Journal of Health Economics 32 (2013) 881–894

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Do cost-sharing and entry deregulation curb pharmaceutical innovation?

Volker Grossmann[a,b,c,*]

[a] University of Fribourg, Switzerland
[b] CESifo, Munich, Germany
[c] Institute for the Study of Labor (IZA), Bonn, Germany

ARTICLE INFO

Article history:
Received 7 June 2011
Received in revised form 2 June 2013
Accepted 3 June 2013
Available online 18 June 2013

JEL classification:
I10
L10
O30

Keywords:
Cost-sharing
Entry deregulation
Health insurance
Patent breadth
Pharmaceutical innovation

ABSTRACT

This paper examines the role of both cost-sharing schemes in health insurance systems and the regulation of entry into the pharmaceutical sector for pharmaceutical R&D expenditure and drug prices. The analysis suggests that both an increase in the coinsurance rate and stricter price regulations adversely affect R&D spending in the pharmaceutical sector. In contrast, entry deregulation may lead to higher R&D spending of pharmaceutical companies. The relationship between R&D spending per firm and the number of firms may be hump-shaped. In this case, the number of rivals which maximizes R&D expenditure per firm is decreasing in the coinsurance rate and increasing in labor productivity.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Dramatically rising health expenditure costs in the last decades, in particular for prescription pharmaceuticals, have triggered ongoing debates about cost-sharing between health insurers and beneficiaries.[1] For instance, in the US, a reform of Medicare (a federal program which provides health insurance for the elderly) which went into effect in 2006 (Medicare Part D) introduced coverage of prescription drug expenditure for Medicare beneficiaries. There is, however, a coinsurance rate (the fraction of expenditure on medical services paid by the insured patient) of 25 percent.[2]

It is typically argued that, compared to full coverage, cost-sharing schemes help to keep health insurance premiums in check. There is a large empirical literature on the effects of prescription drug cost-sharing on health costs and health care utilization. Empirical estimates suggest that a 10 percent increase in patients' prescription drug charge (through higher coinsurance or higher copayment) reduces prescription drug spending by 1–6 percent (see, e.g., Goldman et al., 2007; Gemmill et al., 2008).

In contrast to such short-run demand effects of prescription drugs cost-sharing, long-run supply effects on pharmaceutical innovation are underresearched. Generally, a major concern in designing health insurance systems and regulating the pharmaceutical sector is the tension between keeping prices of pharmaceuticals low and ensuring that they treat illnesses effectively. The main issue therefore is the joint impact of cost-sharing schemes and regulation measures on price-setting behavior and the incentives of pharmaceutical companies to conduct R&D. As pointed out by Berndt (2002, p.45): "The resolution of this static

* Correspondence address: University of Fribourg, Department of Economics, Bd. de Pérolles 90, G424, 1700 Fribourg, Switzerland. Tel.: +41 026 3009383.
E-mail address: volker.grossmann@unifr.ch

[1] In the EU, the average annual real growth rate of spending for pharmaceuticals was 4.7 percent (3.8 percent in Germany) between 1998 and 2008 (OECD, 2010). In the US, there was a more than fivefold increase in spending for prescription drugs between 1990 and 2008 from 40.3 to 234.1 billion USD (see "The Kaiser Family Foundation, Prescription drug trends, May 2010", available at http://kaiserfamilyfoundation.files.wordpress.com/2013/01/3057-08.pdf).

[2] The rate applies after some deductible, up to an initial coverage limit. After a "catastrophic" coverage limit is reached, the coinsurance rate drops to 5 percent. In Switzerland basically all health insurance contracts have a coinsurance rate of 20 percent for branded prescription drugs and 10 percent for generic drugs.

0167-6296/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhealeco.2013.06.001

versus dynamic efficiency conflict is likely the single most important issue facing the pharmaceutical industry".

This paper attempts to shed light on the nature of the price-quality relationship in pharmaceutical markets. It examines the role of cost-sharing in health insurance systems, price regulations and deregulation of entry for both pharmaceutical R&D and drug prices.

The proposed theoretical model builds on the "ideal variety" framework, originated by Lancaster (1979). Although the framework has never been applied to the context of pharmaceutical markets and R&D (to the best of my knowledge),[3] it captures well the notion that patients seek the ideal drug for their type of illness.[4] The horizontal location of a pharmaceutical firm is interpreted as the type of illness to which the drug that the firm produces is targeted to, represented as a point on the circumference of a circle. That is, pharmaceuticals are imperfect substitutes to each other.[5] Firms choose their horizontal location along with prices and R&D spending.

We show that introducing insurance coverage of prescription drug expenditure (like Medicare Part D) raises both drug prices and pharmaceutical R&D spending, whereas an increase in the coinsurance rate within an existing cost-sharing scheme has the opposite effect. Intuitively, a lower coinsurance rate makes demand for pharmaceuticals less price-sensitive and therefore allows firms to charge higher price-cost margins. This, in turn, boosts the return to R&D. In fact, recent empirical evidence by Blume-Kohout and Sood (2013) suggests that Medicare Part D has raised R&D spending of pharmaceutical companies for prescription drugs used by the elderly. They find that the number of drugs entering early-phase clinical testing in a given therapeutic class and given year is higher, the larger the Medicare market share after the year 2004.

By contrast, deregulation of entry may foster pharmaceutical innovation. The result suggests that the repeated claim by pharma lobbyists – that anything which raises profits in the pharmaceutical sector would be conducive to R&D – is potentially erroneous. Appropriate policy measures to foster entry include encouraging entry of foreign firms, restricting marketing practices, and reducing patent breadth with respect to the design of a pharmaceutical product. In fact, patent breadth with respect to product design has a natural representation in the proposed model, as a segment on the circumference of the circle of illnesses which includes the point targeted by a pharmaceutical firm. Patent protection means that potential rivals are prohibited to locate on this segment. Our analysis suggests that the relationship between pharmaceutical innovation and the number of firms may be hump-shaped, i.e., is positive (negative) if the intensity of competition is low (high). In this case, the R&D-maximizing number of firms decreases in the coinsurance rate and increases in the stage of development, captured by the productivity of labor.

We also examine the role of two kinds of price regulations for pharmaceuticals.[6] First, we suppose that prices are directly be set by the government, as practiced in France and Italy. We focus on the simple case where such price controls ignore R&D costs and show that stricter direct price regulation unambiguously reduces R&D expenditure. Second, we study the effects of a price cap – a limit amount of a patients' expenses for a drug which is reimbursed by an insurer. Such cost-sharing device is common in the public health insurance system of Germany and Japan. We show that a stricter price cap reduces both R&D spending on pharmaceuticals and drug prices. The results on the effects of price regulations on R&D expenditure are consistent with a large body of empirical evidence (e.g., Scherer, 1993; Vernon, 2005; Giaccotto et al., 2005).

The paper is organized as follows. Section 2 discusses the relation of our analysis to the literature. Section 3 sets up and analyzes the basic model which focusses on coinsurance policy and restricted entry, where the number of pharmaceutical firms is given. It also discusses the relationship of competition and innovation, first, by allowing for a competitive fringe which can imitate pharmaceuticals and, second, by endogenizing the number of firms which enter at some fixed costs as long profits are non-negative (ruling out imitation). Section 4 examines the effects of price regulations for pharmaceuticals. Section 5 analyzes how the "optimal" number of firms, defined as maximizing R&D expenditure per firm, depends on the coinsurance rate and the (exogenous) productivity of labor. The last section concludes.

2. Related literature

This paper is not the first one to study the relationship between health policy and innovation incentives of pharmaceutical firms. At the theoretical level, Garber et al. (2006) analyze the case of a single-product monopoly firm which sells a pharmaceutical product. The drug is assumed to have heterogeneous effects on the utility of ill consumers. It is shown that, at a coinsurance rate which ensures efficient drug utilization, profits of the monopoly firm may exceed consumer surplus. Thus, R&D incentives may be excessive. Lakdawalla and Sood (2013) analyze a similar framework and argue that a health insurance contract which sets copayment at marginal costs and where innovators are paid an ex-ante fee equal to consumer surplus may at the same time achieve two goals: it may lead to efficient drug utilization and provide efficient incentives for introducing the drug into the market. Lakdawalla and Sood (2009) argue that a public health insurance system with some price-negotiation by the government is welfare-improving, particularly when coupled with an increase in patent length.

The framework proposed in this paper is different to this literature in several respects. First, it captures both horizontal and vertical differentiation of pharmaceuticals. Second, it analyzes product market competition among pharmaceutical companies rather than a monopoly firm. While monopoly situations may exist in some pharmaceutical markets, the exclusive focus on these situations may be less appropriate to capture markets like those for cancer medication, hypertension medication, pain killers, and antibiotics. In such markets there is some substitutability within product groups and pharmaceutical companies engage in price competition. Third, and related, the main contribution of this paper is to examine the price-quality relationship in pharmaceutical markets by contrasting health insurance policy and competition policy like the patent breadth. The salient feature to analyze competition policy is to depart from the monopoly assumption.

At the empirical level, Acemoglu et al. (2006) examine whether the first Medicare program (the "Social Security Act of 1965") had an impact on pharmaceutical innovation. They find no evidence that drug spending of the elderly (aged 65–74) relative to that of the non-elderly (55–64) went up. Similarly, there was no significant

[3] The ideal variety model is sometimes used in the international trade literature (e.g. Helpman, 1981; Wong, 1995; Hummels and Lugovsky, 2009).

[4] Besides realism in this respect, the ideal variety framework also has the attractive feature that the price elasticity of demand depends on the competitive environment of firms. Notably the standard version of the alternative (and far more often applied) "love of variety" model of monopolistic competition by Dixit and Stiglitz (1977) and Ethier (1982) predicts that the price elasticity of demand for a good – and thus the price mark-up – is constant. However, the empirical support for this prediction is generally weak. Under a constant price elasticity, the health insurance system could not have any effect on prices for pharmaceuticals.

[5] Examples are pain killers, antibiotics, hypertension medication, and pharmaceutical cancer therapy.

[6] For an overview on price regulations in the market for pharmaceuticals, see Sood et al. (2009).

effect on the number of new molecular entity approvals, as drug spending was not covered by Medicare before 2006. Our theoretical analysis suggests that the 2006 Medicare reform may spur pharmaceutical innovation.

The present paper is also closely related to the literature on the relationship between entry regulation and innovation. Economides (1993) analyzes an oligopoly model which suggests that entry of new firms lowers R&D investments. Like in the present paper, firms are located on a circumference of a circle of horizontal product attributes. There are two main differences to the present study. First, his model focusses on a single-good economy, rather than distinguishing "regular" consumption goods from health and demand for pharmaceuticals. Second, in contrast to Economides (1993), we allow for the possibility that quality-improvements are more valuable to consumers when targeted closer to their ideal variety (i.e., to their illness). Consequently, the competition-R&D relationship may be positive if the patent breadth is not too narrow rather than unambiguously negative. This prediction of our model is consistent with evidence that a higher intensity of competition may increase R&D spending (e.g., Blundell et al., 1999; Aghion et al., 2009). A similar result has been shown in Schumpeterian growth theory, surveyed by Aghion and Howitt (2005, 2009), including a hump-shaped relationship between competition and innovation suggested by our analysis. However, there the possibility that competition fosters innovation rests on the feature that firms can preserve a monopoly by innovating. Firms' incentives to search for a superior technology rises when the entry threat is enlarged. In our theory, neither heterogeneity nor the prospect of acquiring a monopoly are needed to obtain the result that entry deregulation can spur innovative effort.

More specifically, this paper is related to the literature on optimal patent breadth. Klemperer (1990) proposes a model in which a larger patent breadth is captured by a larger distance between the consumers' most preferred variety and the patentholder's product. A larger patent breadth is associated with stronger price-setting power of incumbents. If the "transport costs" per unit distance are identical among consumers, patents should be narrow. Similarly, Gilbert and Shapiro (1990) suggest that, if the deadweight loss of extending the patent breadth is increasing, then a narrow patent breadth combined with an infinite length of patent is optimal. Gallini (1992) shows that when allowing for costly imitation (i.e., the possibility for rivals to enter the market at some costs, enabling them to produce the same goods with similar production technology) may lead to the opposite result, i.e., broad but short-lived patents. Wright (1999) proposes a framework which derives both of these differing results as special cases. Like in the present paper, he associates larger patent breadth with a smaller number of firms in the market, endogenously leading to a larger market share and a higher price-cost margin (thus higher profits).

In this literature, an innovation requires a fixed R&D investment and the social planer operates under the constraint that the present discounted value of a future profit flow exceeds this fixed level. By contrast, Vincenzo (1996) allows for patent races. Thus, like in the present paper, he explicitly models a R&D decision. He shows that optimal patent breadth is large if the static welfare gain from increased competition is low compared to the social costs of R&D disincentives in a patent race.[7]

The previous literature on patent breadth has emphasized the trade-off between the goal to foster R&D activity and the goal to foster competition, by hypothesizing that larger profits from increased market share implies higher R&D expenditure.[8] By contrast, in our analysis, increased competition induced by narrowing patent breadth may foster R&D. As in this case there is no trade-off between deregulation of entry and R&D, we focus on a positive analysis and investigate the R&D-maximizing patent breadth, by abstracting from patent length. We distinguish two different implications of larger patent breadth: the effects of a change in the number of firms or market share, like in Wright (1999), and the marginal production costs of potential imitators, like in Vincenzo (1996). Moreover, the previous literature did not attempt to examine how the optimal patent breadth in the market for pharmaceuticals depends on the coinsurance rate and the ability of patients to pay for pharmaceuticals (being determined by the productivity of labor).

3. The basic model

3.1. Set up

There is a unit mass of individuals, indexed by j. Individuals draw utility $U(j)$ from consumption of a homogenous (numeraire) good, $C(j)$, and their health status, $H(j)$, according to utility function

$$U(j) = u(C(j), H(j)), \tag{1}$$

with partial derivatives $u_C > 0$, $u_H > 0$, $u_{CC} < 0$, $u_{HH} \leq 0$, and $u_{CH} \geq 0$ (i.e., the marginal utility from consumption is non-decreasing in the health level).

An individual becomes ill with probability s. Illness has two consequences. First, whereas the health level without illness is normalized to unity, it drops below one when an individual becomes ill; health can be improved by consuming a pharmaceutical. Second, we allow labor supply to positively depend on health status.[9] Formally, an individual with health level H inelastically supplies $g(H)$ units of labor, with $g' \geq 0$, $g'' \leq 0$, and $g(1) = 1$; that is, labor supply is unity if an individual stays healthy. For simplicity, the wage rate per unit of labor, w, is exogenous.[10]

There are n pharmaceutical firms, indexed by i. Each firm produces one drug with identical technology in a monopolistically competitive environment. Firms cannot engage in

[7] Brekke and Straume (2009) model a patent race in a model suitable to the pharmaceutical industry. They argue that stronger patent protection may lead to higher strategic spending of an incumbent on advertising relative to R&D investment in order to deter R&D effort of a potential entrant.

[8] In dynamic models, patent breadth may also be defined on a vertical dimension, as minimum size of an innovation to which a new patent is awarded or as "leading breadth" which specifies a range of quality improvements an innovator is prohibited to commercialize without a license. Scotchmer (1991) provides an excellent discussion on patent policy if innovations are cumulative, arguing that the first innovation should not be too broadly protected in order to leave incentives for follow-up innovations. For formal accounts on this issue in an industrial organization context, see e.g. Scotchmer (1996), O'Donoghue (1997), and O'Donoghue et al. (1998). Providing a general equilibrium perspective, Li (2001) examines a quality-ladder endogenous growth model. He finds that extending patent breadth, which is defined as "the extent of quality improvement to which a product is protected from the infringement of its patent by lower-quality goods producers" (p. C166), raises profits of innovators and therefore fosters R&D. O'Donoghue and Zweimüller (2004) focus on protection against future innovators in several classes of endogenous growth models. They show that patent policy with asymmetric effects across industries can have substantial static efficiency implications. This paper analyzes horizontal (rather than vertical) patent breadth which is related to the number of rivals offering differentiated goods.

[9] Empirical support for this assumption is provided by Cai et al. (2008). They find that individuals who experience health shocks respond by incremental reductions in labor supply rather than by leaving the labor force.

[10] Grossmann (2011) endogenizes the wage rate by allowing for endogenous innovation also outside the pharmaceutical sector. In this setting, a better health status raises aggregate productivity and wages, which gives rise to a two-way interaction between health status and economic well-being. Related theoretical papers on the health-productivity relationship include Van Zona and Muysken (2001) as well as Sanso and Aísa (2006).

price-discrimination. Marginal production costs are constant and denoted by c; that is, to produce one unit of any pharmaceutical product requires c units of the numeraire. In the basic model, the number of firms, n, is exogenous (restricted entry). (See Section 3.3.3 for the case where n is endogenous.) For simplicity, suppose that profits of firms accrue to investors outside the economy.

Pharmaceuticals differ in one horizontal dimension of attributes. A variety is a pharmaceutical which is targeted to a specific type of illness. Illnesses are represented by points on the circumference of a circle with unit length. Ill individuals are characterized by their location on the circumference and are uniformly distributed on it. Firms choose to which illness their drug is targeted to (i.e., choose a location on the circumference of the circle). Different kinds of drugs are imperfectly substitutable. For instance, some pain killers that help well for some kinds of headache work less for other types but still have an effect, some work better for rheumatism than for headache, and so on. A certain kind of chemotherapy may improve the health status for various forms of cancer but particular substances may be particularly well-suited for a specific type of cancer. The same is true for illnesses caused by bacteria, which can be treated with various kinds of antibiotics. Typically, a specific kind of antibiotic kills or prevents breeding of a rather wide spectrum of bacteria but is more effective against certain types of bacteria than others. As a final example, there are several classes of medication against hypertension. Products are substitutable quite well, targeting different sources of high blood pressure and differing with respect to side effects. We capture the notions suggested by these examples by assuming that the structure of the pharmaceutical market is characterized by oligopolistic competition on prices for differentiated goods.

Price setting power arises because pharmaceutical products cannot be imitated, e.g., because of patent protection. Patent breadth has a natural representation in the model. It is defined as the sum of the lengths of the segments on the circumference of the circle of illnesses to the left and right of the location of firm i (representing the closest substitutes to product i) where rivals are not allowed to locate. Consider a symmetric situation where the distance between the location of each firm on the circumference of the circle (with unit length) is $1/n$. This is also the size of the segment on the circle of each firm ($0.5/n$ on both sides of a firm's location) which is protected by patent law. Thus, if the patent breadth is at least $1/n$, then no additional firm is allowed to enter. The restricted entry case may therefore be interpreted as a situation where no firm can enter despite positive profits because it would infringe a patent. An increase in the firm number n may thus reflect a change in the patent law which narrows the patent breadth such that more firms can enter (Wright, 1999). Apart from this interpretation, a policy instrument to raises n could be facilitating entry of foreign firms.

Pharmaceutical firms can affect the "quality" (i.e., the vertical dimension) of drugs by incurring R&D costs. Higher quality means that the health status particularly improves for the type of illness to which the drug is targeted to and possibly for related illnesses. For instance, if two drugs are supposed to mitigate headache, the one which works better is said to have higher quality. Both drugs may be chemically very different to each other although being represented by the same horizontal location. Formally, suppose that health status of an ill individual j when consuming one unit of drug i is

$$H(j) = h(\delta_i(j), Q_i), \tag{2}$$

where $\delta_i(j)$ is the shorter (arc) distance between the illness of consumer j and the horizontal location of firm i's product on the circumference of the circle of illnesses. Q_i is the quality of drug i.

We assume that function h has partial derivatives $h_\delta < 0$ (i.e., the health level is lower when the drug is less suited), $h_Q > 0$ with $\lim_{Q\to\infty} h(0, Q) \leq 1$ (recall that unity is the upper limit of the health level by definition), and $h_{QQ} \leq 0$ (i.e., the marginal gain in health from a quality-improvement is non-increasing in the quality level); moreover, suppose $h_{\delta\delta} \leq 0$ and $h_{\delta Q} < 0$. Property $h_{\delta Q} < 0$ implies a ranking of the impact of higher R&D on health improvement for different patients. For instance, consider a drug which contains antibiotics. Suppose the drug is best suited to fight (a specific form of) pneumonia but also works against some other illnesses caused by bacteria. However, also suppose the bacteria which cause pneumonia have developed some antibiotic resistance. Then $h_{\delta Q} < 0$ means that an increase in R&D spending directed to overcome antibiotic resistance of bacteria which cause pneumonia has a larger effect on health for patients with pneumonia than for patients with other bacterial infections.[11]

To supply a drug with quality Q_i, firm i has to incur R&D costs $B(Q_i)$ which are strictly convex in Q_i, $B' > 0$, $B'' > 0$. Following the "endogenous sunk cost" approach (e.g., Sutton, 1991, 1998) and "quality ladder" models of endogenous growth (e.g., Grossman and Helpman, 1991), R&D costs are not reflected in marginal production costs.[12]

Illnesses are assumed to be perfectly and costlessly detectable by diagnostic tests. Moreover, individuals know the horizontal and vertical location of each firm, as well as function h, and therefore are capable of choosing the product which maximizes their utility. Alternatively, one may assume that physicians choose on behalf and in the interest of patients. To abstract from informational constraints greatly simplifies the analysis.[13]

To distinguish pharmaceuticals from "regular" consumption goods, we assume that more is not better. More precisely, ill individuals do not gain from consuming more than one dose of a drug. For simplicity, they also do not gain from consuming different drugs.[14]

For reasons of tractability, we follow the common assumption in ideal variety models that firms simultaneously choose price and their "location" on the circumference of the product circle to maximize profits. In the present context, they also choose the quality Q_i of a drug at the same time.[15]

Finally, suppose that there exists a health insurance system which covers the risk of needing drug treatment. However, patients themselves have to pay a fraction $\tau \in [0, 1]$ of the price of medication – the coinsurance rate.[16] Health insurance is assumed to be fair, i.e., the insurance premium, T, is equal to the expected reimbursement

[11] See The Economist (2011) for a discussion on the efforts to tackle the resistance of antibiotics via R&D.

[12] We abstract from uncertainty in the R&D process. Nothing would change, however, if firms are successful in innovating and entering the market only with some probability, as long as there are many potential innovators which are risk-neutral. In this case, neither supply of R&D funds is affected by uncertainty (due to the law of large numbers) nor is demand.

[13] A priori, it seems unclear whether and how the nature of price competition in the pharmaceutical market would change under asymmetric information between physicians and patients and/or under limited information of both. These are challenging issues which are beyond the scope of the present paper.

[14] For some diseases treatment is more effective when several drugs are combined, like for attacking HIV, the virus that causes AIDS. In our context, this would be captured by defining a drug as a combination of active pharmaceutical ingredients in one dose. Many drugs contain several active ingredients, so it does not matter if those are combined in, say, one injection/pill or provided via several different injections/pills.

[15] Assuming that price and quality are chosen simultaneously draws on the seminal paper on R&D choice under imperfect competition in Dasgupta and Stiglitz (1980). Appendix (C) discusses the case of a two-stage decision process.

[16] We abstract from moral hazard – although sometimes being the alleged reason for implementing coinsurance schemes in the first place. This argument is

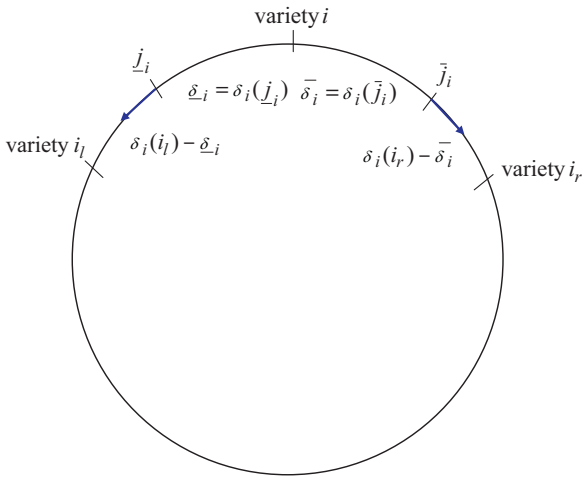

Fig. 1. The product circle of differentiated pharmaceutical goods and the impact of quality-improvements (indicated by arrows).

of patients' medication expenses from the insurance. In the next section we examine the effect of an increase in the coinsurance rate τ on the R&D expenditure of pharmaceutical firms and on prices of their products.

3.2. Equilibrium analysis

3.2.1. Equilibrium conditions

Consider the location of firm i on the circumference of the circle of illnesses. Denote the firm to the left of i by i_l and the firm to the right of i by i_r. The shorter (arc) distance between the location of i and i_l is denoted by $\delta_i(i_l)$ and the one between between i and i_r by $\delta_i(i_r)$. $D_i \equiv \delta_i(i_l) + \delta_i(i_r)$ is the distance between i_l and i_r. Denote by $\underline{j}_i$ the patient with the ideal variety to the left of firm i's location, who is indifferent between buying from firm i and i_l. Similarly, consumer $\bar{j}_i$ is indifferent between buying from i and i_r.

As shown in Fig. 1, $\underline{\delta}_i = \delta_i(\underline{j}_i)$ is the distance between i and $\underline{j}_i$ whereas $\bar{\delta}_i = \delta_i(\bar{j}_i)$ is the distance between i and $\bar{j}_i$. Define by

$$\underline{H}_i \equiv h(\underline{\delta}_i, Q_i), \quad \overline{H}_i \equiv h(\bar{\delta}_i, Q_i), \tag{3}$$

$$\underline{H}_{i_l} \equiv h(\delta_i(i_l) - \underline{\delta}_i, Q_{i_l}), \quad \overline{H}_{i_r} \equiv h(\underbrace{D_i - \delta_i(i_l)}_{=\delta_i(i_r)} - \bar{\delta}_i, Q_{i_r}), \tag{4}$$

the health levels of consumer $\underline{j}_i$ and $\bar{j}_i$, respectively, when consuming drug i (Eq. (3)) and the alternative drugs i_l, i_r (Eq. (4)). The associated consumption levels are given by

$$\underline{C}_i \equiv wg(\underline{H}_i) - \tau p_i - T, \quad \overline{C}_i \equiv wg(\overline{H}_i) - \tau p_i - T, \tag{5}$$

$$\underline{C}_{i_l} \equiv wg(\underline{H}_{i_l}) - \tau p_{i_l} - T, \quad \overline{C}_{i_r} \equiv wg(\overline{H}_{i_r}) - \tau p_{i_r} - T, \tag{6}$$

where p_i, p_{i_l}, p_{i_r} denote the price of drug i, i_l, i_r, respectively. $wg(H)$ is the wage income of an individual with health level H, τp_i is the coinsurance payment when consuming the drug supplied by firm i, and T is the insurance premium.[17] Note that $(\underline{\delta}_i + \bar{\delta}_i)s$ is the mass of consumers buying from firm i (recall that a mass s of consumers is ill). Thus, with a fair health insurance, the insurance premium for each individual is

$$T = (1-\tau)s\sum_{i=1}^{n}(\underline{\delta}_i + \bar{\delta}_i)p_i. \tag{7}$$

Moreover, the profit of a pharmaceutical firm i, taking into account R&D costs B, reads as

$$\pi_i = (p_i - c)(\underline{\delta}_i + \bar{\delta}_i)s - B(Q_i). \tag{8}$$

For individual $\underline{j}_i$, who is indifferent to buy from firm i and i_l, we have

$$0 = u(\underline{C}_i, \underline{H}_i) - u(\underline{C}_{i_l}, \underline{H}_{i_l}). \tag{9}$$

Substituting the respective first equations of (3)–(6) into (9) reveals that $\underline{\delta}_i$ is implicitly given as function of the price and quality of firm i, price and quality of its competitor i_l to the left, the distance to this competitor, $\delta_i(i_l)$, and parameters. We write this function as $\underline{\delta}_i = \underline{\Delta}_i(p_i, Q_i, p_{i_l}, Q_{i_l}, \delta_i(i_l), \tau, T, w)$. Similarly, for individual $\bar{j}_i$, who is indifferent to buy from firm i and i_r, we have

$$0 = u(\overline{C}_i, \overline{H}_i) - u(\overline{C}_{i_r}, \overline{H}_{i_r}). \tag{10}$$

Using the respective second equations of (3)–(6) in (10) implies that, analogously to the function $\underline{\Delta}_i$ which characterizes $\underline{\delta}_i$, we can write $\bar{\delta}_i = \overline{\Delta}_i(p_i, Q_i, p_{i_r}, Q_{i_r}, D_i - \delta_i(i_l), \tau, T, w)$.

According to (8), the profit maximization problem of firm i can then be written as

$$\max_{\delta_i(i_l), Q_i, p_i} (p_i - c)s\left[\underline{\Delta}_i(p_i, Q_i, \delta_i(i_l), p_{i_l}, Q_{i_l}, \tau, T, w) + \overline{\Delta}_i(p_i, Q_i, \underbrace{D_i - \delta_i(i_l)}_{=\delta_i(i_r)}, p_{i_r}, Q_{i_r}, \tau, T, w)\right] - B(Q_i), \tag{11}$$

taking as given p_{i_l}, Q_{i_l}, p_{i_r}, Q_{i_r}, D_i and T.

An equilibrium in the basic model (with restricted entry) is defined as locational choices, drug prices, and drug quality levels, in which firms simultaneously choose these variables to maximize profits, taking as given (and being fully informed about) the choices of the other firms, and ill consumers choose the drug which yields the highest utility given the choices of firms. Using expression (7) for the premium T in the first-order conditions which result from (11), we can derive the following lemma. (All proofs are relegated to Appendix (A).)

Lemma 1. *In a Nash symmetric equilibrium for a given number of firms, where $\underline{\delta}_i = \bar{\delta}_i = 1/2n$, $\delta_i(i_l) = \delta_i(i_r) = 1/n$, $Q_i = Q^*$ and $p_i = p^*$ for all i, equilibrium values (Q^*, p^*) are simultaneously given by*[18]

$$0 = -(p^* - c)s\frac{h_Q(1/2n, Q^*)}{h_\delta(1/2n, Q^*)} - B'(Q^*) \equiv F(Q^*, p^*, n), \tag{12}$$

$$0 = wg'\left(h\left(\frac{1}{2n}, Q^*\right)\right) + MRS(Q^*, p^*, n, w, \tau) - \frac{\tau n B'(Q^*)}{s h_Q(1/2n, Q^*)} \equiv G(Q^*, p^*, n, w, \tau), \tag{13}$$

unconvincing in the case of severe illness like cancer or AIDS, however. Health insurance systems are exogenous in our analysis.

[17] A healthy individual takes no decision in the model and consumes $C = w - T$.

[18] Parameters s and c are suppressed in functions F, G and MRS.

as functions of n, w, τ, *where*

$$MRS(Q^*, p^*, n, w, \tau) \equiv \frac{u_H(wg(h(1/2n, Q^*)) - [\tau(1-s)+s]p^*, h(1/2n, Q^*))}{u_C(wg(h(1/2n, Q^*)) - [\tau(1-s)+s]p^*, h(1/2n, Q^*))} \quad (14)$$

is the marginal rate of substitution between consumption and health in symmetric equilibrium. The equilibrium insurance premium, T^*, *reads as*

$$T^* = (1-\tau)sp^*, \quad (15)$$

The first summand on the right-hand side of Eq. (12) is the marginal benefit of raising quality Q which, in profit maximum, must be equal to the marginal cost of improving quality, $B'(Q)$. The marginal benefit of raising Q is higher, the higher the price-cost margin, $p-c$, the larger the total market size of the pharmaceutical market, s, the higher the effectiveness of R&D for health, h_Q, and the lower is the impact on health status of deviating from the ideal variety of a patient, $|h_\delta|$. To see intuitively that the ratio $h_Q/|h_\delta|$ matters for R&D incentives consider again the example of antibiotics. The innovation incentive is higher, the larger is for a given type of bacterium which causes an illness the effect of higher quality Q on health status, h_Q, and the wider the spectrum of bacteria and illnesses affected by the antibiotic (i.e., $|h_\delta|$ is lower).

According to (13), not surprisingly, the marginal rate of substitution between consumption and health affects demand for pharmaceuticals and therefore matters for the price-setting behavior of firms. Prices are also affected by the marginal impact of an increase in health on wage income, wg', which determines demand for pharmaceuticals as well. Moreover, demand becomes more price-sensitive when the coinsurance rate, τ, increases.

3.2.2. Comparative-static results

A symmetric equilibrium exists and is unique under weak conditions. An example of sufficient conditions on the primitives of the model, implying property $G_Q < F_Q G_p/F_p$, is spelled out in Appendix (B). Uniqueness of equilibrium allows us to derive comparative-static results.

Proposition 1 (Comparative-statics in the restricted entry case). *Suppose that* $G_Q < F_Q G_p/F_p$ *and there exists a unique equilibrium. (a) A higher coinsurance rate* τ *lowers the quality of pharmaceuticals,* Q^*, *their price,* p^*, *and the insurance premium,* T^*. *(b) A higher wage rate* w *raises* Q^*, p^* *and* T^*. *(c) An increase in the number of firms, n, may raise* Q^*. *For instance,* Q^* *is increasing in n if* $u_{CH} = u_{HH} = 0$, $g'' = 0$ *and*

$$\varepsilon\left(\frac{1}{2n}, Q\right) \equiv \left[-\frac{\delta h_{\delta Q}(\delta, Q)}{h_Q(\delta, Q)}\right]_{\delta=1/2n} \geq 1. \quad (A1)$$

If these properties hold, p^* *is increasing in n.*

The intuition for the negative impact of a higher coinsurance rate, τ, on R&D spending and prices (part (a) of Proposition 1) is simple. An increase in the fraction of the drug price which a patient has to copay implies that she becomes more price-sensitive. Thus, by raising the price of its drug, a pharmaceutical company loses more customers to rivals. This induces firms to lower the prices as an equilibrium response. In turn, the reduced mark-up over marginal costs, $p-c$, lowers the marginal benefit of R&D. Thus, raising τ achieves the goal of lowering pharmaceutical prices and the insurance premium, however, at the costs of reducing R&D spending.

A higher wage rate, w, raises the willingness to pay for drugs of ill consumers because of two effects. First, the price sensitivity declines after an increase in w if better health raises the supplied labor units ($g' > 0$). This effect arises since the marginal impact of better health on wage income rises with w. Second, the marginal rate of substitution between consumption and health, u_H/u_C, is increasing in w. Both effects go in the same direction and explain part (b).

Part (c) of Proposition 1 can be understood as follows. Elasticity ε measures by how much the marginal benefit of higher drug quality on health (h_Q) declines if δ increases by one percent. Recall that a higher δ means that the drug is less suited to the particular illness of a patient. Now suppose that the number of pharmaceuticals, n, increases. As a result, a firm loses customers for a given R&D spending since, on average, δ decreases. Consequently, for a given price of a drug and implied by $h_{\delta Q} < 0$, there is a higher incentive to conduct R&D in order to retain some of the customers. This effect is large if ε is high. Thus, under assumption (A1), firms may conduct more R&D.[19] The effect of an increase in n on the equilibrium price p^* is generally ambiguous. On the one hand, market share $1/n$ declines if n increases. Because of fewer customers, price-setting power of firms is reduced. On the other hand, on average, patients are closer to their ideal variety, which raises price setting power. The presumptions of part (c), which imply that quality Q^* rises with n, are also sufficient for p^* to rise with n. This reflects a positive relationship between price and quality in equilibrium.

3.3. Discussion

In this subsection, we relate the relationship between entry of firms and R&D effort to the literature. Next, we discuss it from two different angles. First, we allow for a competitive fringe which can imperfectly imitate the production process of incumbents by operating at higher marginal costs whereas the patent breadth on the product is very narrow. Second, we endogenize the number of firms, by allowing them to freely enter the market (which they do as long as profits are positive).

3.3.1. The relationship between competition and innovation

Part (c) of Proposition 1 shows that weaker protection against entry (e.g., by limiting patent breadth or fostering foreign entry) may foster R&D incentives. Hence, our analysis contributes to the recent debate on the relationship between competition and innovation. The literature has suggested that heterogeneity of firms with respect to their distance to the technology frontier is critical for the result that increased competition or entry deregulation can spur innovative effort (Aghion et al., 2005, 2009; Aghion and Howitt, 2005, 2009). The basic argument runs as follows. Incumbents operating at the technology frontier can escape competition or entry, i.e., secure a monopoly position, by innovating. Increased competition means that pre-innovation profits decline whereas post-innovation profits, which are pure monopoly profits by assumption, do not depend on the number of rivals. Facilitating entry thus raises R&D expenditure. By contrast, firms below the technology frontier see the difference between post- and pre-innovation profits decline if competitive pressure rises, as they cannot escape competition.

The present paper gives complementary insights on the competition-innovation relationship. It shows that the prospect of gaining pure monopoly power from innovation is not required for the result that increased entry can spur innovative effort. Rather, the result may even hold in an environment with strategically

[19] One can also show that higher market size, s, typically raises R&D incentives. This result is consistent with empirical evidence by Acemoglu and Linn (2004). According to their study, an increase in potential market size for drugs - measured by exploiting demographic trends in the US - has fostered pharmaceutical innovation.

interacting firms which possess similar technology. This is in contrast to Economides (1993), who also analyzes a model where firms locate on a circumference of a circle of horizontal product attributes. Focussing on a single-good economy, he implicitly assumes that second derivatives of the utility function u are zero, whereas the present model distinguishes health from regular consumption spending. Moreover, Economides (1993) implicitly assumes that the cross-derivative between distance to the ideal variety and product quality in the utility function is zero, which means $h_{\delta Q} = 0$ in the present context. Thus, in his model, increased entry cannot give firms incentives to raise R&D spending to compensate the resulting loss in market share associated with a larger number of firms.

3.3.2. Imitation

Consider the possibility of other firms to imitate existing pharmaceuticals in the sense that they can locate (without entry costs) at the same points as incumbents on the circumference of the circle of illnesses. This kind of imitation captures that patent breadth is very narrow with respect to product design (e.g., Vincenzo, 1996). However, suppose that incumbents have superior technological capability, possibly affected by the patent breadth, which determines how difficult it is to imitate the production process. Formally, the competitive fringe can imitate each existing pharmaceutical by locating at the same point as incumbents and producing each unit by cost $\chi \in (c, p^*)$. Whenever an incumbent firm sets a price $p > \chi$, it will be driven out of the market. If $p \leq \chi$, the competitive fringe would not make positive profits and thus has no reason to enter. Hence, each incumbent will set price $p = \chi < p^*$ to maximize profits by deterring entry. (If $\chi \geq p^*$, the incumbent would set price p^* in equilibrium and the existence of a competitive fringe would not affect the previous analysis.) We find the following result.

Proposition 2 (Imitation). *Suppose there is a competitive fringe which can locate at the same points as incumbents on the circumference of the circle of illnesses. The more cost-efficient the competitive fringe is (better imitation of the production process as captured by a decrease in χ), the lower is the equilibrium quality of pharmaceuticals.*

The intuition of Proposition 2 is simple. The existence of a competitive fringe limits the price-cost margin, as $\chi - c < p^* - c$, and therefore retards R&D incentives. Thus, although Proposition 1 suggests that encouraging entry or reducing patent breadth on the product may not curb pharmaceutical innovation, this does not mean that lower patent breadth of the production process (in the sense that the competitive fringe can better imitate the production process of the incumbents) spurs innovation in the case where there is zero patent breadth on the product.

3.3.3. Endogenous number of firms

We now show that the basic insights of Proposition 1 are not critically affected by allowing for an endogenous number of firms (unrestricted entry). We rule out the possibility of a competitive fringe to imitate pharmaceuticals without entry costs. Rather, pharmaceutical companies can freely enter the market by incurring $f > 0$ units of the numeraire. Thus, the number of firms, n, is endogenous and firms enter until profits become zero.

Analysis of the unrestricted entry case implicitly assumes that the patent breadth is smaller than the equilibrium value of $1/n$. A decrease in f may capture, for instance, lower administrative costs associated with weaker entry regulation (for examples and measurement, see Djankov et al., 2002). Also extensive marketing effort of pharma firms for branded prescription drugs via sales representatives (who directly contact physicians) erect entry barriers for potential rivals. Such entry barriers could be reduced (again, captured by a decrease in f) by regulating the activities of sales representatives in the pharmaceutical sector like restraining gift-giving to physicians. Also prohibiting that drug makers use doctors' prescribing data to develop marketing strategies could lead to a decrease in f.

According to profit function (8), in a symmetric situation, the zero-profit condition which results from free entry holds if

$$0 = \frac{(p-c)s}{n} - B(Q) - f \equiv Z(Q, p, n, f). \tag{16}$$

The equilibrium quality, price, and number of firms, denoted by (Q^{**}, p^{**}, n^{**}), are simultaneously given by equation system $F = G = Z = 0$. We define matrix

$$\mathbf{M} \equiv \begin{pmatrix} F_p & F_Q & F_n \\ G_p & G_Q & G_n \\ Z_p & Z_Q & Z_n \end{pmatrix}. \tag{17}$$

As shown in the proof of the following proposition, the determinant of $\mathbf{M}$ is positive under weak conditions.

Proposition 3 (Comparative-statics with an endogenous firm number). *Suppose that* $\det(\mathbf{M}) > 0$ *and there exists a unique equilibrium. Then, similar to the restricted entry case, (a) a higher coinsurance rate, τ, lowers equilibrium quality, Q^{**}, equilibrium prices of drugs, p^{**}, and the insurance premium $T^{**} = (1-\tau)sp^{**}$. (b) A higher wage rate, w, has the opposite effects. (c) If $u_{CH} = u_{HH} = g'' = 0$ and assumption (A1) holds, entry deregulation (decrease in f) tends to promote entry (n^{**} increases) and raises Q^{**}.*

Proposition 3 shows that, in unique equilibrium, the impact of an increase in the coinsurance rate and in the wage rate on R&D spending, prices of pharmaceuticals and the insurance premium as suggested by Proposition is robust to allowing for endogenous entry of firms. Moreover, not surprisingly, the number of firms typically declines if entry costs go up. Consistent with the effects of a change in the number of firms under restricted entry (part (c) of Proposition 1), retarding entry by higher fixed costs may be associated with reduced R&D spending.

4. Price regulations

This section examines the effects of price regulations. We distinguish between direct price controls and price caps on drug expenditure reimbursement. In its simplest form, on which we focus in this section, a price control means that the government sets a fixed, maximum drug price, $p_{\max}$, in a regime where health insurance does not cover drug expenses, i.e., the coinsurance rate is 100 percent ($\tau = 1$). A decrease in $p_{\max}$ captures stricter direct price regulation. By contrast, a price cap is a cost-sharing scheme which imposes a limit amount on the costs incurred by an insured patient which is reimbursed. Like coinsurance schemes, a price cap intends to keep insurance premiums low. The limit price is denoted by $\overline{p}$. We relate to a decrease in $\overline{p}$ as stricter price cap.

4.1. Price control

Suppose there is no health insurance, $\tau = 1$. A binding direct price control means that the price set by the government is below the equilibrium price which would result in absence of government intervention. We therefore implicitly assume in this section that the maximum drug price $p_{\max}$ is below the equilibrium price p^* (restricted entry case) and p^{**} (unrestricted entry case) which

results when $\tau = 1$. We focus on the simplest case where the government ignores R&D costs when setting $p_{\max}$.[20]

With restricted entry, the equilibrium drug quality, Q^*, is given by the first-order condition with respect to R&D,

$$F(Q^*, p_{\max}, n) = 0, \tag{18}$$

where function F was defined in (12). Under free entry, which implies that firms' equilibrium profits are zero, equilibrium drug quality, Q^{**}, and the number of firms, n^{**}, are simultaneously given by

$$F(Q^{**}, p_{\max}, n^{**}) = Z(Q^{**}, p_{\max}, n^{**}, f) = 0, \tag{19}$$

where function Z was defined in zero-profit condition (16).

Proposition 4 (Price controls). *Suppose there is a binding direct price control. (a) Stricter price regulation (decrease in $p_{\max}$) lowers the equilibrium quality of pharmaceuticals; with unrestricted entry, it also reduces the number of firms, n^{**}. (b) Under restricted entry, an increase in the number of firms, n, unambiguously raises the quality of drugs, Q^*. With unrestricted entry, entry deregulation (decrease in f) raises both the quality of drugs, Q^{**}, and n^{**}.*

A stricter price control limits the price-cost margin, $p_{\max} - c$, and therefore retards R&D incentives. The profit squeeze also retards entry (part (a) of Proposition 4). Deregulation of entry, which allows for a larger number of competitors, unambiguously raises R&D expenditure under direct price controls (part (b)). Similar to the discussion of the last result in Proposition 1, an increased number of drugs induces pharmaceuticals companies to retain some of its customers by raising R&D. As there is no counteracting effect on R&D incentives through reduced price-setting power, the result suggests than an increase in the competition intensity can unambiguously lead to more R&D in the proposed framework. Again, note that the mechanism is different to the existing literature on competition and R&D, by not resting on the prospect of monopoly.

4.2. Price caps

A health system which reimburses prescription drug expenses up to a price $\overline{p} > 0$ typically raises the demand for drugs vis-à-vis a free market without any insurance (which is captured by setting $\tau = 1$ in Section 3). To see this, first note that the fair insurance premium under a binding price cap (i.e., $\overline{p}$ is lower than the equilibrium price with full insurance) is given by $T = s\overline{p}$. Thus, total health expenditures for a customer of firm i is $p_i - \overline{p} + T = p_i - (1-s)\overline{p} < p_i$. Hence, a stricter price cap is not an intervention in a free market but restricts the subsidy on drug expenditure to beneficiaries. Demand faced by pharmaceutical companies is lowered by a stricter price cap, since a decrease in $\overline{p}$ lowers the marginal rate of substitution, $MRS = u_H/u_C$, in equilibrium with symmetric firms. This can be seen as follows. A customer of firm i with health status H has a consumption level of $wg(H) - p_i + (1-s)\overline{p}$. Thus,

$$MRS = \frac{u_H(wg(H) - p_i + (1-s)\overline{p}, H)}{u_C(wg(H) - p_i + (1-s)\overline{p}, H)}. \tag{20}$$

The right-hand side of (20) is increasing in $\overline{p}$. We find the following result.

Proposition 5 (Price caps). *A stricter price cap (decrease in $\overline{p}$) lowers both the price and quality of pharmaceuticals in symmetric equilibrium. Entry regulations have similar effects as in the basic model.*

Since the marginal rate of substitution decreases with a stricter price cap, firms have less price setting power. In turn, this is associated with a decrease in R&D spending. Regarding entry regulations, the same discussion as for Propositions 1 and 3 applies.

5. The "optimal" number of firms

We now return to the basic model with restricted entry and a coinsurance scheme. Proposition 1 suggests that entry deregulation like a decreased patent breadth may encourage incumbents to invest in R&D. However, intuitively, if the number of rivals becomes very large, the effect of a decreased price-cost margin on the return to R&D should dominate the effect that firms attempt to retain some customers by raising quality. This suggests that there may be an "optimal" number of firms in the sense of maximizing R&D effort per firm. This section examines such a case and characterizes the optimal number of firms in terms of the important parameters of the model.[21]

To simplify the analysis, let us assume the following functional forms which are in line with the properties assumed in Section 3. On the technological side, specify R&D-technology as $B(Q) = 0.5Q^2$ and suppose for notational simplicity that marginal costs are zero, $c = 0$. The utility function is specified to

$$u(C, H) = \ln C + H, \tag{21}$$

i.e., $u_{CH} = u_{HH} = 0$ and $u_{CC} < 0$. Moreover, let us introduce a maximum distance to the ideal variety such that a drug is effective, denoted by $\delta^{\max} \in (0, 1)$. Suppose that individual labor supply is equal to the health status, which itself is a Cobb–Douglas function of product quality, Q, and the difference between $\delta^{\max}$ and the actual distance to the ideal variety, δ:

$$g(H) = H = h(\delta, Q) = \begin{cases} Q^{\alpha}(\delta^{\max} - \delta)^{\beta} \text{ if } \delta \le \delta^{\max}, \\ 0 \text{ otherwise}, \end{cases} \tag{22}$$

$\alpha, \beta \in (0, 1]$. Thus, $g'' = 0$, $h_{QQ} \le 0$, $h_{\delta\delta} \le 0$ and $h_{\delta Q} < 0$. Also suppose that the quality choice of firms is restricted to $0 \le Q \le Q^{\max}$, where $Q^{\max} \equiv (\delta^{\max})^{-\beta/\alpha}$ solves $h(0, Q^{\max}) = 1$ (recall normalization $H = 1$ if an individual is healthy).[22] With these functional forms, we can prove existence and uniqueness of equilibrium.

Lemma 2. *Under the assumed specifications, there exists a unique equilibrium quality level, Q^*.*

Moreover, according to part (c) of Proposition 1, properties $u_{CH} = u_{HH} = 0$ and $g'' = 0$ imply that *equilibrium quality Q^* is increasing in the number of firms, n, if assumption (A1) holds*. According to (A1) and (22), this is the case when

$$\varepsilon\left(\frac{1}{2n}, Q\right) = \left[\frac{\delta\beta}{\delta^{\max} - \delta}\right]_{\delta = 1/2n} \ge 1 \Leftrightarrow n \le \frac{1+\beta}{2\delta^{\max}} \equiv \overline{n}. \tag{23}$$

Thus, only if n is sufficiently high, product quality Q^* can be *non-increasing* in firm number n.

[20] Price controls follow a redistributive goal, aiming to reduce the financial burden of the ill vis-à-vis the healthy. Pharmaceutical prices in this regime are typically negotiated between pharmaceutical companies and the government. Critics of price controls argue that negotiated sales prices insufficiently account for R&D costs. The analysis would become more complicated if R&D costs and the effectiveness of drugs played a role in the setting $p_{\max}$.

[21] The focus is still on a positive analysis which is why the term "optimal" is set in parenthesis. A normative analysis would presume the choice of an appropriate welfare function, balancing the interests between ill and healthy individuals, between individuals and firms, and within the group of patients. Such an analysis is beyond the scope of the current paper.

[22] Moreover, we implicitly assume that $n \ge 1/2\delta^{\max}$, which ensures that $\delta \le \delta^{\max}$ at $\delta = 1/2n$.

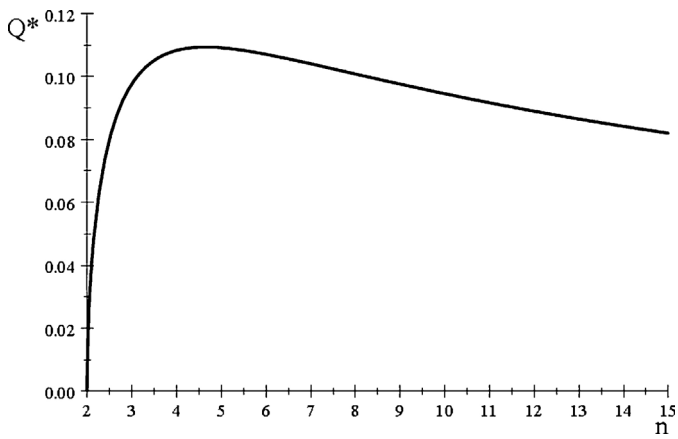

Fig. 2. Hump-shaped relationship between the equilibrium product quality, Q^*, and the number of firms, n. Note: $\alpha = \beta = w = 1$, $\tau = 0.2$, $s = 0.1$, $\delta^{\max} = 0.25$.

The optimal number of firms is defined as $n^{opt} \equiv \underset{n}{argmax} Q^*$. Ignoring the integer problem, inequality (23) and part (c) of Proposition 1 suggest that $n^{opt} > \overline{n}$.

Proposition 6 (Optimal number of firms). *Under the assumed specifications, the following holds. (a) The relationship between the equilibrium quality level Q^* and the number of firms may be hump-shaped. (b) Any interior level of n^{opt} is unique, decreasing in the coinsurance rate, τ, and increasing in the wage rate (labor productivity), w.*

Part (a) of Proposition 6 is illustrated in Fig. 2. It is based on a coinsurance rate of 20 percent ($\tau = 0.2$) and a fraction of ill individuals of 10 percent ($s = 0.1$). In Fig. 2, Q^* reaches its maximum at $n = 4.65 > \overline{n} = 4$. As visible from Fig. 2, the optimal number of firms, as an integer, is $n^{opt} = 4$. If the coinsurance rate is reduced to 10 percent ($\tau = 0.1$), then the optimal number of firms, as an integer, rises to $n^{opt} = 5$ (not shown). The intuitive reason for this result, qualitatively suggested by part (b) of Proposition 6, is the following. Suppose that the number of firms, n, initially maximizes R&D per firm, Q^*. In other words, competition is already quite fierce (i.e., the patent breadth is quite narrow) such that an increase in n would reduce Q^*. Now suppose that the coinsurance rate is reduced and therefore patients become less price-sensitive. As a consequence, firms raise prices and, because the profits increase ceteris paribus, they also raise Q^* (part (a) of Proposition 1). A slight increase in n now means that they have more to lose to rivals which is why they want to raise R&D further to respond to the fiercer competition. Graphically, the curve in Fig. 2 then shifts to the North-East (not shown). The same occurs if labor productivity, w, increases, implying that the ability to pay of patients and therefore profits of firms rise ceteris paribus.

6. Conclusion

This paper has examined the role of cost-sharing schemes in health insurance systems and deregulation of the pharmaceutical industry for prices of pharmaceuticals and pharmaceutical R&D expenditure per firm. According to the proposed model, extending coinsurance or applying stricter price regulations adversely affects pharmaceutical R&D spending while lowering drug prices. For instance, the 2006 Medicare reform in the US introduced coverage of expenses for prescription drugs, effectively reducing the coinsurance rate from 100 percent to 25 percent and less. Consistent with empirical evidence (Blume-Kohout and Sood, 2013), our analysis suggests that the reform has induced pharmaceutical firms to intensify their innovative effort.[23]

Possibly more surprising at the first glance, lifting entry barriers may also spur pharmaceutical innovations. The possibility arises when better suitability of a drug for patients, resulting from increased variety, leads to a sufficient increase in the effectiveness of R&D on health. If this is the case and the number of firms is not too large, increased entry implies that firms attempt to retain some of their customers by quality-improvements in response to a loss in their market share. Examples for appropriate entry deregulation policies in the pharmaceutical sector would be to limit non-informative marketing expenses, to promote access of foreign pharmaceutical companies to domestic markets, or to reduce patent breadth with respect to the design of a pharmaceutical product.

The analysis also suggests, however, that if competition becomes too fierce, firms may reduce their pharmaceutical R&D spending. That is, the relationship between the number of firms and R&D effort per firm may be hump-shaped. In this case, the R&D maximizing number of firms is decreasing in the coinsurance rate and increasing in the stage of development (captured by the productivity of labor). Finally, if patent breadth is very narrow with respect to product design, improving the possibility to imitate the production process of incumbents would unambiguously curb pharmaceutical innovation.

Future research may build on the proposed theory to quantify welfare effects and to derive socially optimal cost-sharing schemes. For instance, limiting the coinsurance rate on drug expenditures may be warranted for a number of reasons. First, there is the standard positive welfare effect of providing health insurance to risk-averse households. Second, as focussed upon in this paper, R&D spending may be adversely related to the coinsurance rate. Thus, limiting cost-sharing may enlarge standard intertemporal "standing-on-shoulders" externalities from pharmaceutical R&D. For simplicity, such spillover effects have been ignored in our static set up but should be allowed for in future research, along with other, possibly negative, R&D externalities (see, e.g., Jones and Williams, 2000). Another important issue which is worth investigating is to allow firms to discriminate prices. How increased competition affects R&D incentives if patients are charged lower prices than others when a drug is less suited for them is an open question. It would be interesting to examine how the nature of price competition changes in the pharmaceutical sector along with R&D incentives when drug-prescribing physicians have superior information and pursue own interests. Finally, this paper calls for an empirical investigation of the effect of changes in patent breadth on pharmaceutical R&D which allows for a non-linear relationship.

Acknowledgements

I am grateful to two anonymous referees as well as to seminar participants at the University of Fribourg, the CESifo Area Conference in Public Economics 2011 in Munich, the Second LEPAS Workshop on "The Economics of Ageing" 2011 in Alicante, and the annual meeting of the European Economic Association 2011 in Oslo for valuable comments and suggestions.

[23] In a similar vein, Microsimulations of the so-called "Global Pharmaceutical Policy Model" (Lakdawalla et al., 2009) suggest that stricter price controls and higher coinsurance rates have adverse effects on longevity.

Appendix A. Proofs

Proof of Lemma 1. The first-order conditions associated with profit-maximization problem (11) with respect to $\delta_i(i_l)$, Q_i, p_i are

$$\frac{\partial \underline{\Delta}_i}{\partial \delta_i(i_l)} + \frac{\partial \overline{\Delta}_i}{\partial \delta_i(i_l)} = 0, \tag{24}$$

$$(p_i - c)s\left(\frac{\partial \underline{\Delta}_i}{\partial Q_i} + \frac{\partial \overline{\Delta}_i}{\partial Q_i}\right) - B'(Q_i) = 0, \tag{25}$$

$$\underline{\Delta}_i + \overline{\Delta}_i + (p_i - c)\left(\frac{\partial \underline{\Delta}_i}{\partial p_i} + \frac{\partial \overline{\Delta}_i}{\partial p_i}\right) = 0. \tag{26}$$

Using the first equations of (3)–(6) in (9) and applying the implicit function theorem, we obtain

$$\frac{\partial \underline{\Delta}_i}{\partial \delta_i(i_l)} = \frac{[u_C(\underline{C}_{i_l}, \underline{H}_{i_l})wg'(\underline{H}_{i_l}) + u_H(\underline{C}_{i_l}, \underline{H}_{i_l})]h_\delta(\delta_i(i_l) - \underline{\delta}_i, Q_{i_l})}{\underline{\Omega}_i}, \tag{27}$$

$$\frac{\partial \underline{\Delta}_i}{\partial Q_i} = -\frac{[u_C(\underline{C}_i, \underline{H}_i)wg'(\underline{H}_{i_l}) + u_H(\underline{C}_i, \underline{H}_i)]h_Q(\underline{\delta}_i, Q_i)}{\underline{\Omega}_i}, \tag{28}$$

$$\frac{\partial \underline{\Delta}_i}{\partial p_i} = \frac{u_C(\underline{C}_i, \underline{H}_i)\tau}{\underline{\Omega}_i}, \tag{29}$$

where

$$\begin{aligned}\underline{\Omega}_i \equiv\ & [u_C(\underline{C}_i, \underline{H}_i)wg'(\underline{H}_i) + u_H(\underline{C}_i, \underline{H}_i)]h_\delta(\underline{\delta}_i, Q_i) \\ & + [u_C(\underline{C}_{i_l}, \underline{H}_{i_l})wg'(\underline{H}_{i_l}) + u_H(\underline{C}_{i_l}, \underline{H}_{i_l})]h_\delta(\delta_i(i_l) - \underline{\delta}_i, Q_{i_l}).\end{aligned} \tag{30}$$

Similarly, using the second equations of (3)–(6) in (10), we obtain

$$\frac{\partial \overline{\Delta}_i}{\partial \delta_i(i_l)} = -\frac{[u_C(\overline{C}_{i_r}, \overline{H}_{i_r})wg'(\overline{H}_{i_r}) + u_H(\overline{C}_{i_r}, \overline{H}_{i_r})]h_\delta(\delta_i(i_r) - \overline{\delta}_i, Q_{i_r})}{\overline{\Omega}_i}, \tag{31}$$

$$\frac{\partial \overline{\Delta}_i}{\partial Q_i} = -\frac{[u_C(\overline{C}_i, \overline{H}_i)wg'(\overline{H}_i) + u_H(\overline{C}_i, \overline{H}_i)]h_Q(\overline{\delta}_i, Q_i)}{\overline{\Omega}_i}, \tag{32}$$

$$\frac{\partial \overline{\Delta}_i}{\partial p_i} = \frac{u_C(\overline{C}_i, \overline{H}_i)\tau}{\overline{\Omega}_i}, \tag{33}$$

where

$$\begin{aligned}\overline{\Omega}_i \equiv\ & [u_C(\overline{C}_i, \overline{H}_i)wg'(\overline{H}_i) + u_H(\overline{C}_i, \overline{H}_i)]h_\delta(\overline{\delta}_i, Q_i) \\ & + [u_C(\overline{C}_{i_r}, \overline{H}_{i_r})wg'(\overline{H}_{i_r}) + u_H(\overline{C}_{i_r}, \overline{H}_{i_r})]h_\delta(\delta_i(i_r) - \underline{\delta}_i, Q_{i_r}).\end{aligned} \tag{34}$$

In a symmetric situation, where $\underline{\delta}_i = \overline{\delta}_i = 1/2n$, $\delta_i(i_l) = \delta_i(i_r) = 1/n$, $Q_i = Q$ and $p_i = p$ for all i, and thus $\underline{C}_i = \overline{C}_i = \underline{C}_{i_l} = \overline{C}_{i_r} = \overline{C}$ as well as $\underline{H}_i = \overline{H}_i = \underline{H}_{i_l} = \overline{H}_{i_r} = \overline{H}$, we have $\partial \underline{\Delta}_i / \partial \delta_i(i_l) = 0.5$ and $\partial \overline{\Delta}_i / \partial \delta_i(i_l) = -0.5$, according to (27) and (31), respectively, using (30) and (34). Thus, (24) holds. Moreover, using (28) and (32) as well as again (30) and (34) in (25) confirms (12). To prove (13), use analogously (29), (33), (30) and (34) in (26) to obtain, under symmetry,

$$\frac{1}{n} + \frac{(p-c)\tau}{(wg'(\overline{H}) + (u_H(\overline{C}, \overline{H})/u_C(\overline{C}, \overline{H})))h_\delta(1/2n, Q)} = 0. \tag{35}$$

Also note that using $\underline{\delta}_i + \overline{\delta}_i = 1/n$ and $p_i = p^*$ for all i in (7) implies (15). Thus,

$$\overline{C} = wg(\overline{H}) - [\tau(1-s) + s]p, \quad \text{with } \overline{H} = h\left(\frac{1}{2n}, Q\right), \tag{36}$$

according to (5) and (3). Finally, rewriting (12) to

$$p = c - \frac{B'(Q)}{s}\frac{h_\delta(1/2n, Q)}{h_Q(1/2n, Q)}, \tag{37}$$

substituting (37) into (35) and using (36) confirms (13). □

Proof of Proposition 1. We have $F_p > 0$ and

$$F_Q = -(p-c)s\frac{h_{QQ}h_\delta - h_Q h_{\delta Q}}{(h_\delta)}^2 - B'' < 0, \tag{38}$$

according to (12) and the properties of functions h and B. Thus, the $F = 0$-locus in $Q - p$-space is upward-sloping. Moreover,

$$G_p = MRS_p = -\frac{[u_{CH}u_C - u_H u_{CC}][\tau(1-s) + s]}{(u_C)}^2 < 0, \tag{39}$$

according to (13), (14), and the properties of utility function u.

The determinant of matrix $\begin{pmatrix} F_p & F_Q \\ G_p & G_Q \end{pmatrix}$ is negative if and only if

$$G_Q = wg''h_Q + MRS_Q - \frac{\tau n}{s}\frac{B''h_Q - B'h_{QQ}}{(h_Q)^2} < \frac{F_Q G_p}{F_p}. \tag{40}$$

A sufficient condition for this inequality to hold is $G_Q \leq 0$. Appendix (B) discusses when this is the case. If $G_Q < F_Q G_p / F_p$, the $F = 0$-locus is steeper than the $G = 0$-locus in $Q - p$-space and any equilibrium is unique. Using $G_Q < F_Q G_p / F_p$ and applying the implicit function theorem, we have

$$sgn\left(\frac{\partial Q^*}{\partial \tau}\right) = sgn\left(F_p G_\tau - F_\tau G_p\right), \tag{41}$$

$$sgn\left(\frac{\partial p^*}{\partial \tau}\right) = sgn\left(F_\tau G_Q - F_Q G_\tau\right), \tag{42}$$

$$sgn\left(\frac{\partial Q^*}{\partial w}\right) = sgn\left(F_p G_w - F_w G_p\right), \tag{43}$$

$$sgn\left(\frac{\partial p^*}{\partial w}\right) = sgn\left(F_w G_Q - F_Q G_w\right), \tag{44}$$

$$sgn\left(\frac{\partial Q^*}{\partial n}\right) = sgn\left(F_p G_n - F_n G_p\right). \tag{45}$$

$$sgn\left(\frac{\partial p^*}{\partial n}\right) = sgn\left(F_n G_Q - F_Q G_n\right), \tag{46}$$

Now, note that $F_\tau = F_w = 0$, $G_\tau < 0$, $G_w > 0$, according to (12)–(14). Moreover,

$$F_n = (p-c)s\frac{h_{\delta Q}h_\delta - h_Q h_{\delta\delta}}{(h_\delta)}^2 \frac{1}{2n^2} > 0, \tag{47}$$

according to (12) and the properties of function h. Together with $F_p < 0$ and $G_p < 0$, comparative-static results regarding changes in τ and w then follow from (41) to (44). This confirms parts (a) and (b).

To prove part (c) note that $\partial Q^*/\partial n > 0$ and $\partial p^*/\partial n > 0$ if $G_n \geq 0$, according to (45) and (46), respectively. Using (13), we have

$$G_n = -\frac{wg''(\overline{H})h_\delta}{2n^2} + MRS_n + \frac{\tau B'(Q)}{sh_Q}\left[\varepsilon\left(\frac{1}{2n}, Q\right) - 1\right], \tag{48}$$

where

$$MRS_n = -\frac{h_\delta}{2n^2 u_C}\left[u_{CH} wg' + u_{HH} - \frac{u_H}{u_C}\left(u_{CC} wg' + u_{CH}\right)\right], \tag{49}$$

according to (14). Thus, $MRS_n \geq 0$ if $u_{CH} = u_{HH} = 0$. If, in addition, $g'' = 0$ and $\varepsilon \geq 1$, then $G_n \geq 0$, according to (48). This concludes the proof. □

Proof of Proposition 2. Incumbents maximize the expression in (11) with respect to $\delta_i(i_l)$ and Q_i while $p_i = p_{i_l} = p_{i_r} = \chi$. Thus, rather than by $F(Q^*, p^*, n) = 0$, the equilibrium quality is given by $F(Q, \chi, n) = 0$. As $\chi < p^*$ by assumption, properties $F_p > 0$ and $F_Q < 0$ imply that the equilibrium quality in presence of a competitive fringe fulfills $Q < Q^*$. □

Proof of Proposition 3. First, note that $Z_Q < 0$, $Z_p > 0$, $Z_n < 0$, $Z_f < 0$ and $Z_\tau = Z_w = 0$, according to (16). As a remark,

$$\det(\mathbf{M}) = F_p G_Q Z_n + F_Q G_n Z_p + F_n G_p Z_Q - F_Q G_p Z_n - F_p G_n Z_Q - F_n G_Q Z_p \tag{50}$$

is positive if $G_Q \leq 0$ (again, see Appendix (B) for sufficient conditions) and G_n is small in magnitude. (Clearly, these are not necessary conditions for $\det(\mathbf{M}) > 0$.) According to (14), if $g' = g'' = u_{CH} = u_{HH} = 0$ and $\varepsilon = 1$ in equilibrium, then $MRS_Q = MRS_n = G_n = 0$ and therefore $\det(\mathbf{M}) > 0$ (recall that $F_p > 0$, $F_Q < 0$, $F_n > 0$ and $G_p < 0$).

Using $F_\tau = F_w = F_f = G_f = Z_\tau = Z_w = 0$, if $\det(\mathbf{M}) > 0$, the implicit function theorem implies that

$$sgn\left(\frac{\partial Q^{**}}{\partial \tau}\right) = -sgn\left(F_p G_\tau Z_n - F_n G_\tau Z_p\right), \tag{51}$$

$$sgn\left(\frac{\partial p^{**}}{\partial \tau}\right) = -sgn\left(F_n G_\tau Z_Q - F_Q G_\tau Z_n\right), \tag{52}$$

$$sgn\left(\frac{\partial Q^{**}}{\partial w}\right) = -sgn\left(F_p G_w Z_n - F_n G_w Z_p\right), \tag{53}$$

$$sgn\left(\frac{\partial p^{**}}{\partial w}\right) = -sgn\left(F_n G_w Z_Q - F_Q G_w Z_n\right), \tag{54}$$

$$sgn\left(\frac{\partial Q^{**}}{\partial f}\right) = -sgn\left(F_n G_p Z_f - F_p G_n Z_f\right), \tag{55}$$

$$sgn\left(\frac{\partial n^{**}}{\partial f}\right) = -sgn\left(F_p G_Q Z_f - F_Q G_p Z_f\right). \tag{56}$$

From (51) to (54) we can confirm the impact of an increase in τ and w on Q^{**}, p^{**} and T^{**} (parts (a) and (b)). Concerning part (c), from (55), we find that $\partial Q^{**}/\partial f < 0$ if $G_n \geq 0$, which holds under the presumptions of part (c) of Proposition 1. Finally, (56) implies that $\partial n^{**}/\partial f < 0$ if $G_Q \leq 0$. □

Proof of Proposition 4. To prove the results for the restricted entry case, recall that $F_Q < 0$, $F_p > 0$, $F_n > 0$ and apply the implicit function theorem. To prove the results in the case of unrestricted entry, recall that $Z_Q < 0$, $Z_p > 0$, $Z_n < 0$, $Z_f < 0$, $F_f = 0$. Thus, the determinant of the matrix $\begin{pmatrix} F_Q & F_n \\ Z_Q & Z_n \end{pmatrix}$ is positive. Consequently, by applying the implicit function theorem to (19), we find that

$$sgn\left(\frac{\partial Q^{**}}{\partial p_{\max}}\right) = -sgn(F_p Z_n - F_n Z_p) > 0, \tag{57}$$

$$sgn\left(\frac{\partial n^{**}}{\partial p_{\max}}\right) = -sgn(F_Q Z_p - F_p Z_Q) > 0, \tag{58}$$

$$sgn\left(\frac{\partial Q^{**}}{\partial f}\right) = sgn(F_n Z_f) < 0, \tag{59}$$

$$sgn\left(\frac{\partial n^{**}}{\partial f}\right) = -sgn(F_Q Z_f) < 0. \tag{60}$$

This concludes the proof. □

Proof of Proposition 5. As argued in Section (4.2), health expenditures for a customer of firm i are $p_i - \overline{p} + T$ (with $T = s\overline{p}$) rather than $\tau p_i + T$ compared to the basic model with coinsurance. Thus, (5) and (6) modify to

$$\underline{C}_i \equiv wg(\underline{H}_i) - p_i + (1-s)\overline{p}, \quad \overline{C}_i \equiv wg(\overline{H}_i) - p_i + (1-s)\overline{p}, \tag{61}$$

$$\underline{C}_{i_l} \equiv wg(\underline{H}_{i_l}) - p_{i_l} + (1-s)\overline{p}, \quad \overline{C}_{i_r} \equiv wg(\overline{H}_{i_r}) - p_{i_r} + (1-s)\overline{p}, \tag{62}$$

We can now employ the conditions for a profit maximum of firms under restricted entry in Lemma 1, Eqs. (24)–(34), except that we have to set $\tau = 1$ in (29) and (33) and evaluate at (61) and (62) instead of (5) and (6), respectively. Making use of the facts that $H = h(1/2n, Q)$ and $p_i = p$ hold under symmetry in (20), we find that equilibrium values (Q^*, p^*) under restricted entry are simultaneously given by

$$0 = F(Q^*, p^*, n), \tag{63}$$

$$0 = wg'\left(h\left(\frac{1}{2n}, Q^*\right)\right) + MRS(Q^*, p^*, n, w, \overline{p}) - \frac{nB'(Q^*)}{sh_Q(1/2n, Q^*)}$$
$$\equiv \overline{G}(Q^*, p^*, n, w, \overline{p}), \tag{64}$$

where

$$MRS(Q^*, p^*, n, w, \overline{p}) \equiv \frac{u_H(wg(h(1/2n, Q^*)) - p^* + (1-s)\overline{p}, h(1/2n, Q^*))}{u_C(wg(h(1/2n, Q^*)) - p^* + (1-s)\overline{p}, h(1/2n, Q^*))}. \tag{65}$$

If the number of firms is endogenous, equilibrium values (Q^{**}, p^{**}, n^{**}) are given by

$$F(Q^{**}, p^{**}, n^{**}) = \overline{G}(Q^{**}, p^{**}, n^{**}, w, \overline{p}) = Z(Q^{**}, p^{**}, n^{**}, f) = 0. \tag{66}$$

Since MRS is increasing in $\overline{p}$, we have $\overline{G}_{\overline{p}} > 0$. The remainder of the proof is then analogous to the proofs of Proposition 1 and 2 (where we now use properties $F_{\overline{p}} = 0$ and $\overline{G}_{\overline{p}} > 0$ instead of $F_\tau = 0$ and $G_\tau < 0$, respectively). Thus, a decrease in $\overline{p}$ has similar effects than an increase in τ in the basic model. A change in the number of firms has a similar effect on function $\overline{G}$ than on function G of the basic model. This concludes the proof. □

Proof of Lemma 2. Using (21) and (22) in (14), we find

$$MRS(Q^*, p^*, n, w, \tau) = w(Q^*)^\alpha\left(\delta^{\max} - \frac{1}{2n}\right)^\beta - [\tau(1-s) + s]p^*. \tag{67}$$

Using $c = 0$, $B'(Q) = Q$, and (22) in (37), we have

$$p^* = \frac{\beta}{\alpha s}\frac{Q^*}{\delta^{\max} - 1/2n}. \tag{68}$$

Define $N \equiv \delta^{\max} - 1/2n$ such that $n = 1/2(\delta^{\max} - N)$. Using this together with (67) and (68) in (13) and rearranging terms, we find that equilibrium quality Q^* solves

$$0 = \underbrace{\frac{w[1+Q^{\alpha}N^{\beta}] - \tau Q^{2-\alpha}/2s\alpha(\delta^{\max}-N)N^{\beta}}{\tau(1-s)+s}}_{\equiv\Phi(Q,N,w,\tau)} - \underbrace{\frac{\beta}{\alpha s}\frac{Q^2}{N}}_{\equiv\Psi(Q,N)} \equiv \Lambda(Q,N,w,\tau). \tag{69}$$

Since $\Phi(0,N,w,\tau) > 0 = \Psi(0,N)$, Φ is strictly concave as a function of Q, and Ψ is strictly increasing and strictly convex as a function of Q, there exists a unique level $Q^* = Q^*(N,w,\tau)$,[24] which solves $\Phi(Q,N,w,\tau) = \Psi(Q,N)$ for Q. This concludes the proof. □

Proof of Proposition 6. To show that the relationship between Q^* and n may be hump-shaped, an example suffices. For instance, suppose that $\alpha = 1$. In this case, using (69), it is easy to show that

$$Q^* = \frac{wN^{\beta}}{2T(N,\tau)} + \left(\frac{w^2N^{2\beta}}{4T(N,\tau)}^{2} + \frac{w}{T(N,\tau)}\right)^{1/2},$$

where $T(N,\tau) \equiv (\tau(1-s)+s/s)(\beta/N) + (\tau/2s(\delta-N)N^{\beta})$ and $N = \delta - 1/2n$. Fig. 2 plots Q^* as a hump-shaped function of n for reasonable parameter values. This confirms part (a).

We now show generally that any interior solution n^{opt} is unique. Note that there is a positive relationship between n and $N = \delta - 1/2n$ which does neither depend on wage rate w nor on coinsurance rate τ. We can define the optimal value of N as

$$N^{opt}(w,\tau) \equiv \underset{N}{argmax}\, Q^*(N,w,\tau) \tag{70}$$

and use $n^{opt} = 1/2(\delta^{\max} - N^{opt})$ to derive comparative-static effects of an increase in τ and w on n^{opt}. Using (69), it is easy to show that

$$\Lambda_Q(Q^*,N,w,\tau) = -\frac{1}{Q^*}\left(\frac{(2-\alpha)w + 2(1-\alpha)w\cdot(Q^*)^{\alpha}N^{\beta}}{\tau(1-s)+s} + \frac{\beta}{s}\frac{Q^*}{N}\right) < 0. \tag{71}$$

As $\partial Q^*/\partial N = -\Lambda_N(Q^*,N,\cdot)/\Lambda_Q(Q^*,N,\cdot)$, the first-order condition of maximization problem (70) which gives us N^{opt}, reads as

$$\Lambda_N(Q^*(N^{opt},w,\tau),N^{opt},w,\tau) = 0. \tag{72}$$

The second-order condition holds if $\Lambda_{NN}(Q^*(N^{opt},\cdot),N^{opt},\cdot) < 0$. According to (69),[25]

$$\Lambda_N = \frac{w\beta Q^{\alpha}N^{\beta-1} - (\tau Q^{2-\alpha}/2s\alpha)((1+\beta)N - \beta\delta^{\max}/(\delta^{\max}-N)^2N^{1+\beta})}{\tau(1-s)+s} + \frac{\beta}{\alpha s}\frac{Q^2}{N^2}. \tag{73}$$

Using the definition of Q^* as given by (69) in (73) implies that

$$\Lambda_N(Q^*,N,w,\tau) = \frac{w/N + w(1+\beta)(Q^*)^{\alpha}N^{\beta-1} - (\tau(Q^*)^{2-\alpha}/2s\alpha)\Theta(N)}{\tau(1-s)+s}, \tag{74}$$

where $\Theta(N) \equiv (\beta N - (1-\beta)\delta^{\max})/((\delta^{\max}-N)^2N^{1+\beta})$. Thus, if $\Theta'(N^{opt}) \geq 0$, then the second-order condition holds. Moreover, $\Lambda_N(Q^*,N^{opt},w,\tau) = 0$ can hold only if $N^{opt} > ((1-\beta)\delta^{\max}/\beta)$. As $\delta^{\max} > N$ for all $n > 0$, $N > ((1-\beta)\delta^{\max}/\beta)$ means that $\Lambda_N = 0$ can hold only if $\beta > 0.5$. One can show that $\Theta'(N^{opt}) \geq 0$ indeed holds for all $\beta \in (0.5, 1]$. Thus, any N^{opt} which solves first-order condition (72) is a unique solution for maximization problem (70).

To prove the comparative-static results in part (b), apply the implicit function theorem to (72), employ $\partial Q^*/\partial N|_{N=N^{opt}} = 0$ by the envelope theorem, use $\partial Q^*/\partial\tau = -\Lambda_{\tau}(Q^*,\cdot)/\Lambda_Q(Q^*,\cdot)$ and $\partial Q^*/\partial w = -\Lambda_w(Q^*,\cdot)/\Lambda_Q(Q^*,\cdot)$, and recall $\Lambda_Q(Q^*,\cdot) < 0$ to find that

$$\frac{\partial N^{opt}}{\partial\tau} < 0 \Leftrightarrow [\Lambda_Q\Lambda_{N\tau} - \Lambda_{NQ}\Lambda_{\tau}]_{N=N^{opt}} > 0, \tag{75}$$

$$\frac{\partial N^{opt}}{\partial w} > 0 \Leftrightarrow [\Lambda_Q\Lambda_{Nw} - \Lambda_{NQ}\Lambda_{w}]_{N=N^{opt}} < 0. \tag{76}$$

Using (69), we find

$$\Lambda_{\tau} = -\frac{w[1+Q^{\alpha}N^{\beta}]}{[\tau(1-s)+s]^2} + \frac{s\Phi(Q,N,w,\tau)}{\tau(1-s)+s}. \tag{77}$$

As $\beta/\alpha s(Q^*)^2/N = \Phi(Q^*,N,w,\tau)$, according to the definition of Q^* in (69), we can write

$$\Lambda_{\tau}(Q^*,N^{opt},w,\tau) = \frac{1}{\tau(1-s)+s}\left(\frac{\beta}{\alpha}\frac{(Q^*)^2}{N^{opt}} - \frac{w[1+(Q^*)^{\alpha}(N^{opt})^{\beta}]}{\tau(1-s)+s}\right). \tag{78}$$

Using (73), we find

$$\Lambda_{NQ} = \frac{1}{Q}\left(\frac{w\alpha\beta Q^{\alpha}N^{\beta-1} - (2-\alpha)(\tau Q^{2-\alpha}/2s\alpha)((1+\beta)N - \beta\delta^{\max}/(\delta^{\max}-N)^2N^{1+\beta})}{\tau(1-s)+s} + 2\frac{\beta}{\alpha s}\frac{Q^2}{N^2}\right), \tag{79}$$

$$\Lambda_{N\tau}(Q^*,N^{opt},w,\tau) = -\frac{w\beta(Q^*)^{\alpha}(N^{opt})^{\beta-1} + sZ(Q^*,N^{opt})}{[\tau(1-s)+s]^2}, \tag{80}$$

where $Z(Q,N) \equiv (Q^{2-\alpha}/2s\alpha)((1+\beta)N - \beta\delta^{\max}/(\delta^{\max}-N)^2N^{1+\beta})$. Recalling that $N^{opt} > \beta\delta^{\max}/(1+\beta)$, we have $Z(Q^*,N^{opt}) > 0$. Using (72) and (73), we can rewrite (79) to

$$\Lambda_{NQ}(Q^*,N^{opt},w,\tau) = \frac{1}{Q^*}\left(\frac{\beta}{s}\frac{(Q^*)^2}{(N^{opt})^2} - \frac{2(1-\alpha)\beta(Q^*)^{\alpha}(N^{opt})^{\beta-1}w}{\tau(1-s)+s}\right). \tag{81}$$

Substituting (71), (78), (80) and (81), into (75), rearranging terms and using (69) confirms $\partial N^{opt}/\partial\tau < 0$. Next, note from (69) and (73) that

$$\Lambda_w = \frac{1+Q^{\alpha}N^{\beta}}{\tau(1-s)+s} \quad \text{and} \quad \Lambda_{Nw} = \frac{\beta Q^{\alpha}N^{\beta-1}}{\tau(1-s)+s}, \tag{82}$$

respectively. Substituting (71), (81) and (82) into (76) and rearranging terms confirms $\partial N^{opt}/\partial w > 0$. This concludes the proof. □

Appendix B. Existence and uniqueness of equilibrium

Define the right-hand side of (37) as $P(Q,n)$ and note that, by this definition, the slope of the $F = 0$-locus in $Q-p$-space is given by $P_Q > 0$. To show that an equilibrium exists and is unique under weak conditions, consider the following case. Suppose that $B'(0) = 0$

[24] The notation indicates that Q^* as defined by (69) is a function of N, w, τ.

[25] According to (73), $\Lambda_N = 0$ can hold only if $N > \beta\delta^{\max}/(1+\beta)$. Using $N = \delta^{\max} - \frac{1}{/}2n > \beta\delta^{\max}/(1+\beta)$ implies $n^{opt} > (1+\beta)/2\delta^{\max} = \bar{n}$. This is consistent with Proposition 1(c), which suggests that $\Lambda_N > 0$ if $\varepsilon\left(1/2n, Q\right) \geq 1$. According to (23), $n > \bar{n}$ is equivalent to $\varepsilon(1/2n, Q) < 1$.

(thus, $P(0, n) = c$), $\lim_{Q\to\infty} B'(Q) \to \infty$, and $\lim_{C\to 0} u_C \to \infty$. Let $\tilde{Q}$ be given by

$$wg'\left(h\left(\frac{1}{2n}, \tilde{Q}\right)\right) = \frac{\tau n B'(\tilde{Q})}{s h_Q(1/2n, \tilde{Q})}. \tag{83}$$

Properties $g'' \le 0$ and $h_Q > 0$ imply that the left-hand side of (83) is non-increasing, whereas $h_{QQ} \le 0$ and $B'' > 0$ imply that the right-hand side of (83) is increasing. Thus, in view of the boundary conditions on B', an interior and unique level of $\tilde{Q}$ exists. Moreover, define

$$\tilde{p} \equiv \frac{wg(h(1/2n, \tilde{Q}))}{\tau(1-s)+s}. \tag{84}$$

Thus, at $(\tilde{Q}, \tilde{p})$ the consumption level is zero and $\lim_{C\to 0} u_C \to \infty$ implies $MRS(\tilde{Q}, \tilde{p}, \cdot) = 0$. Hence, $G(\tilde{Q}, \tilde{p}, \cdot) = 0$, according to (13) and (83). Consequently, if $\tilde{p} > P(\tilde{Q}, n)$, then at $Q = \tilde{Q}$ the $G = 0$-locus is above the $F = 0$-locus in $Q - p$-space. If, in addition, $G_Q < F_Q G_p / F_p$, the latter is steeper than the former. In this case, there is exactly one intersection point of function $P(Q, n)$ and the $G = 0$ locus, i.e., the equilibrium exists and is unique.

What are sufficient conditions for $\tilde{p} > P(\tilde{Q}, n)$ and $G_Q < F_Q G_p / F_p$ in terms of the primitives of the model? From $F_Q < 0$, $G_p < 0$ and $F_p > 0$ that $G_Q < F_Q G_p / F_p$ always holds if $G_Q \le 0$. A sufficient (but not necessary) condition for $G_Q \le 0$ is $MRS_Q \le 0$. For instance, $MRS_Q \le 0$ holds if g' is small or $|g''|$ is large, according to (14). ($MRS_Q < 0$ if $g' = 0$.) Moreover, note from (83) and (84) that $\tilde{Q}$ and $\tilde{p}$ do not depend on marginal cost c. Thus, using (37), we have $\tilde{p} > P(\tilde{Q}, n)$ if c is sufficiently small.

Appendix C. Two-stage decision

Suppose that, alternatively to the analysis in the main body of the paper, firms engage in a two-stage decision process. At stage 1, they choose the type of horizontal differentiation along with the vertical quality component. At stage 2, they choose prices (product market competition). There are two ways to analyze the model under a two-stage decision framework. First, firms foresee the Bertrand equilibrium for any vector of horizontal and vertical location of firms and take the related equilibrium responses into account at stage 1. Unfortunately, in this case, the analysis becomes intractable.[26]

The second way to analyze the two-stage problem is to assume that at stage 1 firms take prices of other firms as given (along with product quality and horizontal location) and therefore only foresee the impact of their choices on their price setting power for given prices of rivals. In this case, the behavior of firms is exactly the same as in the case where there is just one decision stage.

To see this, note that at stage 2 the optimal price of each firm fulfills first-order condition (26), which gives us the optimal price of firm i. Recalling that $\underline{\Delta}_i$ is a function of $p_i, Q_i, \delta_i(i_l), p_{i_l}, Q_{i_l}$ and $\overline{\Delta}_i$ is a function of $p_i, Q_i, D_i - \delta_i(i_l), p_{i_r}, Q_{i_r}$, we see that (26) gives us p_i implicitly as a function of $Q_i, \delta_i(i_l), p_{i_l}, p_{i_r}, Q_{i_l}, Q_{i_r}$. Write $p_i = \tilde{P}_i(Q_i, \delta_i(i_l), p_{i_l}, p_{i_r}, Q_{i_l}, Q_{i_r})$. Now, the optimization problem at stage 1 is:

$$\max_{\delta_i(i_l), Q_i} [\tilde{P}(Q_i, \delta_i(i_l), \cdot) - c] s [\underline{\Delta}_i(\tilde{P}(Q_i, \delta_i(i_l), \cdot), Q_i, \delta_i(i_l), \cdot) + \overline{\Delta}_i(\tilde{P}(Q_i, \delta_i(i_l), \cdot), Q_i, D_i - \delta_i(i_l), \cdot] - B(Q_i) - f, \tag{85}$$

where firms take as given $p_{i_l}, p_{i_r}, Q_{i_l}, Q_{i_r}, D_i$. The first-order condition with respect to vertical differentiation Q_i is:

$$0 = (p_i - c)s\left(\frac{\partial \underline{\Delta}_i}{\partial Q_i} + \frac{\partial \overline{\Delta}_i}{\partial Q_i}\right) - B'(Q_i) + \frac{\partial \tilde{P}_i}{\partial Q_i} s\left[\underline{\Delta}_i + \overline{\Delta}_i + (p_i - c)\left(\frac{\partial \underline{\Delta}_i}{\partial p_i} + \frac{\partial \overline{\Delta}_i}{\partial p_i}\right)\right]. \tag{86}$$

Applying the envelope theorem, the term is squared brackets of (86) becomes zero, according to stage 2 first-order condition (26). Thus, (86) coincides with first-order condition (25) of the profit maximization problem (11). An analogous argument holds for the first-order condition with respect to horizontal differentiation $\delta_i(i_l)$ associated with profit maximization problem (85); it coincides with (24). This confirms the claim.

[26] Lancaster (1979) focusses on simultaneous choices of horizontal location of firms and prices as well.

References

Acemoglu, D., Cutler, D., Finkelstein, A., Linn, J., 2006. Did medicare induce pharmaceutical innovation? American Economic Association Papers and Proceedings 96, 103–107.

Acemoglu, D., Linn, J., 2004. Market size in innovation: theory and evidence from the pharmaceutical industry. Quarterly Journal of Economics 119, 1049–1090.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and innovation: an inverted U relationship. Quarterly Journal of Economics 120, 701–728.

Aghion, P., Blundell, R., Griffith, R., Howitt, P., Prantl, S., 2009. The effects of entry on incumbent innovation and productivity. Review of Economics and Statistics 91, 20–32.

Aghion, P., Howitt, P., 2005. Growth with quality-improving innovations: an integrated framework. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. North-Holland, Amsterdam, pp. 67–110.

Aghion, P., Howitt, P., 2009. The Economics of Growth. MIT Press, Cambridge, MA.

Berndt, E.R., 2002. Pharmaceuticals in U.S. Health Care: determinants of quantity and price. Journal of Economic Perspectives 16, 45–66.

Blume-Kohout, M.E., Sood, N., 2013. Market size and innovation: effects of medicare part D on pharmaceuticals research and development. Journal of Public Economics 97, 327–336.

Blundell, R., Griffith, R., Van Reenen, J., 1999. Market share, market value and innovation in a panel of British manufacturing firms. Review of Economic Studies 66, 529–554.

Brekke, K.R., Straume, O.R., 2009. Pharmaceutical patents: incentives for research and development or marketing? Southern Economic Journal 76, 351–374.

Cai, L., Mavromaras, K., Oguzoglu, U., 2008. The Effects of Health and Health Shocks on Hours Worked, IZA Discussion Paper No. 3496.

Dasgupta, P., Stiglitz, J., 1980. Industrial structure and the nature of innovative activity. Economic Journal 90, 266–293.

Dixit, A., Stiglitz, J.E., 1977. Monopolistic competition and optimum product diversity. American Economic Review 67, 297–308.

Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The regulation of entry. Quarterly Journal of Economics 116, 1–37.

Economides, N., 1993. Quality variations in the circular model of variety-differentiated products. Regional Science and Urban Economics 23, 235–257.

Ethier, W.J., 1982. National and international returns to scale in the modern theory of international trade. American Economic Review 72, 389–405.

Gallini, N.T., 1992. Patent policy and costly imitation. RAND Journal of Economics 23, 52–63.

Garber, A.M., Jones, C.I., Romer, P., 2006. Insurance and incentives for medical innovation. Forum for Health Economics & Policy 9 (2), Article 4.

Gemmill, M.C., Thomson, S., Mossialos, E., 2008. What impact do prescription drug charges have on efficiency and equity. International Journal for Equity in Health 7, 12.

Giaccotto, C., Santerre, R.E., Vernon, J.A., 2005. Drug prices and research and development investment behavior in the pharmaceutical industry. Journal of Law and Economics 48, 195–214.

Gilbert, R., Shapiro, C., 1990. Optimal patent length and breadth. RAND Journal of Economics 21, 106–112.

Goldman, D.P., Joyce, G.F., Zheng, Y., 2007. Prescription drug cost sharing: associations with medication and medical utilization and spending and health. Journal of the American Medical Association 298, 61–69.

Grossman, G.M., Helpman, E., 1991. Quality ladders in the theory of growth. Review of Economic Studies 58, 43–61.

Grossmann, V., 2011. Do Cost-sharing and Entry Deregulation Curb Pharmaceutical Innovation?, CESifo Working Paper No. 3439.

Helpman, E., 1981. International trade in the presence of product differentiation, economies of scale and monopolistic competition. Journal of International Economics 11, 305–340.

Hummels, D., Lugovsky, V., 2009. Trade in ideal varieties: theory and evidence. Journal of Money, Credit and Banking 41, 3–33.

Jones, C.I., Williams, J.C., 2000. Too much of a good thing? The economics of investment in R&D. Journal of Economic Growth 5, 65–85.
Klemperer, P., 1990. How broad should the scope of patent protection be? RAND Journal of Economics 21, 113–130.
Lakdawalla, D.N., Sood, N., 2009. Innovation and the welfare effects of public drug insurance. Journal of Public Economics 93, 541–548.
Lakdawalla, D.N., Sood, N., 2013. Health insurance as a two-part pricing contract. Journal of Public Economics 102, 1–70.
Lakdawalla, D.N., Goldman, D.P., Michaud, P.-C., Sood, N., Lempert, R., Cong, Z., de Vries, H., Gutierrez, I., 2009. US pharmaceutical policy in a global marketplace. Health Affairs 28, 138–150.
Lancaster, K., 1979. Variety, Equity, and Efficiency. Columbia University Press, New York, NY.
Li, C.-W., 2001. On the policy implications of endogenous technological progress. Economic Journal 111, C164–C179.
O'Donoghue, T., 1997. A patentability requirement for sequential innovation. RAND Journal of Economics 29, 654–679.
O'Donoghue, T., Scotchmer, S., Thisse, J.-F., 1998. Patent breadth, patent life, and the pace of technological progress. Journal of Economics and Management Strategy 7, 1–32.
O'Donoghue, T., Zweimüller, J., 2004. Patents in a Model of Endogenous Growth. Journal of Economic Growth 9, 81–123.
OECD, 2010. Health at a Glance: Europe 2010. OECD, Paris.
Sanso, M., Aísa, R.M., 2006. Endogenous longevity, biological deterioration and economic growth. Journal of Health Economics 25, 555–578.
Scherer, F.M., 1993. Pricing, profits, and technological progress in the pharmaceutical industry. Journal of Economic Perspectives 7, 97–115.
Scotchmer, S., 1991. Standing on the shoulders of giants: cumulative research and patent law. Journal of Economic Perspectives 5, 29–41.
Scotchmer, S., 1996. Protecting early innovators: should second generation products be patentable? RAND Journal of Economics 27, 117–126.
Sood, N., de Vries, H., Gutierrez, I., Lakdawalla, D.N., Goldman, D.P., 2009. The effect of pharmaceutical regulation on revenues. Health Affairs 28, 125–137.
Sutton, J., 1991. Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration. MIT Press, Cambridge, MA.
Sutton, J., 1998. Technology and Market Structure. MIT Press, Cambridge, MA.
The Economist, 2011. Resistance to Antibiotics: The Spread of Superbugs., pp. 70–72, April 2nd.
Van Zona, A., Muysken, J., 2001. Health and endogenous growth. Journal of Health Economics 20, 169–185.
Vernon, J.A., 2005. Examining the link between price regulation and pharmaceutical R&D investment. Health Economics 14, 1–16.
Vincenzo, D., 1996. Patent races and optimal patent breadth and length. Journal of Industrial Economics 44, 249–265.
Wong, K.-y., 1995. International Trade in Goods and Factor Mobility. MIT Press, Cambridge, MA.
Wright, D.J., 1999. Optimal patent breadth and length with costly imitation. International Journal of Industrial Organization 17, 419–436.