Medicine Worse than the Malady: Cure Rates, Population Shifts, and Health Insurance*

Robert F. Graboyes**

Federal Reserve Bank of Richmond Working Paper No. 00-6

August 2000

JEL Nos. I11, G22, D81

Keywords: health insurance, technological change, heterogeneity, indemnities

Abstract

We examine the welfare effects of the interaction of three types of technological progress in medicine and health insurance; some paradoxes emerge. The model specifies three types of people: W (well); H (sick with high cure rate κ_H if treated); and L (sick with low cure rate κ_L if treated); they comprise proportions π_W, π_H, π_L of the population. There are four insurance modes: Indemnity (**I**): fully covered treatments for Hs, cash bribes for Ls to forgo treatment); Deductible (**D**): partially covered treatments for Hs, no treatments for Ls); Zero (**Z**): no insurance and no treatments); and Full (**F**): fully covered treatments for Hs and Ls). The three types of technological progress are represented as population shifts from sicker to healthier classes of people; for brevity, we call the shifts L→W, H→W, and L→H, and describe each as follows:"

L→W: Improved ability to prevent illness among Ls—π_L falls as π_W rises. L→W unambiguously improves welfare and seems to yield intuitive mode sequences.

H→W: Improved ability to prevent illness among Hs—π_H falls as π_W rises. H→W unambiguously improves welfare but sometimes yields surprising mode sequences. Examples: **F-Z** (full insurance when there are many Hs, no insurance when there are fewer Hs); and **D-F-D** (Hs partially covered, then fully covered, then only partially covered once again. Ls not treated, then treated, then not treated once again.).

L→H: Some would-be Ls become more highly treatable Hs—π_L declines as π_H rises. Here, technological progress not only yields surprising mode shifts (e.g., **D-Z-I-Z**), but the welfare effects of progress are ambiguous. This is because L→H may lead to more people being treated and cured (a welfare gain), but at a cost of higher premiums for all subscribers (a welfare loss).

The paradoxical results are in part explained by the fact that utility is a concave function of wealth and a linear function of health." L→W, H→W, or L→H could also be interpreted as autonomous demographic shifts rather than as technological progress.

---

* This paper does not necessarily represent the views of the Federal Reserve System or the Federal Reserve Bank of Richmond. This paper comprised a portion of a doctoral dissertation at Columbia University. The author thanks the following for their assistance in writing this and two companion articles: Sherry Glied, Ken Leonard, Dan O'Flaherty, Abigail Tay, and Josh Zivin and also Mark Babunovic, Dolores Clement, Doug Hadley, Bob Hurley, Jeff Lacker, Roice Luke, Lou Rossiter, Rich Schieken.

** Federal Reserve Bank of Richmond, Robert.Graboyes@rich.frb.org and Virginia Commonwealth University (Medical College of Virginia) rgraboyes@hsc.vcu.edu

1. Introduction

Can better-quality health care hurt? This paper models conditions under which shifts in population from sicker to healthier categories may lead to some surprising phenomena. In these models, welfare can decline as curative powers increase and outcomes improve.[1]

The model also examines how such population shifts may change optimal insurance contract parameters (premiums, deductibles, indemnities) or change the "mode" of the optimal insurance policy (modes are described below). In some possible mode sequences, increasingly healthy (or treatable) populations will induce insurers to drop some or all individuals from coverage. In some cases, the market undulates back and forth between covering and not covering (and therefore treating and not treating) different classes of individuals.

This paper follows from Graboyes (2000a) and parallels the work in Graboyes (2000b). Graboyes (2000a) examines the relative desirability of deductibles and indemnities as tools for deterring those with poor chances of cure from seeking expensive medical care. Graboyes (2000b) asks how welfare and the optimal insurance policy change as cure rates improve for some population groups. The current paper continues to explore the relative merits of indemnity and deductible contracts and, as in Graboyes (2000b), adds an element of technological progress in medical treatment. Here, we model technological progress as three types of population shifts from sicker to healthier categories. We find that such changes alter the optimal contract and welfare in some paradoxical ways.

[1] Graboyes (2000b) yields similar results, but only in cases where no one enjoys the fruits of the medical progress. Here, welfare can decline even as patients are actually benefiting from better outcomes.

Graboyes (2000a) assumes it is socially beneficial to treat “Hs” (patients with higher probability of cure) but not “Ls” (those with low probability of cure). The remainder of the population consists of “Ws”—people who are well. The current paper extends this analysis by examining how the optimal insurance contract changes with three types of improvement in medical technology. Graboyes (2000a) assumes that first-best welfare (expected utility across agents) occurs when insurance provides 100% coverage for Hs and 0% for Ls. However, first-best is infeasible because in the model, one’s H/L status is observable to all, but not legally verifiable. So, the only way to stop Ls from seeking treatment is to require them to bear a marginal cost (through deductibles paid or indemnities forgone) that exceeds the marginal benefit of treatment. In determining the optimal contract, the market can choose from among four feasible modes:

- **I:** An indemnity policy leading Hs to seek treatment (with 100% coverage) and Ls to forgo treatment in exchange for cash indemnities;[2]
- **D:** A deductible policy leading Hs to seek treatment (with less than 100% coverage) and deterring Ls from seeking treatment by charging deductibles;
- **Z:** Zero insurance for anyone; and
- **F:** Full insurance for Hs *and* Ls.

The current paper begins where Graboyes (2000a) leaves off. Here, as in Graboyes (2000b), we derive the relative desirability of all four modes. And we allow the efficacy of medical science to vary (represented as shifts in π_W, π_H, and π_L) and ask how the optimal contract and associate welfare change in response.

We look here at three types of improvement in medical technology: L$\rightarrow$W: Greater ability to prevent the disease among Ls (π_L declines, matched by an increase in π_W); H$\rightarrow$W: Greater ability to prevent the disease among Hs (π_H declines, matched by an increase in π_W); and L→H: Some would-be Ls become Hs (π_L declines, matched by an increase in π_H), who are more highly curable than Ls.

Like Graboyes (2000b), this paper also deals with technological progress and how it alters the optimal contract parameters and/or mode. The two types of technological change are different, however. Graboyes (2000b) is concerned with marginal improvements for a discrete class of people (Hs or Ls), whereas this paper, "Medicine Worse than the Malady," looks at discrete improvements for a marginal number of people. To repeat, in the former, an entire class of people become a little better off, while in the latter, a few people become a lot better off. The former can be thought of as a case in which the medical profession gradually improves its ability to treat the illness—a social learning curve, in other words. In the latter, there is a breakthrough that only affects a marginal number of patients.

In this paper, several paradoxical results emerge as some people shift from sicker to healthier states. At times, such shifts result in a decline in utility, even though health improves unambiguously. Other times, utility increases, even though health declines unambiguously. The paradox emerges from the fact that the utility gain from improved health can be more than offset by the utility loss from the increased cost of treatment. This, in turn, is related to the fact that utility is a concave function of wealth and a linear function of health.

There are real-life analogues for L→W, H→W, and L→H progress. L→W is perhaps the most counterintuitive of these three types of progress. This implies that a new treatment technique prevents illness only among the worst-off sufferers of the disease,

---

[2] Traditional fee-for-service policies are also referred to as "indemnification" policies. In contrast, this

but not those less seriously affected. A real-world example might be the effect of gastric bypasses on hypertension and its associated cardiovascular ills. The gastric bypass is a radical surgical procedure that reduces the appetite and reduces the absorption into the body of ingested foods. Excess body weight profoundly influences the likelihood of hypertension and, therefore, of serious cardiovascular illness. The heavier the individual, the less treatable is the hypertension. By significantly reducing body weight, the gastric bypass can sometimes deter the onset of hypertension. But, the catch is that the gastric bypass is only appropriate for those suffering from obesity and not those who are mildly overweight. Thus, the procedure prevents hypertension among Ls, but not among Hs. (This example is also appropriate to the model presented here in that whether obesity will induce hypertension in a particular individual depends on factors not generally known until the hypertension is diagnosed. In other words, adverse selection is minimized because individuals do not know beforehand whether their obesity will or will not induce hypertension. Other, less radical weight-loss techniques fall more into the H→W and L→H categories of technological progress.[3]

Within each of the four modes, there is an optimal contract (i.e., optimal set of parameters), and the optimal insurance contract is the best of these four optima. The following tables summarize the changes that occur within each of the four modes as we experience each of the three types of technological progress (i.e., population shifts):

---

paper uses "indemnity" contract to indicate a plan where patients are reimbursed in cash for a diagnosis, rather than for a treatment.

[3] Weight loss techniques as examples of the three types of progress were suggested by Dr. Richard Schieken, Chairman of the Department of Pediatric Cardiology at the Medical College of Virginia Campus of Virginia Commonwealth University.

Table 1: L® W

Preventive medicine improves for Ls: $p_L<0$ and $p_W=-p_L$

resulting direction of change under indemnity, deductible, zero, and full modes

	indemnity or deductible	**Premium**	**Utility**
I	increases	decreases	increases
D	no change	no change	increases
Z	—	—	increases
F	—	decreases	increases

L® W: Some Ls become Ws—progress we can interpret as improved preventive medicine whose benefits accrue only to Ls. The most important result visible in this table is that L→W always increases welfare. In other words, fewer sick people (Ls, in this case) is an unambiguous good. One implication is that L→W will always improve welfare, regardless of the sequence of modes through which the market passes. (We will discuss mode shifts later.)

Table 2: H® W

Preventive medicine improves: $p_H<0$ and $p_W=-p_H$

resulting direction of change under indemnity, deductible, zero, and full modes

	Indemnity or deductible	**Premium**	**Utility**
I	increases	decreases	increases
D	no change	no change	increases
Z	—	—	increases
F	—	decreases	increases

H® W: Some Hs become Ws—progress we can interpret as improved preventive medicine whose benefits accrue only to Hs. As with L→W, H→W always increases welfare. Once again, fewer sick people is an unambiguous good, no matter what sequence of modes the market passes through.

Table 3: L® H **Treatment improves: $p_L<0$ and $p_H=-p_L$** resulting direction of change under indemnity, deductible, zero, and full modes			
	Indemnity or deductible	**Premium**	**Utility**
I	increases: if $\hat{U}_{i*}>\hat{U}_z$ decreases: if $\hat{U}_z>\hat{U}_{i*}$	decreases: if $\hat{U}_{i*}>\hat{U}_z$ increases: if $\hat{U}_z>\hat{U}_{i*}$	ambiguous: if $\hat{U}_{i*}>\hat{U}_z$ increases: if $\hat{U}_z>\hat{U}_{i*}$
D	increases	decreases	ambiguous
Z	—	—	no change
F	—	no change	increases
Note: if $\hat{U}_z>\hat{U}_{i*}$, then changes under **I** are irrelevant because the market will select some mode other than **I**.			

L® H: Some Ls become Hs. We can interpret this as an improvement in curative medicine that allows some would-be Ls to become Hs instead. Alternatively, we could interpret this change as an improved ability to prevent some co-morbid condition that causes Hs to become Ls. The striking result in Table 3 is that the welfare effects of L→H are ambiguous. With zero insurance (**Z**), there is no change in utility, because no one is treated, and the number of sick people does not change. Under full coverage for Hs and Ls (**F**), welfare increases because more people are cured and expenditures are unchanged. But, under indemnity (**I**) or deductible (**D**) policies, L→H can either increase or decrease welfare. The reason is that as we treat more and more patients, wealth effects may eventually overtake health effects as marginal influences on utility. Utility is a linear function of health but an increasing function of wealth. Later, when we examine the impact of potential mode changes under L→H, we will find surprising mode sequences and welfare implications.

The structure of the paper is as follows: Section 2 discusses literature related to this paper.[4] Section 3 reviews the assumptions, notation, and results from Graboyes (2000a);

[4] The literature review here is identical to that used in Graboyes (2000b). The two papers are to be published as separate working papers and this literature review is appropriate to both.

this review serves to set up the problem addressed in the current paper; this section also briefly describes the results of Graboyes (2000b). Sections 4, 5, and 6 examine how insurance contract parameters, welfare, and modes change in response to L→W, H→W, and L→H, respectively. Section 7 presents the conclusions and suggestions for further research. Most mathematical proofs are in the Appendix.

2. Related Literature

Moral hazard, which is central to this paper, has long been linked with efficiency in health care production and with the direction of technological progress in medicine. Zeckhauser (1970) described why moral hazard is an inevitable by-product of health insurance contracts that spread risks and why moral hazard creates disincentives for efficient production. In creating the optimal health insurance policy, he wrote, “The best that can be done, as we would suspect, is to find a happy compromise with some risk-spreading and some incentive.”

Feldstein (1973) refined the notion that moral hazard-induced inefficiencies would lead to overspending on health care itself. In doing so, he estimated the level of patient copayment (deductible) that would achieve the equivalent of Zeckhauser's happy medium between risk-sharing and efficiency.

To this framework, Goddeeris (1984) added technological innovation and found that under the right circumstances, scientific progress could reduce welfare. His paper addressed the ways in which insurance could bias the direction of technological progress (research and development, technology diffusion). Baumgardner (1991) carried this farther by examining the relationships between technical change, welfare, and optimal class of insurance contract ("mode" in this paper), with a focus on asymmetric information and imperfect agency. He contrasted how these relationships would appear under conventional (fee-for-service) insurance policies and under managed care policies. A similar comparison of demand-side and supply-side incentives is the theme of Ellis and McGuire (1993) who link technological progress to increasing medical expenditures in the United States. They ask how supply-side incentives might hold down the rate of technological progress and, therefore, of overall costs—with an implicit assumption that progress is cost-increasing. Cutler and Sheiner (1997) similarly ask how managed care might hold down the rate of technological progress and, therefore, costs.

In many of these papers, an explicit or implicit idea is that of the cost-increasing technological imperative. That is, writers assume or discover the validity of the technological imperative. This concept, described by Pauly (1986, p. 664) holds that a health care technology, once it exists, tends naturally and unstoppably to diffuse throughout the economy. Sometimes, the technology is used well beyond its optimal level of provision. A corollary to many of these findings is that moral hazard and imperfect agency bias technological progress toward cost-increasing innovations. Many writers blame the development and diffusion of such technologies for the rapid rise in health care expenditures in the United States since the 1960s. Cutler (1996, p. 35) writes that, "the medical care marketplace is driven by overuse of medical resources, and the rapid development and diffusion of new technologies.

The current paper looks entirely at demand-side incentives for efficiency. Whereas Zeckhauser and Feldstein postulate deductible policies, we compare deductible policies with indemnity policies (and with full insurance policies and with zero insurance).

Indemnity policies have played a small role in health insurance in recent years decades, but they were mentioned by Arrow (1963, p. 962), and their renewed use has been suggested by Gianfrancesco (1983) and by Feigenbaum (1992).

In some ways, the model developed here goes in the opposite direction from those of Goddeeris and Baumgardner. Goddeeris stresses the influence of insurance on the directions of technological change. Here, we look in the opposite direction, focusing on the effect of autonomous technological change on health insurance. Baumgardner compares traditional insurance with managed care. Here, we look only at different demand-side mechanisms.

The conclusion discusses how our results might link back to affect the rate and direction of progress. Mutual determination of technological progress, health care provision, and insurance contracts was described by Weisbrod (1991). We identify some conditions under which insurance might *impede* rather than encourage technological advance. This might happen because in the model presented here, certain technological advances will eventually lead to abandonment of health insurance altogether and, along with it, usage of the previously insured treatments. So, it is natural to suppose that forward-looking investors might shy away from investing in technologies that will eventually be dropped from coverage and usage. We might find that insurance biases progress toward treatments which do not appear likely to be abandoned due to changes in the primitive assumptions listed in the notation section below.

The model developed here bears some resemblance to the market for lemons postulated by Akerlof (1970) in that the market is characterized by bimodal heterogeneity. Here, though, information is symmetric; rather, it is the ability to *respond* to information that is asymmetric.

The current paper more closely resembles the literature on “tagging.” In Akerlof (1978), the goal is to construct the optimal feasible redistribution of wealth; lump-sum welfare payments are made to “deserving” people, financed by general taxation. The model optimizes by restricting welfare payments to a small group, defined (“tagged”) by a variable that acts as a proxy for “deserving.” (Race-based set-asides are an example.) By limiting the number of eligible recipients, tagging allows high welfare payments to be paid to the deserving group, financed by the rest of the population at low marginal tax rates. In the current paper, high-benefit patients are “deserving” while low-benefit patients are part of a “non-deserving” population that also includes well people. Here, indemnities and deductibles induce people to tag themselves. The distribution of population between high- and low-benefit patients determines whether a high-lump sum benefit (the treatment) can be financed through a low marginal “tax” rate (namely, the insurance premium.)

Finally, this paper bears some resemblance to the “loyalty” or “shirking” literature, as in Akerlof (1983). In that literature people must decide whether or not to shirk on the job; people who shirk run some probability of being fired. One way to generate honesty is to require workers to post a surety bond that is forfeited if fired. In the current paper, deductibles and indemnities serve essentially the same purpose. Paying a deductible or forfeiting an indemnity serves as the bond guaranteeing the value for treatment that the patient claims. In discussing “loyalty filters,” Akerlof (1983) describes how experience changes one's loyalty, in turn, affecting one's economic strategies. Here, it is the quality of medical practice rather than personal experience that changes behavior, but with similar results.

3. Setup: Review of Graboyes (2000a) and (2000b)

The setup for this paper comes from the assumptions and results of Graboyes (2000a). That paper asks when lump-sum indemnities are more efficient than deductibles at deterring Ls from seeking expensive treatment. This review serves as the point of departure of the current paper. We also review the results of Graboyes (2000b), which parallels the current paper.

3.1 Assumptions

Graboyes (2000a) begins with the following assumptions:

(1) *Ex post* utility is a state-dependent Von Neumann-Morgenstern function where $U(y; w)=U(y; s)+k$, with $U_y>0$ and $U_{yy}<0$. y is *ex post* monetary wealth, w and s are the two values of a binary variable representing well and sick states, and k is a constant denoting the difference in utility between the two states for any y. This functional form means that utility is state-dependent, but marginal utility is not.

(2) The insurance policy protects against a single illness. It is a carveout—similar to a dread disease policy, although dread disease policies' benefits are often contingent upon a hospital stay or other medical service.

(3) Adverse selection is not an issue. All agents are equally likely to contract the illness. That probability is known both to subscribers and insurers.

(4) There is no *ex ante* moral hazard; the presence or lack of insurance does not influence the behavior of insured parties *before* they contract the illness or, hence, the incidence of disease.

(5) Diagnosis is binary and unambiguous and requires no costly monitoring.

(6) Sick people are classified as Hs or Ls, based on their probability of cure if treated. An individual's likelihood of cure, κ_H or κ_L, is costlessly observable by both the patient and the insurer. However, the prognosis is not legally verifiable, so patients can act on the information, but insurers cannot. The insurer cannot, for example, promise to pay for chemotherapy if the probability of cure is 5%, but not if it is 1%, though the patient may accept or decline treatment on the same basis. This is because patients cannot bind themselves to forgo treatment if they are Ls.

(7) There are no loading costs or other fixed costs.

(8) The cost of treatment is large enough that no one can purchase it without insurance. In other words, there is no borrowing or capital market.

3.2 Notation

Both Graboyes (2000a) and the current paper use the following notation:

Initial conditions: These parameters define the state of the world:

π_W	percent of subscribers who are well
π_H	percent of subscribers who are sick and will experience a high cure rate if treated
π_L	percent of subscribers who are sick and will experience a low cure rate if treated
π_S	percent of subscribers who are sick: $\pi_H + \pi_L$
κ_H	the cure rate for Hs
κ_L	the cure rate for Ls
y_0	initial wealth of all agents
x	the cost of treatment
k	the welfare loss of having the disease; it is completely reversed if cured

Contract parameters (indemnities, deductibles, premiums) and *ex post* wealth:

i	A cash indemnity large enough to deter Ls from seeking treatment
i*	The minimum cash indemnity large enough to deter Ls from seeking treatment
d	A deductible large enough to deter Ls from seeking treatment
d*	The minimum deductible large enough to deter Ls from seeking treatment
p_{i*}	The insurance premium paid by all subscribers under the indemnity contract
p_{d*}	The insurance premium paid by all subscribers under the deductible contract
p_f	The insurance premium paid by all subscribers under the full-insurance contract
y	*ex post* wealth; y_0 minus premiums and deductibles paid or indemnities received

Welfare under different modes: Mode **H** is infeasible because insurers cannot be legally bound to refuse treatment if they are found to be Ls. **I**, **D**, **Z**, and **F** are feasible:

$\hat{U}_h$	Mode **H**: Hs 100% covered, Ls not treated; this mode is infeasible.
$\hat{U}_i$	Suboptimal indemnity; deters Ls, but not Hs, from seeking treatment.
$\hat{U}_{i*}$	Mode **I**: Optimal indemnity; deters Ls, but not Hs, from seeking treatment.
$\hat{U}_d$	Suboptimal deductible; deters Ls, but not Hs, from seeking treatment.
$\hat{U}_{d*}$	Mode **D**: Optimal deductible; deters Ls, but not Hs, from seeking treatment.
$\hat{U}_z$	Mode **Z**: Zero insurance; neither Hs and Ls are treated
$\hat{U}_f$	Mode **F**: Full insurance; treatment for Hs and Ls 100% covered
$\hat{U}$	MAX[$\hat{U}_{i*}$, $\hat{U}_{d*}$, $\hat{U}_z$, $\hat{U}_f$]; the optimal policy across all modes
U(-;w)	State-dependent utility function in well state
U(-;s)	State-dependent utility function in sick state

3.3 Results: General Case

The above assumptions yield the following results:

(1) $\hat{U}_{i*} > \hat{U}_i \; \forall \; i > i*$ The minimum deterrent indemnity is the optimal indemnity. [ch.2, (P.2)]

(2) $\hat{U}_{d*} > \hat{U}_d \; \forall \; d > d*$ The minimum deterrent deductible is the optimal deductible. [ch.2, (P.4)]

(3) $i* \overset{>}{\underset{<}{=}} x \Rightarrow \hat{U}_z \overset{>}{\underset{<}{=}} \hat{U}_{i*}$ (The desirability of **I** versus **Z** depends on the relative size of i* and x.) [ch.2, (P.17)]

(4) $\hat{U}_h = U(y_0 - p_H; w) - (\pi_S - \pi_H \kappa_H)k$, where $p_H = \pi_H x$

Equation [4] shows the unattainable utility that would prevail if Ls could be costlessly deterred from receiving treatment.

(5) $$\begin{aligned} \hat{U}_{i*} &= (1-\pi_L)U(y_0 - p_{i*}; w) + \pi_L U(y_0 - p_{i*} + i*; w) - (\pi_S - \pi_H \kappa_H)k \\ &= U(y_0 - p_{i*}; w) - (\pi_S - \pi_H \kappa_H - \pi_L \kappa_L)k \end{aligned}$$,where

$p_{i*} = \pi_H x + \pi_L i*$ [ch.2, (1.5)]

(6) $$\begin{aligned} \hat{U}_{d*} &= (1-\pi_H)U(y_0 - p_{d*}; w) + \pi_H U(y_0 - p_{d*} - d*; w) - (\pi_S - \pi_H \kappa_H)k \\ &= U(y_0 - p_{d*}; w) - (\pi_S - \pi_H \kappa_H + \pi_H \kappa_L)k \end{aligned}$$, where

$p_{d*} = \pi_H x - \pi_H d*$ [ch.2, (3.5)]

(7) $\hat{U}_z = U(y_0; w) - \pi_S k$

Equation [7] shows the utility prevailing if no insurance exists and no one is treated.

(8) $\hat{U}_f = U(y_0 - p_f; w) - (\pi_S - \pi_H \kappa_H - \pi_L \kappa_L)k$, where $p_f = p_S x$

Equation [8] shows the utility prevailing if everyone is insured and treated.

(9) $\hat{U}_{d*} \gtreqless \hat{U}_{i*}$ iff $U(y_0 - \pi_H x + \pi_H d^*; w) - U(y_0 - \pi_H x - \pi_L i^*; w) \gtreqless \pi_S \kappa_L k]$ [ch.2, (P.5)]

3.4 Results: Logarithmic Specification

We obtain stronger results by restricting the utility function to a logarithmic specification, where U(y; w)=ln(y) and U(y; s)=ln(y)-k.[5] In results (4)-(9), U(·) can be replaced by ln(·). The logarithmic specification also yields the following results:

(10) $i^* = \dfrac{(f-1)(y_0 - p_H x)}{1+(f-1)p_L}$, where $f = e^{k_L k}$ (the optimal indemnity) [ch.2, (P.6)]

(11) $d^* = \dfrac{(f-1)(y_0 - p_H x)}{f-(f-1)p_H}$, where $f = e^{k_L k}$ (the optimal deductible) [ch.2, (P.7)]

(12) i*>d* [ch.2, (P.16)]

(13) If $\pi_L \geq \pi_H$, then $U_{d*} > U_{i*}$ under all circumstances [ch.2, (P.9)]

(14) $\hat{U}_{d*} \gtreqless \hat{U}_{i*}$ iff $\dfrac{1+(\phi-1)\pi_L}{[\phi-(\phi-1)\pi_H]\phi^{\pi_H+\pi_L-1}} \gtreqless 1$ [ch.2, (P.8)]

This is the boundary condition that determines the preference ordering between **D** and **I**.

(15) In the limit, as $\kappa_L \to 0$, $\hat{U}_{i*} \gtreqless \hat{U}_{d*}$ iff $\pi_H \gtreqless \pi_L$. [ch.2, (P.12)]

3.5 Extensions in the Current Paper

The current paper extends the results of Graboyes (2000a) by varying the proportions of the population falling into categories L, H, and W (either because of technological or autonomous demographic changes). As in results (10) through (15) above, we limit the analysis to a logarithmic utility function. Graboyes (2000a) established criteria for choosing between modes **I, D,** and **Z**, holding π_W, π_H, π_L, κ_L, and κ_H constant (assuming that **Z** is always preferable to **F**). Graboyes (2000b) examined how the relative desirability of modes **I, D, Z**, and **F** changes as cure rates κ_L and κ_H change; once again, π_W, π_H, π_L were fixed. The current paper, in contrast, holds the cure rates κ_L and κ_H constant but varies the population distributions π_W, π_H, and π_L.

In Graboyes (2000b), an improvement in medical science represented by an increase in κ_L can have perverse effects. If κ_L rises, but not enough to warrant treating Ls, then no one actually receives the benefits of this medical progress. However, we now have to pay Ls a higher bribe to induce them to reveal the low value they place on treatment. This carries us farther from the ideal of equal *ex post* monetary wealth across agents and, thus, reduces welfare.

In contrast, the current paper stipulates conditions under which technological progress may in fact lead to improved health but, under certain conditions, will still have perverse welfare effects. Here, it may be that we treat and cure an increasing number of people, but the marginal utility cost of financing these treatments rises until it exceeds the marginal utility benefit from improved health.

4. Effects of Reduced Number of Ls (p_L): L® W Technological Progress

This section describes changes in contract parameters and in welfare resulting from a change in π_L under each of the four feasible insurance modes. We implicitly hold π_H

[5] This specification is similar to that used in Neipp and Zeckhauser (1985).

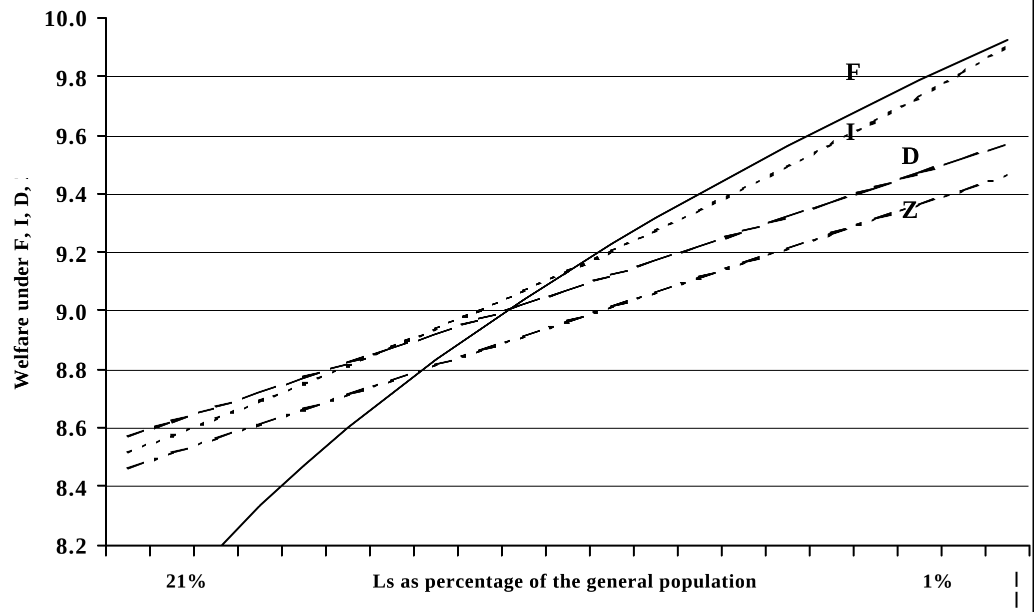
constant so that a change in π_L is offset by an equal and opposite change in π_W. We thus refer to this as L→W technological progress.

L→W might occur when a single illness has two different etiologies—one that causes a person to become an L and another that causes the person to become an H. The technological progress here is that medical science finds a way to reduce the incidence of L-causing factors but not of H-causing factors. For example, some kinds of lung cancer are more likely to respond to treatment than others. In particular, lung cancer associated with smoking is less curable than some other types. Hence reduced smoking can be said to be a technological change that qualifies as L→W.[6] Earlier in the paper, we mentioned that the gastric bypass has a similar relationship to hypertension in that the procedure is only appropriate for the most obese patients.

Figure 1 applies arbitrary parameters to the model, enabling us to examine welfare under each of the four feasible insurance contract modes as a function of π_L.[7] Since progress is measured by the *reduction* in the number of Ls, π_L *declines* from 21% on the left to 1% on the right. As indicated by Table 1 (and by Table 4 below), welfare climbs steadily under all four modes and, hence, under the optimal contract, which shifts along the way from a deductible contract to an indemnity contract, and then to full insurance for both Hs and Ls.

[6] Smoking and lung cancer as an example of L→W progress was suggested by my advisor, Sherry Glied, Columbia University.
[7] The parameters here are κ_L=40%, κ_L=70%, k=5, y_0=\$100,000, x=\$150,000, π_H=40%, and $\pi_L \in$ (1%,21%).

Figure 1
Welfare, 4 Contracts, # of Ls varies



Now, referring to Table 4, we explore how welfare, contract parameters, and the optimal mode change under general conditions. Table 4 shows the partial derivatives of contract parameters and of welfare with respect to $-\pi_L$. The paragraphs immediately following explain the intuition behind the signs of the entries in Table 4.

Table 4
Contract parameter and welfare changes under L® W technological progress
(preventive medicine improves: $\Delta\pi_L<0$ and $\Delta\pi_W=-\Delta\pi_L$)

	Indemnity (i*) or Deductible (d*)	**Premium (p_{i*}, p_{d*}, p_f)**	**Utility ($\hat{U}_{i*}$, $\hat{U}_{d*}$, $\hat{U}_z$, $\hat{U}_f$)**
I	**1:** $\frac{i^{*2}}{y_0 - p_H x} = \frac{(f-1)^2(y_0 - p_H x)}{[1+(f-1)p_L]^2} > 0$	**2:** $-\frac{i^*}{1+(f-1)p_L} = -\frac{(f-1)(y_0 - p_H x)}{[1+(f-1)p_L]^2} = -i^*\left[\frac{y_0 - p_H x - p_L i^*}{y_0 - p_H x}\right] < 0$	**3:** $\frac{i^*}{y_0 - p_H x} + (1-k_L)k > 0$
D	**4:** 0	**5:** 0	**6:** k>0
Z	—	—	**7:** k>0
F	—	**8:** -x<0	**9:** $\frac{x}{y_0 - p_S x} + (1-k_L)k > 0$

Numbers 1 through 9 correspond with Propositions 1 through 9 in the Appendix.

4.1 L® W Technological Progress: No mode changes

The entries in Table 4 consist of partial derivatives of the optimal indemnities, deductibles, premiums, and utility functions with respect to $-\pi_L$. The table ignores mode changes, which are discussed in Section 4.2.

Indemnity mode (I): L→W raises the optimal indemnity, reduces the optimal premium, and raises welfare. These changes occur via the following linkages: (1) L→W means fewer Ls requiring indemnities; (2) Fewer Ls requiring indemnities means lower premiums; (3) Lower premiums mean greater post-premium wealth, thus increasing the willingness to pay for treatment (i.e., the optimal indemnity). (4) The rising indemnity partially, but not completely, reverses the initial decline in the premium. The reversal is only partial because the number of Ls declines faster than the indemnity rises. (5) In the end, all subscribers pay lower premiums, there are fewer Ls suffering illness, and the remaining Ls receive higher indemnities. Hence all agents are better off than before, so welfare unambiguously rises.

Deductible mode (D): L→W leaves the optimal deductible and premium unchanged and raises welfare, as follows: (1) If the deductible does not change, then Ls' post-premium wealth does not change. (2) If Ls' post-premium wealth is unchanged, then Ls' willingness to pay (and thus the optimal deductible) is unchanged. (3) If the number of Hs is unchanged and the premium is unchanged, then the treatment costs incorporated in the premium are unchanged. (4) Hence, welfare is unchanged for Ws, Hs, and those who remain Ls after L→W occurs. (5) However, those whom progress changes from Ls to Ws are better off because they no longer experience illness. Hence, higher welfare is consistent with unchanged deductible and premium.

Zero-insurance mode (Z): L→W always causes welfare to rise. With no indemnities, deductibles, or premiums, *ex post* wealth remains constant for all agents. Since fewer Ls are ill, welfare increases unambiguously.

Full-insurance mode (F): L→W causes premiums to decline and welfare to rise. Premiums decline because there are fewer sick people whose treatments must be covered by insurance, so all agents' post-premium wealth rises. In addition, there are fewer Ls and, thus, fewer sick people. So, welfare rises through health and wealth improvements.

4.2 L® W Technological Progress: Mode changes

Here, we look at some ways in which the contract mode can change in response to a change in π_L. In Figure 1, the mode sequence is **D**-**I**-**F**. Mode shifts always occur where two welfare functions intersect, and always toward the mode whose utility function has the steeper slope at that point. Knowing this lets us rule out certain shifts. For instance, Table 4 shows that a decline in π_L will never prompt a mode shift from **D** to **Z** or from **Z** to **D**, because the slopes of the relevant curves are always equal. Similarly, we can rule out a shift from **F** to **I** because the slope of the **F**-curve is always greater than the slope of the **I**-curve.

Table 5 shows some possible mode sequences under different sets of arbitrary parameters; no particularly unusual sequences are apparent. While we have not rigorously demonstrated all the possible and impossible sequences, tests over a wide range of parameters failed to indicate any sequences other than those shown here. More unusual mode sequences will appear when we explore the implications of H→W and L→H technological progress.

Table 5
Some mode shifts under L® W technological progress
(preventive medicine improves: $\Delta\pi_L<0$ and $\Delta\pi_W=-\Delta\pi_L$)

contract sequence	κ_L	κ_H	k	y_0	x	p_L	π_H
Z	30%	90%	2	$ 100,000	$ 200,000	**0%-20%**	10%
I	30%	90%	4	$ 100,000	$ 150,000	**0%-20%**	40%
F	30%	90%	20	$ 100,000	$ 150,000	**0%-20%**	10%
D-I	3%	90%	10	$ 100,000	$ 150,000	**0%-20%**	5%
Z-I	37%	70%	4	$ 100,000	$ 150,000	**0%-20%**	40%
D-F	17%	90%	10	$ 100,000	$ 150,000	**0%-20%**	5%
Z-F	40%	60%	4	$ 100,000	$ 150,000	**0%-20%**	40%
I-F	30%	90%	5	$ 100,000	$ 150,000	**0%-10%**	30%
D-I-F	30%	90%	5	$ 100,000	$ 150,000	**0%-20%**	30%
Z-I-F	40%	70%	4	$ 100,000	$ 150,000	**0%-20%**	40%

5. Effects of Reduced Number of Hs (p_L): H ® W Technological Progress

This section derives changes in contract parameters and in welfare resulting from a change in π_H under each of the four feasible insurance modes. We implicitly hold π_L constant, so a change in π_H is offset by an equal and opposite change in π_W. As in Section 3, the logic here might be that the illness has two different etiologies—one that causes a person to become an L and another that causes the person to become an H. In this section, we assume that medical science has found a way to reduce the incidence of H-causing factors but not L-causing factors.

5.1 H ® W Technological Progress: No mode changes

The entries in Table 6 consist of partial derivatives of the optimal indemnities, deductibles, premiums, and utility functions with respect to $-\pi_H$.[8] The table ignores mode changes, which are discussed in Section 5.2. The following paragraphs explain the intuition behind the signs of the expressions in Table 6.

[8] Table 2, shown earlier in the paper, derives its information from Table 6. The expressions in Table 6 are derived in the Appendix, propositions 10 to 18.

Table 6
Contract parameter and welfare changes under H® W technological progress
(preventive medicine improves: $\Delta\pi_H<0$ and $\Delta\pi_W=-\Delta\pi_H$)

	i*/d*	p	Û
I	**10:** $\frac{xi^*}{y_0 - p_H x} = \frac{(f-1)x}{1+(f-1)p_L} > 0$	**11:** $-\frac{x}{1+(f-1)p_L} = -x\left[\frac{y_0 - p_H x - p_L i^*}{y_0 - p_H x}\right] < 0$	**12:** $\frac{x}{y_0 - p_H x} - k_H k + k > 0$
D	**13:** $\frac{d^*}{y_0 - p_H x}(x - d^*) > 0$	**14:** $-(x-d^*)\left[\frac{y_0 - p_H x + p_H d^*}{y_0 - p_H x}\right] < 0$	**15:** $\frac{x-d^*}{y_0 - p_H x} - k_H k + k_L k + k > 0$
Z	—	—	**16:** k>0
F	—	**17:** -x<0	**18:** $\frac{x}{y_0 - p_S x} - k_H k + k > 0$

Numbers 10 through 18 in cells correspond to Propositions 10 through 18 in the Appendix.

Indemnity mode (I): H→W raises the optimal indemnity, reduces the optimal premium, and raises welfare. The logic is found in the following linkages: (1) Technological progress means fewer Hs requiring treatment; (2) Fewer Hs requiring treatment means lower premiums needed to pay for treatments; (3) Lower premiums mean greater post-premium wealth, thus increasing the willingness to pay for treatment. Dissuading Ls from seeking treatment thus requires higher indemnities. (4) The rising indemnity partially, but not completely, reverses the initial decline in the premium, because the premium must also cover the higher indemnities. The reversal is only partial because treatment costs for Hs decline faster than indemnity costs for Ls rise. (5) In the end, all subscribers pay a lower premium, Ls receive higher indemnities, and fewer Hs suffer the welfare losses from sickness. Hence welfare unambiguously rises with a decline in π_H.

Deductible mode (D): H→W raises the optimal deductible, reduces the optimal premium, and increases welfare. The logic is as follows: (1) Fewer Hs means lower treatment costs and lower premiums needed to cover those treatment costs. (2) Lower premiums increase *ex post* wealth for all subscribers, thus raising the willingness to pay for treatments by Hs and Ls. (3) To prevent Ls from seeking treatment, the deductible must rise, further reducing the premium. (3) The process iterates until the deductible just induces Ls to forgo treatment. (4) Welfare rises because all subscribers pay lower premiums, some would-be Hs never get sick (and thus never pay deductibles or experience a failed cure attempt). Hs who do get sick pay a higher deductible than before, but that negative wealth effect is too small to offset the factors raising expected welfare.

Zero-insurance mode (Z): H→W causes welfare to rise. There are no indemnities, deductibles, or premiums, so *ex post* wealth remains constant. The welfare gain comes from the fact that there are fewer Hs to contract the illness.

Full-insurance mode (F): H→W causes premiums to decline and welfare to rise. Premiums decline for all subscribers because there fewer sick people need treatments, paid for by insurance. In addition, there are fewer Hs and, thus, fewer Hs who are not, in the end, cured. Hence, welfare unambiguously rises with a decline in π_H.

5.2 H® W Technological Progress: Mode changes

Here, we look at some ways in which the contract mode can change in response to a change in π_H. Table 7 shows some possible mode sequences. Here, we see some more unusual possibilities than those that resulted from a change in π_L. Because of nonlinearities, for example, sequences can include either **F**-**Z** or **Z**-**F**. Following are descriptions of several of the more unusual sequences.

I-**F**-**D** is a feasible sequence. Insurance initially bribes Ls (by means of an indemnity) to forgo treatment. After some H→W, insurance begins to fully cover treatments for both Hs and Ls. Finally, additional H→W leads the insurer once again to exclude Ls from treatment, this time by means of a deductible.

D-**F**-**D** is also feasible. Here, we initially treat Hs and use a deductible to deter Ls from seeking treatment. After some H→W, it becomes optimal to treat and fully cover both Hs and Ls. After still more H→W, however, it once again becomes optimal to deter Ls by means of a deductible.

Z-**F**-**Z** entails even more drastic shifts. Initially, no one is treated. π_H drops, and both Hs and Ls are treated. Then π_H drops still more and we return to treating no one. With a slight drop in κ_L, the sequence becomes an even more erratic **Z**-**F**-**Z**-**D**.

However, the critical observation is that despite these rather odds shifts between modes, H→W progress is always welfare-improving. In the last sequence discussed, for example, every infinitessimal decrease in π_H is welfare-improving, despite the shifts from **Z** to **F** to **Z** to **D**.

Table 7
Some mode shifts under H® W technological progress
(preventive medicine improves: $\Delta\pi_H<0$ and $\Delta\pi_W=-\Delta\pi_H$)

contract sequence	κ_L	κ_H	k	y_0	x	p_L	p_H
D	10%	40%	10	$ 100,000	$ 200,000	15%	**0%-20%**
F	20%	90%	10	$ 100,000	$ 150,000	5%	**0%-20%**
Z	50%	70%	2.6	$ 100,000	$ 150,000	5%	**0%-20%**
I	3%	90%	10	$ 100,000	$ 150,000	5%	**10%-20%**
D-F	30%	80%	20	$ 100,000	$ 400,000	5%	**0%-20%**
I-D	3%	90%	10	$ 100,000	$ 150,000	5%	**0%-20%**
Z-D	25%	100%	6	$ 100,000	$ 500,000	1%	**0%-16%**
Z-F	80%	90%	2	$ 100,000	$ 150,000	5%	**0%-16%**
F-Z	50%	80%	2.6	$ 100,000	$ 150,000	5%	**0%-20%**
Z-I-D	1%	80%	4	$ 100,000	$ 250,000	1%	**0%-20%**
I-F-D	20%	100%	7	$ 100,000	$ 250,000	1%	**0%-16%**
D-F-D	18%	80%	20	$ 100,000	$ 400,000	5%	**0%-16%**
Z-F-Z	50%	80%	2.6	$ 100,000	$ 150,000	7%	**0%-20%**
Z-F-Z-D	48%	80%	2.6	$ 100,000	$ 150,000	7%	**0%-20%**

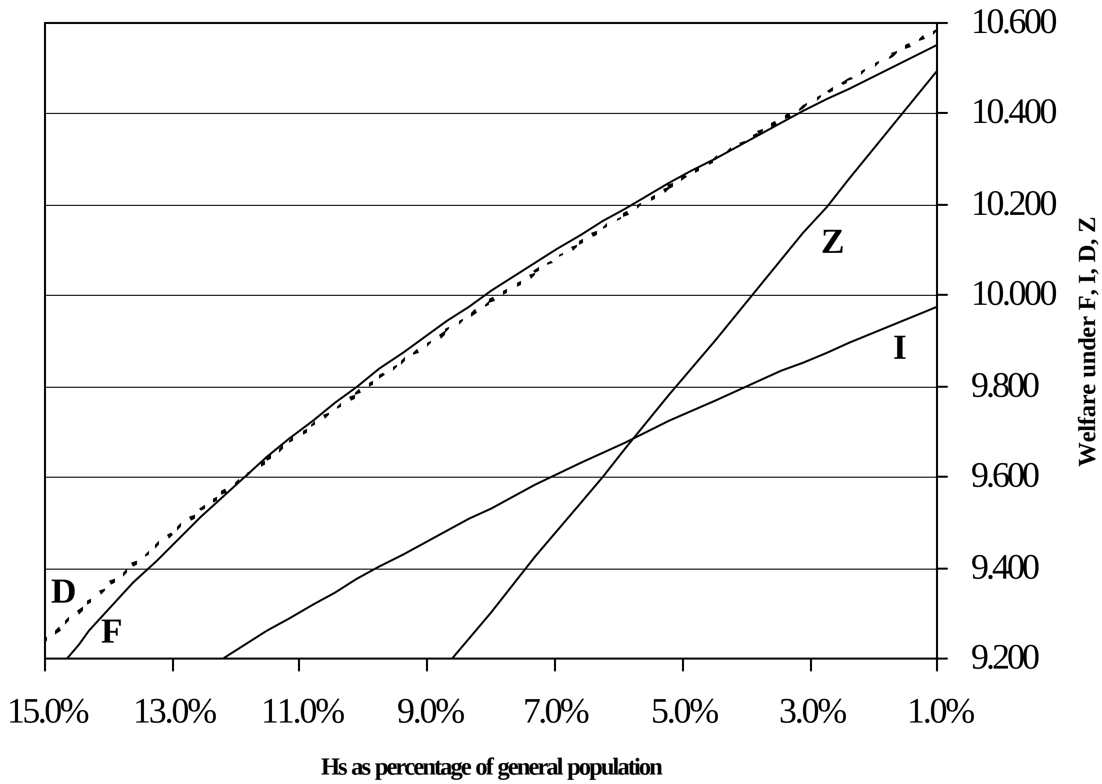
Figure 2 shows how reversals can occur. In this graph, we apply arbitrary parameters to examine welfare under each of the four feasible insurance contract modes as a function of π_H, similar to what we did in Figure 1.[9] Here, π_H declines from 15% on the left to 1% on the right. As in Figure 1, welfare climbs steadily under all contracts as preventive technology improves (this time for Hs instead of for Ls).

With H→W, only the **Z**-curve is linear with respect to π_H. In Figure 2, the nonlinearities cause **F**-curve to cut the **D**-curve from below and then from above, giving rise to the **D**-**F**-**D** contract sequence.

---

[9] The parameters here are κ_L=19%, κ_L=90%, k=17, y_0=$100,000, x=$380,000, π_L=5%, and $\pi_H\in(1\%,15\%)$.

Figure 2
Welfare, 4 Contracts, # of Hs Varies



6. Effects of a Shift from Ls to Hs ($\Delta p_L = -\Delta p_H$): L → H Technological Progress

This section combines the results of Tables 4 and 6 to examine a third type of technological progress. Here, medical science learns to turn some Ls into Hs. In sections 3 and 4, we presumed that progress meant a newfound ability to prevent the onset of disease in certain individuals. Here, there is no reduction in the incidence of disease, but treatment becomes more effective for some people. Analytically, this is equivalent to simultaneous L→W progress and an equivalent reversal of H→W progress.

L→H might occur where the presence of some comorbid condition Y reduces the probability of successfully curing disease X. If medical science reduces the incidence of Y, then some would-be Ls will become Hs instead and enjoy higher cure rates of X. Section 6.1 below explains the intuition behind the cells in Table 8. Later, Section 6.2 will discuss the possible mode changes.

6.1 L® H Technological progress: No mode changes

The entries in Table 8 are partial derivatives of the optimal indemnities, deductibles, premiums, and utility functions as π_H increases (and π_L decreases equally). We could obtain these expressions by taking the derivatives of the equations with respect to π_H and constraining π_L to decline by an equal amount. A simpler method, though, is simply to subtract the expressions in Table 6 from the equivalent cells in Tables 4. [10]

Table 8

Contract parameter and welfare changes under L® H technological progress

(treatment improves: $\Delta\pi_L<0$ and $\Delta\pi_H = -\Delta\pi_L$)

	Indemnity (i*) or Deductible (d*)	**Premium (p)**	**Utility ($\hat{U}$)**
I	$-\frac{i^*(x-i^*)}{y_0-p_H x}$ $\gtreqless 0$ if $\hat{U}_z \gtreqless \hat{U}_{i^*}$	$\frac{x-i^*}{1+(f-1)p_L}$ $\gtreqless 0$ if $\hat{U}_{i^*} \gtreqless \hat{U}_z$	$-\frac{x-i^*}{y_0-p_H x}+(k_H-k_L)k$ >0 if $\hat{U}_{i^*} \leq \hat{U}_z$; sign ambiguous otherwise
D	$-\frac{d^*}{y_0-p_H x}(x-d^*)<0$	$(x-d^*)\left[\frac{y_0-p_H x+p_H d^*}{y_0-p_H x}\right]>0$	$-\frac{x-d^*}{y_0-p_H x}+(k_H-k_L)k$ sign ambiguous
Z	—	—	0
F	—	0	$(k_H-k_L)k>0$

Each expression equals the equivalent cell in Table 4 minus the equivalent cell in Table 6.

[10] Table 3, shown earlier in the paper, derives its information from Table 8. Since the subtractions are obvious, no formal proofs are presented.

Indemnity mode (I): L→H leads to a lower optimal indemnity if **I** is preferred to **F**. (And if **I** is not preferred to **F**, then changes in the optimal indemnity are irrelevant, because the market will not choose **I**.). The lower indemnity occurs because Hs are more costly to the insurer than Ls (because $x>i^*$ anytime that we are in mode **I**). So, as some Ls become Hs, the insurer's expenses rise (along with premium); this reduces Ls' *ex ante* wealth, thus reducing their willingness to pay for treatment (i.e., the optimal indemnity).

If **I** is preferred to **F**, the change in welfare resulting from L→H is ambiguous because Ls become Hs, *ex post* wealth for all agents declines. Ls receive smaller indemnities, and all agents pay higher premiums. In utility terms, the cost of these wealth effects rises with each successive individual who shifts from L to H; the marginal health gain, however, is constant. So at some point, the wealth cost overwhelms the health benefit, and welfare begins to decline.

Deductible mode (D): L→H leads to a lower optimal deductible and a higher optimal premium. The premium rises because for every L who becomes an H, the insurance company must pay the treatment cost $x-d^*$. With a higher premium, Ls' willingness to pay (i.e., the optimal deductible) declines. The declining deductible raises the covered treatment cost, further driving down the deductible until some equilibrium is reached. As with the indemnity contract, an increasing marginal utility of wealth eventually overwhelms the constant marginal utility of health; L→H can either increase or decrease welfare.

Zero-insurance mode (Z): Since there are neither deductibles nor premiums, L→H progress induces no change in any agent's *ex post* wealth. There are no changes in any agent's *ex post* health condition, either. This is because the total number of sick people does not change and, without treatments, no sick person is ever cured. With no patients being treated, the technological progress has no effect on any agent's health or wealth, so welfare is unaffected.

Full-insurance mode (F): L→H leaves the premium unchanged, but raises welfare. The same number of sick people as before are treated, so there is no change in the insurance expenses that must be covered; hence, the premium is static. A higher number of sick people are cured, however, because those would-be Ls who become Hs enjoy a higher cure rate. In sum, no agent's *ex post* wealth changes, but some people end up healthier, so welfare improves.

6.2 L® H technological progress: Mode changes

Here, we examine some possible mode sequences under L→H. Table 9 shows that, as in the case of H→W, some unusual sequence are possible, thanks to nonlinearities. In Table 9, we use arbitrary sets of parameters. In each case, we assume that there are initially Ls but no Hs and, in the end, Hs, but no Ls; we could also have assumed more narrow ranges of variation where we begin and end with some Hs and some Ls.

Table 9
Some mode shifts under L® H technological progress
(preventive medicine improves: $\Delta\pi_L < 0$ and $\Delta\pi_L = -\Delta\pi_H$)

Contract sequence	κ_L	κ_H	k	y_0	x	$p_L + p_H =$
F	2%	80%	10	$ 100,000	$ 200,000	**20%**
Z	20%	30%	6	$ 100,000	$ 225,000	**40%**
D-I	2%	80%	10	$ 100,000	$ 200,000	**20%**
D-F	30%	80%	10	$ 100,000	$ 300,000	**10%**
D-Z	1%	20%	15	$ 100,000	$ 225,000	**40%**
Z-F	25%	35%	10	$ 100,000	$ 250,000	**10%**
D-I-Z	1%	20%	15	$ 100,000	$ 180,000	**40%**
D-Z-I	5%	10%	27	$ 100,000	$ 200,000	**20%**
D-Z-F	30%	40%	15	$ 100,000	$ 225,000	**40%**
D-Z-I-Z	5%	15%	11.8	$ 100,000	$ 150,000	**20%**

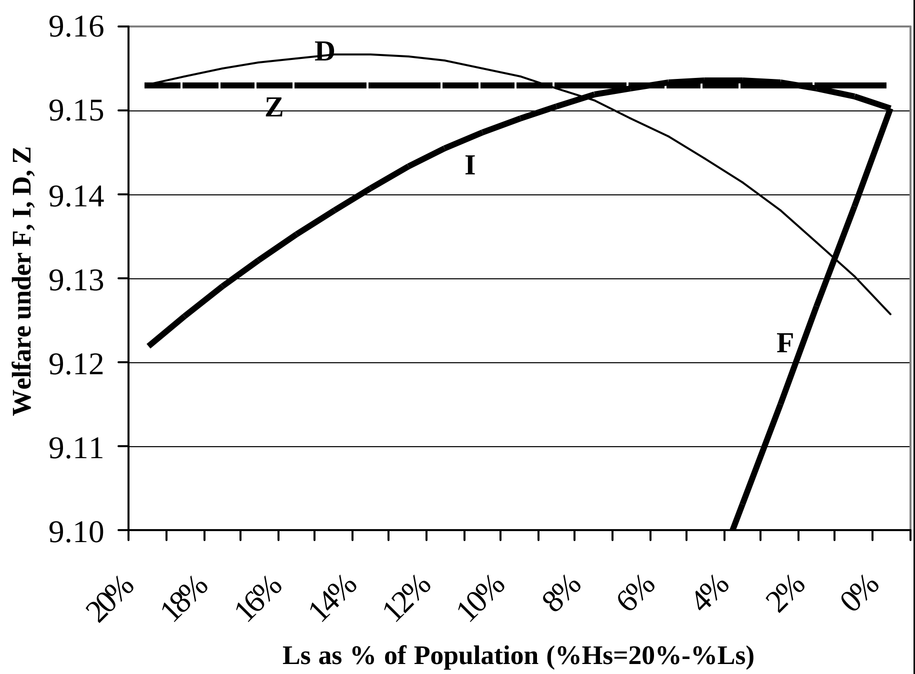
Example: In the last row, π_L=20% and π_H=0% initially. As technological gradually decreases the number of Ls and increases the number of Hs, we proceed through the sequence **D-Z-I-Z**. Eventually, π_L=0% and π_H=20%.

Figure 3 illustrates how one unusual mode sequence comes about.[11] Both the **D**-curve and the **I**-curve have upward- and downward-sloping segments; the **Z**-curve is horizontal; and the **F**-curve (irrelevant in this example) is always upward-sloping. Welfare here takes a roller-coaster ride. At the extreme left, welfare begins at what will prove to be its low point. As some Ls become Hs, welfare rises and then falls back to its minimum level. Then, **D** gives way to **Z** with welfare hovering at its previous minimum. Eventually, the mode shifts to **I**, and welfare rises again (though not as far as it did under **D**) and then falls back again. Finally, the market shifts back to **Z**, and welfare settles down at its global minimum. In this example, as π_L declines from 20% to 0% (and π_H does the opposite), welfare reaches its maximum at around the point where π_L=14% and π_H=6%. Afterwards, technological progress can be considered mostly a losing

[11] The parameters here are κ_L=5%, κ_L=15%, k=11.8, y_0=\$100,000, x=\$150,000, $\pi_L \in$ (20%, 0%), and π_H=20%$-\pi_L$.

proposition. Contracts come and go, welfare undulates, but expected welfare never attains that momentary high.

Figure 3
Welfare, 4 contracts, Ls become Hs



By specifying different parameters, we can create examples where L→H continually improves welfare, or where progress continually reduces welfare; and all sorts of intermediate cases are possible. The central message here, though, is that this type of improvement in medical science yields ambiguous results—sometimes beneficial, sometimes not. In fact, the stretches of progress-induced welfare decline are always in situations where more people are being cured—where health is unambiguously improving. The problem is that the better health comes at too high a financial cost.

7. Conclusion

This paper has explored ways in which several types of technological progress might affect health insurance markets and medical outcomes. Under theoretical conditions specified in our model, as medical science improves its power to prevent illness (L$\rightarrow$W- or H$\rightarrow$W-type progress), welfare will unambiguously rise, but the market may pass through some unusual sequences of insurance contracts (modes). If medicine increases its curative powers (or its curative or preventive powers over some comorbid condition), the welfare effects may be ambiguous (L$\rightarrow$H-type progress). A more highly curable sick population may be a less happy population—the medicine may be worse than the malady in terms of utility. Importantly, this may even be true if some peoples' health is improving and no one's health is deteriorating—i.e., unambiguous health improvement. This is because as treatment outcomes improve, wealth effects may begin to dominate.

Another result of this model is that externalities may be crucial to the workings of a health insurance market that seeks to exclude from treatment those people who benefit least. We can specify cases in which there is no change in demographics, in the inputs required for treatment, or in the cost of inputs, but where welfare and the structure of health insurance contracts may change considerably because of external factors. For instance, research that reduces the incidence of a comorbidity Y may enable physicians to improve their performance (i.e., cure rates) on disease X with the same inputs. The same procedure, at the same price, may become more and more attractive to specific population groups, thus expanding demand and undermining the financial viability of an insurance scheme. We can think of this as a sort of societal learning curve with stochastic elements. The implication is that actuarially sound insurance plans may collapse under the weight of external factors. We can imagine the Centers for Disease Control, the National Institutes of Health, or *JAMA* publishing information that financially destabilizes the plan by making a medical procedure attractive to too many takers.

Finally, the ideas developed here could be expanded by making technological progress endogenous to the model rather than externally imposed. If firms are aware of how progress will influence contract parameters and modes, then this knowledge may feed back on the rate or specifics of technological progress. Then, medical research and technology diffusion would be endogenous.

REFERENCES

Akerlof, George A. 1970. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* 84:488-500.

__________. 1978. The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *American Economic Review*, 68: 8-19.

__________. March 1983. Loyalty filters. *American Economic Review*, 73:54-63.

Arrow, Kenneth J. 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* 53: 941-973.

Baumgardner, James R. 1991. the interaction between forms of insurance contract and types of technical change in medical care. *RAND Journal of Economics* 22-1: 36-53.

Cutler, David and Louise Sheiner. 1997. Managed care and the growth of medical expenditures. *NBER Working Paper Series* #6140.

Cutler, David M. 1996. Public policy for health care. *NBER Working Paper* #5591.

Ellis, Randall B. and McGuire, Thomas G. 1993. Supply-side and demand-side cost sharing in health care. *Journal of Economic Perspectives* 7:135-152.

Feigenbaum, Susan. 1992. "Body shop" economics: What's good for our cars may be good for our health. *Regulation: The Cato Review of Business and Government* 15(4).

Feldstein, Martin S. 1973. The welfare loss of excess health insurance. *Journal of Political Economy* 81:251-280.

Fuchs, Victor R. 1986. *The Health Economy.* Cambridge: Harvard University Press.

Gianfrancesco, Frank D. 1983. A proposal for improving the efficiency of medical insurance. *Journal of Health Economics* 2:175-184.

Goddeeris, John H. 1984. Medical insurance, technological change, and welfare. *Economic Inquiry* 22:56-67.

Graboyes, Robert F. 2000a. Our money or your life: Indemnities vs. deductibles in health insurance. *Federal Reserve Bank of Richmond Working Papers*.

Graboyes, Robert F. 2000b. Getting better, feeling worse: Cure rates, health insurance, and welfare. *Federal Reserve Bank of Richmond Working Papers*.

Neipp, Joachim and Richard Zeckhauser. 1985. Persistence in the choice of health plans. In *Advances in Health Economics and Health Services Research*, ed. R.M. Scheffler and L.F. Rossiter 6 47-72.

Weisbrod, Burton A. 1991. The health care quadrilemma: An essay on technological change, insurance, quality of care, and cost containment. *Journal of Economic Literature* 29:523-552.

Zeckhauser, Richard. 1970. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory* 2:10-26.

APPENDIX

This Appendix has two sets of proofs. Section A.1 derives the expressions found in Table 4, which shows the effects of L→W progress. Section A.2 derives the expressions found in Table 6, which shows the effects of H→W progress. Throughout this Appendix, numbers in square brackets (e.g., [10]) refer to the equations in Section 3 of the main text of this paper. Equations introduced in the Appendix are numbered in the form (x.y). In transferring the expressions derived here to the two tables, we reverse the signs, because L→W and H→W progress implies a *reduction* in π_L and π_H, respectively.

A.1 Table 4 Derivations

The section proves the nine number derivatives from Table 4. The number preceding each derivative corresponds to Propositions 1 through 9 below.

Proposition 1: If π_L declines, the optimal indemnity i* rises [$\partial i^*/\partial \pi_L<0$]. Recall:

[10] $$i^* = \frac{(f-1)(y_0 - \pi_H x)}{1+(f-1)\pi_L}, \text{ where } f = e^{k_L k}$$

As π_L declines, it takes a larger indemnity to deter Ls from seeking treatment. We can prove this as follows:

(1.1)
$$\begin{aligned}
\frac{\partial i^*}{\partial \pi_L} &= (f-1)(y_0 - \pi_H x)\frac{\partial f}{\partial \pi_L}\frac{1}{[1+(f-1)\pi_L]} \\
&= \frac{(f-1)^2(y_0-\pi_H x)}{[1+(f-1)\pi_L]^2} \\
&= -\frac{i^{*2}}{(y_0-\pi_H x)} < 0
\end{aligned}$$

[Q.E.D.]

So as insurance pays premiums to fewer Ls, premiums decline, thereby increasing the Ls' willingness to pay for treatment; so, the indemnity rises in response.

Proposition 2: If π_L declines, the premium p_{i*} also declines [$\partial p_{i*}/\partial \pi_L > 0$]. The premium must cover both the cost x of treating Hs and the indemnity i* paid to Ls in lieu of treatment:

[5] $p_{i*} = \pi_H x + \pi_L i*$

π_L enters through both terms of the right-hand side of this expression i*, so

$$\begin{aligned}
\frac{\partial p_{i*}}{\partial \pi_L} &= \frac{\partial \pi_L}{\partial \pi_L} i* + \pi_L \frac{\partial i*}{\partial \pi_L} \\
&= i* + \pi_L \left[-\frac{i*^2}{(y_0 - \pi_H x)} \right] = i* \left[\frac{y_0 - \pi_H x - \pi_L i*}{y_0 - \pi_H x} \right] > 0 \\
&\textit{or} \\
&= \frac{[1+(f-1)\pi_L]}{[1+(f-1)\pi_L]} \frac{(f-1)(y_0 - \pi_H x)}{[1+(f-1)\pi_L]} - \pi_L \frac{(f-1)^2 (y_0 - \pi_H x)}{[1+(f-1)\pi_L]^2} \qquad (2.1) \\
&= \frac{(f-1)(y_0 - \pi_H x) + (f-1)^2 \pi_L (y_0 - \pi_H x) - (f-1)^2 \pi_L (y_0 - \pi_H x)}{[1+(f-1)\pi_L]^2} \\
&= \frac{(f-1)(y_0 - \pi_H x)}{[1+(f-1)\pi_L]^2} = \frac{i*}{1+(f-1)\pi_L} > 0
\end{aligned}$$

While the indemnity rises, as shown in Proposition 1, this rise is more than offset by the smaller number of Ls needing indemnities. [Q.E.D.]

Proposition 3: If π_L declines, welfare under the optimal indemnity contract rises [$\partial \hat{U}_{i*}/\partial \pi_L < 0$]. In logarithmic form, result (5) from Graboyes (2000a) is:

[5] $\hat{U}_{i*} = \ln(y_0 - p_{i*}) - (\pi_S - \pi_H k_H - \pi_L k_L)k$, where $p_{i*} = \pi_H x + \pi_L i*$

From this we can derive

$$
(3.1)\qquad
\begin{aligned}
\frac{\partial \hat{U}_{i*}}{\partial \pi_L} &= \frac{\partial}{\partial \pi_L}\left[\ln\left(y_0 - p_{i*}\right) - \left(\pi_S - \pi_H \kappa_H - \pi_L \kappa_L\right)k\right] \\
&= \frac{1}{y_0 - p_{i*}}\left[-1\right]\frac{\partial p_{i*}}{\partial \pi_L} - \left(1 - \kappa_L\right)k \\
&= -\frac{1}{y_0 - \pi_H x - \pi_L i^*}\left[\frac{y_0 - \pi_H x - \pi_L i^*}{y_0 - \pi_H x}\right] - \left(1 - \kappa_L\right)k \\
&= -\frac{1}{y_0 - \pi_H x} - \left(1 - \kappa_L\right)k < 0
\end{aligned}
$$

since both terms are negative. Thus, an improved preventive ability (lower π_L) *increases* welfare, as intuition would suggest. [Q.E.D.]

Proposition 4: If π_L rises, then the optimal deductible d* rises. Recall:

$$
[11]\qquad d^* = \frac{\left(f - 1\right)\left(y_0 - \pi_H x\right)}{f - \left(f - 1\right)\pi_H}, \text{ where } f = e^{\kappa_L k}
$$

Since π_L appears nowhere in this expression,

$$
(4.1)\qquad \frac{\partial d^*}{\partial \pi_L} = 0
$$

The number of Hs and the amount spent on treating them does not change, so the premium remains unchanged. Therefore the Ls' post-premium wealth is unaffected, thereby leaving their willingness to pay unaffected.

Proposition 5: If π_L declines, then the premium p_{d*} will remain unchanged. The premium covers the cost x of treating Hs, minus the deductible d* paid out-of-pocket:

[6] $p_{d*} = \pi_H x - \pi_H d^*$

Once again, π_L appears nowhere in this expression, so:

$$
(5.1)\qquad \frac{\partial p_{d*}}{\partial \pi_L} = 0
$$

Proposition 6: If p$_L$ **declines, then welfare under the optimal deductible contract rises [¶$\hat{U}_{d*}$/¶p$_L$<0].** In logarithmic form, result [6] is:

[6] $\hat{U}_{d*} = \ln(y_0 - p_{i*}) - (p_S - p_H k_H + p_H k_L)k$, where $p_{d*} = p_H x - p_H d*$

Since π_L does not enter into the expression:

(6.1) $$\frac{\P\hat{U}_{i*}}{\P p_L} = -k < 0$$

Since there is no change in the deductible, premium, or identity of patients treated, the only welfare change comes from the smaller number of people (Ls, in particular) falling ill in the first place.

Proposition 7: If p$_L$ **declines, then welfare under the zero-insurance mode increases.** Since we have assumed that treatment is prohibitively expensive without insurance, then it is simple to show that in a zero-insurance mode, changes in π_L will have no effects on *ex post* income, but welfare will increase. Formally, we know that without insurance

[7] $\hat{U}_z = \ln(y_0) - p_S k$

π_L only appears as part of π_S, so:

[8] $$\frac{\P\hat{U}_z}{\P p_L} = -k < 0$$

Once again, welfare is only affected by the reduced number of Ls falling ill. Of course, without any insurance contracts whatsoever, there are no indemnities, deductibles, or premiums, so we need not consider those parameters.

Proposition 8: If p$_L$ **declines, then the insurance premium p$_f$ falls.** In this case, all Hs and Ls receive treatment, and all subscribers pay an equal premium. So, a reduction in the number of Ls falling ill will reduce the total premiums paid and, hence, the per capita cost to subscribers. The premium is:

[8] $\pi_f = (\pi_H + \pi_L)x$,

so the change in the premium is

(8.1) $$\frac{\partial \pi_f}{\partial \pi_L} = x$$

Proposition 9: If π_L declines, then welfare rises. We know that utility under the full-insurance mode is:

[8] $\hat{U}_f = \ln(y_0 - \pi_f) - (\pi_S - \pi_H \kappa_H - \pi_L \kappa_L)k$ where $\pi_f = (\pi_H + \pi_L)x$

When π_L declines, there are two welfare effects. There is a wealth effect from the reduced premium demonstrated in Proposition 8. And, there are fewer Ls falling ill.

(9.1) $$\begin{aligned}\frac{\partial \hat{U}_f}{\partial \pi_L} &= \frac{\partial}{\partial \pi_L} \ln(y_0 - \pi_H x - \pi_L x) - (\pi_S - \pi_H \kappa_H - \pi_L \kappa_L)k \\ &= -\frac{x}{y_0 - \pi_H x - \pi_L x} - (1 - \kappa_L)k < 0\end{aligned}$$

We know this is negative, since both of its component expressions are negative.

A.2: Table 6 Derivations

Proposition 10: If π_H declines, the optimal indemnity i* rises [$\partial i^*/\partial \pi_H < 0$]. Recall:

[10] $$i^* = \frac{(f-1)(y_0 - \pi_H x)}{1 + (f-1)\pi_L}, \text{ where } f = e^{\kappa_L k}$$

As π_H declines, it takes a larger indemnity to deter Ls from seeking treatment. We can prove this as follows:

$$(10.1)\quad \frac{\P i^*}{\P p_H} = \frac{(f-1)x}{1+(f-1)p_L} = -\frac{xi^*}{(y_0 - p_H x)} < 0$$

Inspection shows that this expression is negative.

Proposition 11: If p_H declines, the premium p_{i*} also declines [$\P p_{i*}/\P p_H>0$]. The premium must cover both the cost x of treating Hs and the indemnity i* paid to:

[5] $p_{i*} = p_H x + p_L i^*$

π_H enters through both terms of the right-hand side of this expression i*, so

$$(11.1)\quad \frac{\P p_{i*}}{\P p_H} = x + p_L \frac{\P i^*}{\P p_H} = x - \frac{p_L x i^*}{y_0 - p_H x} = x\left[1 - \frac{p_L i^*}{y_0 - p_H x}\right] = x\left[\frac{y_0 - p_H x - p_L i^*}{y_0 - p_H x}\right] > 0$$

We can see by inspection that (A.2.4) is always positive. [Q.E.D]

Proposition 12: If p_H declines, then welfare under the optimal indemnity contract rises [$\P \hat{U}_{i*}/\P p_H<0$]. In logarithmic form, result (5) is:

[5] $\hat{U}_{i*} = \ln(y_0 - p_{i*}) - (p_S - p_H k_H - p_L k_L)k$, where $p_{i*} = \pi_H x + \pi_L i^*$

From this we can derive

$$
(12.1)\quad
\begin{aligned}
\frac{\partial \hat{U}_{i*}}{\partial \pi_H} &= \frac{\partial}{\partial \pi_H}\left[\ln\left(y_0 - p_{i*}\right) - \left(\pi_S - \pi_H k_H - \pi_L k_L\right)k\right] \\
&= -\frac{1}{y_0 - p_{i*}}\frac{\partial p_{i*}}{\partial \pi_H} + k_L k - k \\
&= -\frac{x}{y_0 - \pi_H x - \pi_L i*}\left[\frac{y_0 - \pi_H x - \pi_L i*}{y_0 - \pi_H x}\right] + k_L k - k \\
&= -\frac{x}{y_0 - \pi_H x} + k_L k - k < 0
\end{aligned}
$$

since the first term is negative, as is the sum of the second and third terms. Thus, an improved preventive ability (lower π_H) *increases* welfare, as intuition would suggest. [Q.E.D]

Proposition 13: If π_H rises, then the optimal deductible d* rises. Recall:

[11] $$d* = \frac{(f-1)(y_0 - \pi_H x)}{f - (f-1)\pi_H}$$, where $f = e^{k_L k}$

If we say u= $(y_0 - \pi_H x)$ and v= $f - (f-1)\pi_H$, then:

$$
(13.1)\quad
\begin{aligned}
\frac{\partial d*}{\partial \pi_H} &= (f-1)\frac{\partial}{\partial \pi_H}\frac{u}{v} \\
&= (f-1)\frac{u(f-1) - xv}{v^2} \\
&= \frac{(f-1)}{v}(d* - x) \\
&= -\frac{d*}{y_0 - \pi_H x}(x - d*) < 0
\end{aligned}
$$

A decline in π_H causes the optimal deductible to rise.

Proposition 14: If π_H declines, then the premium p_{d*} will decline. The premium covers the cost x of treating Hs, minus the deductible d* paid out-of-pocket by patients:

[6] $p_{d*} = \pi_H x - \pi_H d*$

From this, we see that:

$$
(14.1) \quad
\begin{aligned}
\frac{\P p_{d^*}}{\P p_H} &= x - d^* - p_H \frac{\partial d^*}{\partial p_H} \\
&= (x - d^*) + p_H \frac{d^*}{y_0 - p_H x}(x - d^*) \\
&= (x - d^*)\left[1 + p_H \frac{d^*}{y_0 - p_H x}\right] \\
&= (x - d^*)\left[\frac{y_0 - p_H x + p_H d^*}{y_0 - p_H x}\right] > 0
\end{aligned}
$$

A decline in π_H leaves p_{d*} lower, as well. [Q.E.D.]

Proposition 15: If p_H declines, then welfare under the optimal deductible contract rises [$\P \hat{U}_{d*} / \P p_H < 0$]. In logarithmic form, result [6] is:

[6] $\hat{U}_{d^*} = \ln(y_0 - p_{i^*}) - (p_S - p_H k_H + p_H k_L)k$, where $p_{d^*} = p_H x - p_H d^*$

We can see that:

$$
(15.1) \quad
\begin{aligned}
\frac{\P \hat{U}_{d^*}}{\P p_H} &= -\frac{1}{y_0 - p_{d^*}} \frac{y_0 - p_{d^*}}{y_0 - p_H x}(x - d^*) + k_H k - k_L k - k \\
&= -\frac{1}{y_0 - p_H x}(x - d^*) + k_H k - k_L k - k < 0
\end{aligned}
$$

So, as π_H declines, utility rises.

Proposition 16: If p_H declines, then welfare under the zero-insurance mode increases. Since we have assumed that treatment is prohibitively expensive without insurance, then it is simple to show that in a zero-insurance mode, changes in π_H will have no effects on *ex post* income, but that welfare will increase. Formally, we know that without insurance Recall:

[7] $\hat{U}_z = \ln(y_0) - p_S k$

π_H only appears as part of π_S, so:

(16.1) $\frac{\partial \hat{U}_z}{\partial \pi_H} = -k < 0$

Once again, welfare is only affected by the reduced number of Hs falling ill. Of course, without any insurance contracts whatsoever, there are no indemnities, deductibles, or premiums, so we need not consider those parameters. [Q.E.D.]

Proposition 17: If π_H declines, then the insurance premium p_f falls. In this case, all Hs and Ls receive treatment, and all subscribers pay an equal premium. So, a reduction in the number of Hs falling ill will reduce the total premiums paid and, hence, the per capita cost to subscribers. The premium is:

[8] $p_f = (\pi_H + \pi_L)x$,

so the change in the premium is

(17.1) $\frac{\partial p_f}{\partial \pi_H} = x$

Proposition 18: If π_H declines, then welfare rises. We know that utility under the full-insurance mode is:

[8] $\hat{U}_f = \ln(y_0 - p_f) - (\pi_S - \pi_H k_H - \pi_L k_L)k$ where $p_f = (\pi_H + \pi_L)x$

When π_H declines, there are two welfare effects. First, there is a wealth effect from the reduced premium demonstrated in Proposition 17. Second, there are fewer Hs suffering the effects of illness.

(18.1)
$$\begin{aligned}\frac{\partial \hat{U}_f}{\partial \pi_H} &= \frac{\partial}{\partial \pi_H} \ln(y_0 - \pi_H x - \pi_L x) - (\pi_S - \pi_H k_H - \pi_L k_L)k \\ &= -\frac{x}{y_0 - \pi_H x - \pi_L x} + k_H k - k < 0\end{aligned}$$

We know this is negative, since both of its component expressions are negative. [Q.E.D.]