) - u(0)))$$

Hence, if doctors can only partly insure, there will exist a unique equilibrium if the following two assumptions (sufficient conditions) are satisfied:

$$\frac{C'(d_{max})}{-\rho'(d_{max})} > \frac{1}{1 - \mu} \left( \mu(u(m) - u(0)) + \frac{\beta}{1 - \beta} \left[ u(w_o) - u(\bar{b}) \right] \right)$$

$$K(d_P^*) < u(w_o) - \mu \left[ (1 - \rho)u(m) + \rho u(0) \right] + \frac{\beta \rho(d(u(m) - u(0)))}{1 - \beta} \left[ u(w_o) - u(\bar{b}) \right] +$$
$$+ (1 - \mu)C(d(u(m) - u(0)))$$

---

[30]Note that the solution for $d$ does not depend on $K$

# E  Notes

## E.1  Mike's notes on doctor utility function

Utility is additively separable in consumption and health. A doctor gets utility from his own consumption $C = w - \gamma c - p$ and from his health $H$. Assume the doctors health is the same regardless of whether he works as a doctor or in an alternative occupation, so $H$ drops out of the decision rule. The decision rule for becoming a doctor is given by the following equation, which will hold with equality in equilibrium:

$$u(w - \gamma c - p) + \mu \left\{ u(I - p_d d) + H(d) \right\} > u(w_o) + \frac{\beta \rho(d)}{1 - \beta} \left[ u(w_o) - u(b - \gamma c) \right]$$

Here w is the doctor wage and $w_o$ is the outside option wage. $\rho(d)$ is the probability of malpractice given diagnostic level $d$, and $b$ is the wage conditional on malpractice. Doctors get utility from the utility of their patients, which is also additively separable in consumption and health, and is given by:

$$U(d) = u(I - p_d d) + H(d)$$

Assume that $\mu > 0$ so that doctors do put some weight on the utility of their patients. Then a higher $U(d)$ will lower the reservation wage of doctors. The non-pecuniary reward that doctors receive is:

$$N^d(d) = \mu \left\{ u(I - p_d d) + H(d) \right\}$$

Non-doctors also get some non-monetary rewards from work, which we can denote by N. In discrete choice models it is standard to normalize non-pecuniary rewards for one occupation to zero, and measure rewards for other occupations relative to that. So suppose we set $N = 0$. It is important to note that by doing this we are implicitly subtracting $N$ from $N^d(d)$. Then, as an example of a particular function form, we might write:

$$N^d(d) = \mu \left\{ \beta + \log(I - p_d d) + a \log(1 + d) \right\}$$

Here the intercept $\beta$ has the interpretation of $N$. So we would expect $\beta$ to be negative. We would expect $N^d(d)$ to be positive because doctors presumably get some non-pecuniary rewards from being doctors, but the sign of $N^d(d)$ is ambiguous. Note that we have three wages w, w0 and b. I suggest we calibrate w0 to equal the value for some comparable group like professionals or college graduates. To calibrate b we need some information on what doctors do after malpractice issues. I suggest we do not calibrate the doctor wage w directly, but rather use it as a target we try to match by manipulating the model parameters (such as $\mu, \beta$ and $a$).

## E.2  Strengths and limitations of the model

This model can address the following issues:

- risk aversion of individuals - general utility function

- differences in subsidies between different countries ($\gamma$)

- differences in the total cost of education, which could take into account differences in the length of education ($C$)

- differences in the risk and consequences of malpractice ($\rho, M$)

- prices of malpractice insurance (replace $p = \rho M$ by e.g. $(1 + \alpha)\rho M$)

- differences in the length of a career - $n$ (are doctors allowed to practice as long as they want?)

- differences in the size of opportunity wage ($w_o$)

The model allows to calculate the effects of different institutional settings on the wage of doctors. In particular, it is easy to calculate the effect of the change in the subsidy rate $\gamma$ on wages (note that the wage function is linear in $\gamma$).

The model does not address:

- differences in the access to funding (initial endowments) for the costs of the tuition fee - we assume all individuals finance tuition fee by bank loans

- wage bargaining - for now we assumed that doctors receive minimum wage that will incentivize them to follow medical education

- wage patterns over time - including a pattern for wages over time complicates the model a lot, because wages show up inside the utility function, so making parametric assumptions on the pattern of wages will not help. To solve the model, we could make assumptions on the utility of wages, which would however not be very plausible and could also put restrictions on other elements of the utility function, such as $\rho M$ and $\gamma c$. Working with the implicit solution of the model is not really informative (total differentiation does not tell much). You can solve the model under the linear utility though, with an assumption $w_t = w(1 - a^t)$.

### E.2.1  General equilibrium issues

The share of health spending in GDP presented here is for now true only in the first period. With malpractice, some of the doctors quit the profession, which gives rise to the following issues:

- If we want to have constant number of doctors in the economy $L^*$, we need to allow inflow of $\rho L$ new doctors per period.

- If we want to allow some of the doctors to retire, the inflow should be $(\rho + \lambda_r)L$ where $\lambda_r$ is the retirement rate.

- with retirement, we also need to allow for the inflow to other sector(s), at the retirement rate.

- this raises a question, if all new individuals become doctors. If we do not allow for retirement, we would have to put such a restriction on the model.

Another issue is where the benefits $b$ come from. Rather not taxes, because these are the benefits that an individual receives after committing malpractice. Do we want to allow these individuals to produce something? Will they show up in the GDP?

## E.3  Barry's comments - empirical issues

Empirical issues we should take into account

- accounting issue - are subsidies included in the measure of health spending?

- nurses are also subsidized. their wages are lower than doctors', but the number of wages is much higher than the number of doctors

- wages can depend on the lenght of the career and the length of the education - additional years in college can play a role

- wages of doctors in the UK are subject to bargaining, unlike in the US

Other

- McCormick and White 2013 use PSSRU data on annuized costs of training: interest rate 3.5%, length of career - they weight each period using probability of survival (see section 6.2)

# F  Health spending as a share of GDP

## F.1  Components of health spending

Health spending has three main components:

1. Qualified labor costs

   Qualified labor costs are costs of health care professionals: doctors and nurses. Consider the case when there are $L_D$ doctors and $L_N$ nurses in the economy, and they earn wages $w_D, w_N$ respectively. Then the labor costs of health care professionals are given by:
   $$w_D L_D + w_N L_N$$

   Note that the size of labor costs depend only on the number of doctors and nurses employed, but not on the number of patients treated.

2. Non-labor costs of treatments - in particular, costs of diagnostics

   Consider the costs of diagnostics per treatment. The level of diagnostics is given by $d$ and the patient's optimum is $d_P^*$. Assume the price per diagnostic is constant and equal to $p_d$.[31] The total costs in the economy will depend on the number of patients treated $n$:
   $$n p_d d = n p_d (d - d_P^*) + n p_d d_P^*$$

   We assume that $p_d$ does not include any labor costs of trained medical professionals.

3. Administrative and fixed costs

   Other costs can include labor and non-labor administrative costs and fixed costs. We assume that the labor component of the admin costs is supplied by the unqualified workers. Moreover, it includes costs of building, infrastructure etc. We denote this costs as $A$ and we assume that the size of the admin cost does not depend on the subsidies to education of health professionals (but it could e.g. depend on the number of patients).

4. Costs of subsidies to medical training

   Depending on the measure of health spending we use, it will or will not include spending on the subsidies to medical training[32]. The total costs of subsidies to doctors' education equals:
   $$(1 - \beta)(1 - \gamma)C L_D$$

---

[31] We could also model it in terms of costs to the patient $C(d)$ but I am not sure if the units should be the same.. (costs of the insurer do not have to be equal to patients costs)

[32] OECD measure dos not include subsidies, national measures do include (check)

This value represents the flow value of the part of the total cost of training $C$ that was subsidized: $(1 - \gamma)C$, multiplied by the number of doctors. [33]

The total health spending is then given by:

$$HS = w_D L_D + w_N L_N + n p_d d + A$$

or, if subsidies are counted as part of the health spending, by:

$$HS = w_D L_D + w_N L_N + n p_d d + A \; (+ \text{ subsidies})$$

Some additional plausible assumptions:

1. Changing subsidies to doctors' training will only affect doctors' wages $w_D$, while other prices in the model $w_N$, $w_U$ and $p_d$ will be unchanged.

2. $n$ - the number of patients is fixed and does not depend on the funding of the professional training. The literature suggests that Americans, if anything, visit doctors less often than other nations. [34]

3. $L_D$ and $L_N$ are determined by the isoelastic demand function. To keep things simple, we can assume no substitution:

$$
\begin{aligned}
L_D &= D_D(w_D) = k w_D^{-\varepsilon} \\
L_N &= D_N(w_N) = k_N w_N^{-\varepsilon_N}
\end{aligned}
$$

It can make sense to assume some substitution - if in a situation when doctors become relatively more expensive, hospitals will substitute some of their work with nurses. On the other hand, there is probably a set of tasks that only doctors can do. If this set is relatively big, we could claim there is no substitution between doctors and nurses.

These assumptions will allow us to keep a few costs components fixed, independent of subsidies to doctors' training: labor costs of nurses and administrative costs (labor and non-labor). Moreover, because the optimal level of diagnostics does not depend on the subsidies, also $n p_d d_P^*$ will be unaffected. Then, the effect on reducing subsidies will only work through changes in $w_D$, followed by possible changed in $L_D$ (through a change in demand), and changes in the number of diagnostics prescribed $d$, above the patient's optimum .

$$HS = w_D L_D + n p_d d \; (+ \text{ subsidies }) + \text{ constant terms}$$

With isoelastic demand:

$$
\begin{aligned}
HS &= w_D L_D + w_N L_N + n p_d d + A \; (+ \text{ subsidies}) \iff \\
HS &= k w_D^{1-\varepsilon} + k_N w_N^{1-\varepsilon_N} + n p_d d \; (+ \text{ subsidies }) + \text{ constant terms}
\end{aligned}
$$

---

[33] Note that if banks charge additional interest rates on private loans then $f < (1 - \beta)C$, so the size of subsidy is lower than what doctors would have to pay without subsidies. $(1 - \gamma)C < (1 - \gamma)\frac{f}{1-\beta}$.

[34] check once again

# References

Anderson, G. F., Frogner, B. K., and Reinhardt, U. E. (2007). Health spending in OECD countries in 2004: an update. *Health Affairs*, 26(5):1481–1489. 1

Anderson, G. F., Reinhardt, U. E., Hussey, P. S., and Petrosyan, V. (2003). It's the prices, stupid: why the United States is so different from other countries. *Health Affairs*, 22(3):89–105. 1, 68

Baicker, K., Fisher, E. S., and Chandra, A. (2007). Malpractice liability costs and the practice of medicine in the Medicare program. *Health Affairs*, 26(3):841–852. 68

Berwick, D. M. and Hackbarth, A. D. (2012). Eliminating waste in US health care. *Jama*, 307(14):1513–1516. 1

Currie, J. and MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *The Quarterly Journal of Economics*, 123(2):795–830. 68

Curtis, L. A. (2013). Unit costs of health and social care 2013. Technical report, Personal Social Services Research Unit. 68

Cutler, D. M. and Ly, D. P. (2011). The (paper) work of medicine: understanding international medical costs. *The Journal of Economic Perspectives*, pages 3–25. 2, 68

Danzon, P. M. (2000). Liability for medical malpractice. *Handbook of health economics*, 1:1339–1404. 68

Danzon, P. M., Pauly, M. V., and Kington, R. S. (1990). The effects of malpractice litigation on physicians' fees and incomes. *The American economic review*, pages 122–127. 68

Farrell, D., Jensen, E., Kocher, B., Lovegrove, N., Melhem, F., Mendonca, L., and Parish, B. (2008). Accounting for the cost of US health care: A new look at why Americans spend more. Technical report, McKinsey & Company, Special Sector Office. 1

Garber, A. M. and Skinner, J. (2008). Is American health care uniquely inefficient? *Journal of Economic Perspectives*, 22(4):27–50. 68

Health Care Cost Institute (2014). 2013 health care cost and utilization report. 68

Jackson Healthcare (2010). Quantifying the cost of defensive medicine: Summary of findings. Technical report, Jackson Healthcare. 68

Kessler, D. and McClellan, M. (1996). Do doctors practice defensive medicine? *The Quarterly Journal of Economics*, pages 353–390. 68

Kessler, D. P. and McClellan, M. B. (2002). How liability law affects medical productivity. *Journal of health economics*, 21(6):931–955. 68

Laugesen, M. J. and Glied, S. A. (2011). Higher fees paid to us physicians drive higher spending for physician services compared to other countries. *Health Affairs*, 30(9):1647–1656. 1, 68

Lewis, J. and Cooper., J. (2015). Expenditure on healthcare in the UK 2013. Technical report, Ofiice for National Statistics. 68

McCormick, B. and Stevens, M. (2014). Financing the training of health professionals: a welfare analysis. unpublished.

McCormick, B. and White, J. (2013). Economic aspects of NHS education and training subsidies and employer levies. Technical report, Center for Health Service Economics and Organisation.

Mello, M. M., Chandra, A., Gawande, A. A., and Studdert, D. M. (2010). National costs of the medical liability system. *Health Affairs*, 29(9):1569–1577. 2, 68

OECD (2013). *Health At a Glance 2013: OECD Indicators*. OECD. 68

Pozen, A. and Cutler, D. M. (2010). Medical spending differences in the United States and Canada: the role of prices, procedures, and administrative expenses. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 47(2):124–134. 68

Quinn, R. (1998). Medical malpractice insurance: the reputation effect and defensive medicine. *Journal of Risk and Insurance*, pages 467–484. 68

Sethi, M. K., Obremskey, W. T., Natividad, H., Mir, H. R., and Jahangir, A. A. (2012). Incidence and costs of defensive medicine among orthopedic surgeons in the United States: a national survey study. *Sports medicine*, 421:35–5. 68

Sloan, F. A. and Shadle, J. H. (2009). Is there empirical evidence for defensive medicine? a reassessment. *Journal of health economics*, 28(2):481–491. 68

Smith, S., Newhouse, J. P., and Freeland, M. S. (2009). Income, insurance, and technology: why does health spending outpace economic growth? *Health Affairs*, 28(5):1276–1284. 1

Studdert, D. M., Mello, M. M., Sage, W. M., DesRoches, C. M., Peugh, J., Zapert, K., and Brennan, T. A. (2005). Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA: The Journal of the American Medical Association*, 293(21):2609–2617. 68

Thomas, J. W., Ziller, E. C., and Thayer, D. A. (2010). Low costs of defensive medicine, small savings from tort reform. *Health Affairs*, 29(9):1578–1584. 68

US Department of Health and Human Services (2013). CMS statistics. *Washington, DC: Centers for Medicare and Medicaid Services (CMS Pub. No. 03504)*. 68

Waxman, D. A., Greenberg, M. D., Ridgely, M. S., Kellermann, A. L., and Heaton, P. (2014). The effect of malpractice reform on emergency department care. *New England Journal of Medicine*, 371(16):1518–1525. 68

Woolhandler, S., Campbell, T., and Himmelstein, D. U. (2003). Costs of health care administration in the United States and Canada. *New England Journal of Medicine*, 349(8):768–775. 2

# G   References explained

Differences in health spending: Anderson et al. (2003), Cutler and Ly (2011), Garber and Skinner (2008), Laugesen and Glied (2011), Lewis and Cooper. (2015), OECD (2013), Pozen and Cutler (2010).

Statistics: Curtis (2013), Lewis and Cooper. (2015), OECD (2013), US Department of Health and Human Services (2013), Health Care Cost Institute (2014).

Malpractice and defensive medicine: Baicker et al. (2007), Currie and MacLeod (2008), Danzon (2000), Danzon et al. (1990), Jackson Healthcare (2010), Kessler and McClellan (1996), Kessler and McClellan (2002), Mello et al. (2010), Quinn (1998), Sethi et al. (2012), Sloan and Shadle (2009), Studdert et al. (2005), Thomas et al. (2010), Waxman et al. (2014).