Vierteljahrshefte
zur Wirtschaftsforschung
71 (2002), 4, S. 468–476

Can Insurance Cause Medical Care Spending to Grow too Rapidly?

By Mark V. Pauly*

Zusammenfassung: Im Mittelpunkt dieses Beitrags steht die Frage, ob wettbewerbsfähige Krankenversicherungsmärkte und wettbewerbsfähige Märkte für medizinische Leistungen zu effizienten Wachstumsraten von qualitätsverbessernden, aber kostensteigernden Technologien führen. Der Beitrag zeigt, dass es auf der einen Seite vielerlei Gründe gibt, warum diese Wachstumsraten unter dem Optimum liegen können. Andererseits ist es schwierig, ökonomische Modelle zu entwickeln, die ein übermäßiges Wachstum und die damit verbundenen Kosten gleichzeitig abbilden. Wenn Krankenkassen es ablehnen können, neue Leistungen zu versichern, werden neue Technologien, die zu einer Wohlfahrtsminderung führen, nicht eingeführt. Dies trifft auch dann zu, wenn die Krankenkassen durch gesetzliche oder administrative Regelungen dazu verpflichtet werden, die gleiche Selbstbeteiligung für alle von der Versicherung abgedeckten Leistungen anzubieten. Die Folge ist, dass der relativ hohe technologiebedingte Kostenanstieg in weniger regulierten Gesundheitssystemen (wie dem der USA) mit einer größeren Wirtschaftlichkeit verbunden sein kann als geringere Wachstumsraten in staatlich regulierten Gesundheitssystemen wie den europäischen.

Summary: This paper investigates the question of whether competitive markets in health insurance and in medical services will lead to efficient rates of growth in quality-improving but cost-increasing technology. It is shown that there are a number of reasons why the rate will be below the optimum, but that it is very difficult to develop models which display excess growth in technology and associated costs. As long as insurers are permitted to reject new services for coverage, new technologies which reduce welfare will not be adopted. This is true even if insurers are constrained by law or administrative to costs to provide the same coinsurance and deductibles for all covered services. The implication is that the relatively high technology driven cost increases in private health care systems (such in the United States) may represent greater efficiency than do the slower rates of growth in government-constrained systems.

1 Introduction

Reform of health insurance and health care markets is discussed frequently in many countries. While the term "reform" can obviously mean almost any kind of change in any direction, in recent years it has taken the flavor of increased emphasis on economic incentives in the payment or reimbursement process, usually associated with attempts to construct more market-like mechanisms for payment, along with some relaxation of government policies or rules to permit more competition among providers and health plans, and more citizen choice of both. "Consumer-driven health care" sounds good to many people, until they start talking to health economists.

* Health Care Systems Department, The Wharton School, University of Pennsylvania, USA, E-Mail: pauly@wharton.upenn.edu

Health economists (with some exceptions) have been uncomfortable with competitive medical care and medical insurance markets in the absence of fairly severe public regulation, and worried about such markets even with regulation. There have been two main sources of concern. One, more emphasized in Europe than in the United States, is that markets could lead to more "risk segmentation" (in some sense) among the citizenry, which is most troublesome in economists' amateur views of equity but also (as we know from the recent Nobel prize winners) can in theory cause problems with efficiency. The other potential fear is more primal and less grounded in any economic theory developed so far: a fear that markets (in contrast to government) can lead to a (worse) cost explosion.

My own view is that the risk segmentation problem has been overemphasized. Not all risk variation in premiums is unjust, and actual insurance markets provide (as far as we know) fairly effective protection against both adverse selection and long-term risk fluctuation through underwriting and guaranteed renewability. However, normative economic analysis of the cost containment problem, though the problem is documented generally in many ways and in many places, usually gets no farther than some grumbling about inefficiency in medical care and medical insurance. That is the topic I wish to discuss, first in theory, and then with evidence from recent U.S. experience.

2 Why Medical Spending Growth?

Some portion of the growth can in real medical spending in almost all developed countries be attributed to payments to suppliers of inputs to this sector (including profits to suppliers of equity capital) rising more rapidly than wages and prices in general, and some small part is in theory attributable to changing demographics, but the largest and most consistent influence on spending growth is what we usually call "technological change." In most statistical studies technological change is measured as a residual after the effects of prices and demographics are accounted for; it therefore represents technology in the literal sense of describing the application of larger quantities of productive inputs to whatever is the output received by the average consumer, but we really do not know what it is.

Technological change is not, of course, limited to the medical services sector of a modern economy; it is ubiquitous in virtually all sectors of the private economy, and has been responsible for the rising productivity and improving quality that describe growing market economies (Baumol 2002). However, public attitudes toward technological change often differ; while progress is praised in manufacturing, agriculture, and most of the services sector, technological change in medical services is thought to be a problem that must be both feared and controlled. To some extent, of course, this is because technological change has been associated with growing spending in this sector, in contrast to falling unit prices and falling relative expenditure shares in manufacturing and agriculture, and in some (though by no means all) service industries. It also occurs because the form of technological change has primarily been one of discovery of a method of treating the symptoms of some conditions that formerly could not be treated well (e.g., migraine headaches, erectile dysfunction) rather than discovery of ways to prevent a given illness altogether.

The key economic question, of course, is not whether technology increases or reduces spending. The key question is whether change is welfare increasing. In this paper I explore the question of the impact of technological change on welfare in a private economy in

which both insurance and medical services are purchased in competitive markets. I ask whether the outcome will be welfare increasing, and whether the outcome will be first best efficient. I find that outcomes are generally welfare increasing and that, if there is a deviation from the optimum, it is generally in the direction of suboptimal levels of both technical change and expenditure growth. While a full normative treatment would require a dynamic multigenerational model, I present here a simple model that finds that competitive markets (under appropriate legal structures) do not lead to excessively high levels of spending growth or technical change, given the instruments available to manage the use of care.

I outline a number of reasons for market failure in the choice of new technology; they all point in the direction of undersupply relative to first best optimum. It appears that welfare cannot be increased by limiting or curtailing the rate of technological change, and costs can never increase too rapidly. I hasten to add that this optimistic theoretical message does not fully dissipate my intuitive worries about medical spending growth, but I mean this paper as a stimulus to yet further effort to find a cause for excessive spending growth in medical markets, if there is such excessive growth.

3 The Basic Model

I assume that at any point in time t there exists a vector of medical products X_{it}, but that new products can be developed over time to add to this set. To develop product X_{it} requires incurring research costs R_{it-1} in the previous period. The product can then be produced at constant marginal cost C_{it}. The market price for the product is set at the profit maximizing price P^*_{it} where MR = C. Welfare is improved if consumers' surplus at P* exceeds R. The question then is whether all products for which this is the case will be brought to the market, or whether growth in technology will be suboptimal (too few products) or superoptimal (too many products).

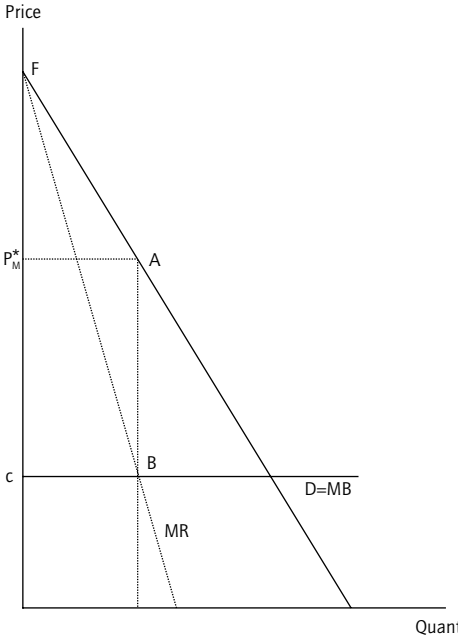
In the competitive market in the literature, the insurer is assumed to cover medical care spending at a uniform coinsurance rate c (Goddeeris 1991). Administrative complexities are invoked as a reason to assume a constant rate – more on this below. Ramsey and I argued that a key assumption was that the insurer had the power to choose whether or not to cover the new product at the fixed coinsurance rate c (Pauly and Ramsey 1999). When it has the freedom to cover or not, it will do so only if, on balance, its insureds have higher utility with coverage than without. All insureds have theoretical gains *ex ante*.

Some of the normative issues are illustrated in figure 1. MB is the expected market level demand or marginal benefit curve from the new product, and c is its marginal cost. If the product is produced by a monopolist, the equilibrium price will be P^*_m, and expected profits per period will be given by the area ABCD. If we convert research costs to an annual flow r_t for the same number of time periods, the product will be supplied if the expected profit equals or exceeds r_t. If so, consumers will obtain consumers' surplus in the amount FAD. It is therefore obviously efficiency-improving to invest in research on this product sold under monopoly; if any new product in this situation is brought to market, it will be efficiency-improving to do so.

I now ask how, compared to this benchmark, insured markets in various circumstances are likely to compare.

Figure 1

Optimal Insurance with No Moral Hazard



Source: Author.

4 Optimal Insurance With No Moral Hazard

If an insurer was contemplating covering this product, the best situation for consumers as consumers, conditional on monopoly pricing, would be to cover it, but to set the quantity at Q*. That is, the insurer and insureds would most prefer that there be no moral hazard, and to choose the quantity which (income effects aside) the consumer would buy if there were no insurance. Of course, the insurer would ordinarily prefer to pay a lower price for the product, but doing so might lead to profits that fell short of R^*_t. In such a case, the consumers' surplus from efficient new products would be lost. Hence, it is not necessarily desirable for insurers to gather buyer market power or governmental support to reduce payments to providers.

Competitive Product Markets with Information as a Public Good

Obviously the outcome will be inferior to this one if product markets are competitive, since short run profits will be zero and thus research costs will not be able to be covered. Competitive markets in new products tend to occur (or to appear more quickly) when the innovation can be copied, and not kept secret or prevented from use by a patent. However, the *optimal* degree of patent protection is not necessarily what prevails in the U.S. or in other countries. For the present, it is sufficient to note that technological change will be suboptimal in competitive markets because of the public good nature of research. Harris pointed this out many years ago, but policymakers interested in cost containment sometimes forget it.

Spillover Effects

Compared to the benchmark happy situation, there are several other reasons why an efficient new product might fail to be offered. An obvious problem for competitive insurers is that there may be spillover effects. Suppose, for example, that the marginal benefit from a new technology represents the avoidance of future illness (and the costs associated with future illness) for a preventive good or service. If there is turnover among health plan members, the plan may fail to capture the benefits from future cost offsets as members move to other plans. The insurer demand curve will then fall below the (social) marginal benefit curve; even under patent protection, some innovative goods and services may fail to be offered.

These spillovers will not occur in a competitive insurance model with perfect information and no regulation, since then the new plan will charge a lower premium to those with prospects of lower future expenses. Those prospective lower future premiums will motivate consumers to choose the plan providing protective care at a higher premium in the first period. It will be a more serious problem if insurers are not allowed to vary premiums with risk. However, even the utility value of the health benefits will be lost if the initial plan is unable to capture these benefits as higher premiums. So I conclude that it is very likely that some new technologies that yield net benefits over time will fail to emerge. In such a case, the rate of growth of medical spending will be too low.

5 Moral Hazard, Monopoly and Efficiency

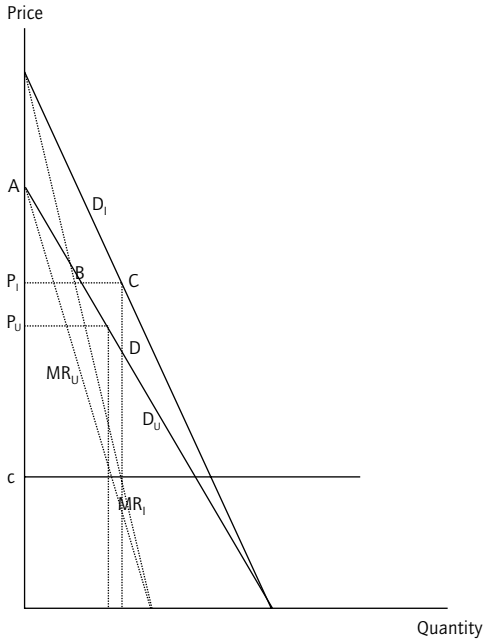
The first two cases are both reasonably obvious, and they lead to a conclusion that, under market arrangements, both improved health outcomes and medical spending may grow too slowly.

But what if supply-side limits cannot control moral hazard, so that both the price of a monopoly product and the quality are affected? What happens depends critically on the form of consumer cost sharing in insurance. Consider the simple case in which it takes the form of proportional coinsurance. A linear demand curve is pivoted about the intersection with the "quantity-axis" as shown in figure 2. Compared to either competitive pricing or control of moral hazard, in such a case provider profit is increased, but consumers' surplus from new products remains positive. In addition, expected profit levels are increased. The implication is that some products with positive consumers' surplus which failed to come to market with no insurance effects on the demand curve will now be placed on the market.

Compared to the monopoly outcome with no moral hazard, the rate of growth in medical care spending will be larger yet in this case, but surprisingly, the change in outcome could well be efficiency improving from the viewpoint of the economy as a whole. The reason is that any product that would have been introduced previously will still be introduced, but (paradoxically) the power to earn higher profits will cause more products with positive consumers' surplus to be brought to market. The efficiency evaluation relative to the competitive benchmark is ambiguous. The quantity is potentially higher than optimal for any given product, versus the monopoly situation in which the quality is lower than optimal. Even if supply-side limits would move the quantity closer to the optimum, there would still be a tradeoff compared to the monopoly case: less efficient use of any given product under

Figure 2

Optimal Insurance with Moral Hazard



Source: Author.

monopoly, but a more efficient supply of new products. Of course, were it possible to provide the optimal subsidy to research via the public sector, and to have optimal supply side limits under competitive pricing, that would be the first-best outcome.

What about the possibility that moral hazard at the uniform coinsurance rate could be so large as to offset any consumers' surplus? In the spirit of Gaynor et al. (2000), this is prevented by the assumption that insurers could and would refuse to cover new products which, at the use level determined by the coinsurance rate, make consumers worse off. An extension of this model could either allow for multiple endogenous coinsurance rates, or model the optimal choice of the coinsurance rate. In either case, the outcome would be an improvement in efficiency.

What is going on here? This is a specific application of the proposition that, in a decreasing cost industry, the only way to be certain that all efficiency gains are reaped is to have the product provided by a discriminating monopolist. The discriminating monopolist's revenues absorb the entire consumers' surplus and the monopolist will supply output (at the efficient level) if those revenues exceed total costs, both fixed and variable. In contrast, the non-discriminating monopolist may not be able to collect enough to cover total costs. Shifting and pivoting the demand curve with insurance moves the outcome closer to that of the discriminating monopolist, though it need not hit exactly the right point. So these gains are only possible, not certain.

6 Too Much Cost and Technology?

Up to this point, all of the results have pointed in the same direction: the rate of growth of spending in competitive insurance markets (conditional on the term of patent protection, if any) can never be higher than optimal and never be below the optimum. In this section, I want to consider three alternative theories that might suggest that growth could be excessive.

Medical Arms Race

A theory from the health economics literature on private health insurance in the U.S. from about 25 years ago suggested that insurance induced providers (especially hospitals) to compete in terms of consumer-perceived quality rather than price. Hospitals supposedly engaged in a medical arms race to attract patients and their doctors with technology which was costly and not very effective, but attractive because it signaled that care in general was modern and up-to-date.

It is generally believed that the cause of this behavior was a requirement by dominant insurers, often enforced by law, that all providers were entitled to be covered by any insurer, and patient cost sharing was low. Unable to compete on the basis of price, hospitals used availability of the latest technology as the lure for patients and physicians. The passage of laws which permitted private insurers to contract with some but not all hospitals reintroduced price competition and is said to have slowed the rate of growth of spending.

The conventional explanation attributes this change to the introduction of price competition, which is surely correct. But the change can also be viewed as one which permitted the insurer to deny coverage for care at institutions with certain costly technologies. In this sense, the cost-increasing properties of U.S. insurers were caused by coverage designs that were extremely restrictive.

Legal Limits

The key to insurance design for efficient rates of spending growth is the power to refuse to cover new technologies which they judge to provide insufficient benefits relative to their costs. U.S. insurers generally have this power, but in specific cases state (not national) legislatures have intervened to compel some insurers to cover improved technologies. Probably the most notorious example of this was in the case of autologous bone marrow transplantation (ABMT) for metastasized breast cancer, a costly and painful procedure that some cancer advocates alleged to be effective. Subsequent research proved it to be positively harmful.

If governments are permitted to override insurer judgments on coverage of new technology, or if the fear of governmental retaliation inhibits such limits, then spending growth can be excessive. While insurers should not be permitted to deny coverage arbitrarily, competition among plans makes such behavior unlikely.

Imperfect Information

The final case is probably the most relevant. If insurers and/or consumers incorrectly believe a new technology to be more beneficial or less costly than it actually is, and especially if increasingly large numbers of such overrated technologies are brought to market, spending and premium growth will be excessive.

In principle, research can provide improved (though never perfect) information on the effectiveness and cost of new technology. The problem, of course, is that such technology evaluation is a public good. Some of the larger HMOs in the U.S. do perform evaluations of new technology, but many do not.

In the absence of good information, would one expect insurance plans to cover more new technologies than otherwise? Ignorance could lead to adoption of different technologies than would occur under full information, but it is not obvious that the rate of growth would be different, or higher if it was different. If no one knows whether a new technology is beneficial, wouldn't an insurer refuse to cover it, keep its premiums down, and try to convince consumers (or the employers who in the U.S. choose insurance for workers) that the unproven new technology is worthless?

One would expect sellers of new technology to provide selective information that points in the direction of adoption, but insurers, aware of this motivation, would not adopt in response to it. It is therefore an empirical question of whether insurers can be systematically induced to offer coverage to high cost new technology.

7 Can a Heavy Hand Be a Better Hand?

The implication of the discussion so far is that the rate of growth of real costs chosen in a competitive insurance market, given the patent system, is unlikely to be excessive. The emphasis is primarily on the rate of addition of new products, but is it not also necessary to constrain the rate at which a newly-introduced product is used by the population? And might such heavy-handed but popular government control not devices, such as fixed budgets or reference prices, be lucky enough to limit that use and thus limit spending growth due to moral hazard?

If we limit competitive insurers to the use of coinsurance (at a constant rate for "approved" products and at a 100 percent rate for new "disapproved" products), the answer is surely positive. All I have shown is that technological change in a competitive market will not be at such a rate as to make consumers worse off than if the technology were never invented; I have not shown that there is not some alternative moral-hazard-limiting devices that could make consumers yet better off. And the administrative cost of discovering and labeling new technology (which often takes the form not of a wholly new device or service but rather newly discovered uses for existing technologies), and of offering varying coinsurance rates, will limit insurers ability to do the ideal.

In theory it is surely possible that there are some supply side limits on budgets or inputs which might bring about improvements. However, virtually all of these devices that governments can use can (legal system permitting) also be used by competing private insurers.

Fixed budgets, capitation payments, classification of technologies into preferred and unpreferred sets are all devices which private managed care plans can (and do in the United States) use to limit use. Without more, it is not obvious that there is any device that a government can use that a private insurer cannot, but the private market has the further advantage that consumers can choose among plans with different degrees of intensity of supply-side control (and associated premiums); the forced uniformity implied by public sector control can be avoided in private markets if consumers wish to do so. So, in theory, the observation that spending growth was contained by a government-imposed spending limit that consumers would not have wanted their health plan to impose implies that the public spending limit was too low.

8 Conclusion

The market can choose a rate of growth in spending that is too low. It is much less clear that it will choose a rate that is too high. People will complain about rising spending, but that spending is the price to be paid for technological process. It remains to be seen whether the ability of markets to limit spending growth to the optimal level or less will be credible in the political arena.

References

Baumol, W. J. (2002): Towards Microeconomics of Innovation: Growth Engine Hallmark of Market Economics. *Atlantic Economic Journal*, 30 (1), 1–12.

Gaynor, M., D. Haas-Wilson and W. B. Vogt (2000): Are Invisible Hands Good Hands? Moral Hazard, Competition, and the 2nd Best in Health Care Markets. *Journal of Political Economy*, 108 (5), 992–1005.

Goddeeris, J. H. (1984): Medical Insurance, Technological Change, and Welfare. *Economic Inquiry*, 22 (1), 56–67.

Pauly, M. V. and S. D. Ramsey (1999): Would You Like Suspenders to Go with That Belt? An Analysis of Optimal Combinations of Cost Sharing and Managed Care. *Journal of Health Economics*, 18 (4), 443–458.