Getting Better, Feeling Worse: Cure Rates, Health Insurance, and Welfare*

Robert F. Graboyes**

Federal Reserve Bank of Richmond Working Paper No. 00-5

August 2000

JEL Nos. I11, G22 D81

Keywords: health insurance, technological change, heterogeneity, indemnities

Abstract

We model a health insurance market where rising cure rates for a disease may paradoxically diminish welfare and even negate the desirability of health insurance altogether. In the model, rising cure rates can affect welfare in two ways: (1) directly, by improving some individuals' health, and (2) indirectly, by influencing the mode and parameters of the optimal insurance contract and, thus, *ex post* financial wealth distribution. ("Mode" refers to the qualitative specifications of the contract—presence or absence of indemnities and full, partial or zero coverage of treatments received. "Parameters" refers to the quantitative features of the contracts—level of nonzero indemnities, deductibles and premiums.)

Graboyes (2000a) compares the relative efficiency of deductibles and indemnities in deterring low-benefit patients (those whom treatment is less likely to cure) from seeking expensive medical care. The current paper asks how optimal insurance and associated welfare change as curative technology improves. Deterring some but not all patients requires deductibles or indemnities because one's H/L status is known to all, but not legally verifiable.

We find that advances in curative power may reduce welfare and even eradicate health insurance. This is because at higher cure rates, higher indemnities and deductibles are needed to deter patients from seeking treatment. Given certain population parameters, higher cure rates can *only* reduce welfare. The findings raise some practical, empirical, and ethical questions; some of these issues are enumerated in the conclusion.

---

* 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. Technological Progress Defined

In a world of expensive medical procedures and competing wants, some treatable patients should not be treated because the opportunity costs are too great. This assertion, harsh to some ears, is implicitly recognized by elements of our health care system and by private choices. In calm moments of good health, all might agree, for example, that an expensive, near-futile cancer treatment is not the best use of resources. The problem is binding ourselves to such an agreement, for when we fall ill, a slim chance at life often trumps prior pledges.

This paper follows from Graboyes (2000a), which compared the relative desirability of four contract structures, or "modes": Indemnity mode (**I**) offers low-benefit patients cash payoffs in lieu of treatment.[1] Deductible mode (**D**) requires anyone seeking treatment to pay an out-of-pocket deductible high enough to deter low-benefit patients. Zero insurance mode (**Z**) provides no insurance, so no one seeks treatment. Full insurance mode (**F**) provides full coverage of treatment for high- and low-benefit patients.

The current paper adds an additional complication—changing efficacy of medical technology. In the model developed here, rising cure rates can paradoxically diminish welfare and even render treatments and health insurance inefficient. Rising cure rates can induce changes in optimal contract parameters (premiums, deductibles, indemnities) within a mode; or they can induce the market to switch modes.

[1] Traditional fee-for-service policies are also referred to as "indemnification" policies. In contrast, this paper uses "indemnity" contract to indicate a plan where patients are reimbursed in cash for a diagnosis, rather than for a treatment.

1.1 A Hypothetical Example

Take a thousand people whose only health concerns are found in actuarial tables and pose the following: "In the next year, 25 of you will contract cancer. Treatments cost $500,000. Because of certain physiological divisions in this population, 5 patients will recover with certainty if treated, but the other 20 will only experience a 5% survival rate. You can agree today to any of three options: (1) Full coverage of treatments for all 25 patients paid for by actuarially fair insurance; (2) Full coverage of treatments for the 5 high-benefit patients ("Hs") but no treatment for the 20 low-benefit patients ("Ls"); or (3) No insurance and no treatments. Which option do you want?" Table 1 provides hypothetical data on these three options:

Table 1
Hypothetical Case, Insurance Options

	1: Ls and Hs treated	2: Hs only treated	3: no insurance
annual premium per capita	$12,500	$2,500	$0
lives saved	6	5	0
% saved by treatment	0.6%	0.5%	0%

From an individual standpoint, $2,500 buys an additional 0.5% chance of surviving the year (assuming this is the only life-threatening prospect). The next $10,000 only adds another 0.1% chance of survival. Given certain utility curves, the 1,000 would agree to option 2 while still behind a veil of good health. The operational problem is legally verifying who is an H and who is an L. When a year passes and 25 people have fallen ill, each of the 20 Ls will have a powerful incentive to portray himself as an H. Without an objectively verifiable criterion for separating Hs and Ls, the insurer will be hard-pressed to limit coverage to Hs, even if the insurer and patients can all distinguish between Hs and Ls. Even if there is an objectively verifiable delimiter between the two classes, legal, political, and public relations concerns may make it difficult for the insurer to deny coverage on the basis of that delimiter.

In practice, there are many ways to exclude Ls from treatment. Utilization review, queuing, moral suasion, redefining and refining diagnostic categories, and advance medical directives all act as filters. Oregon's much-publicized Medicaid plan explicitly excluded certain low-benefit patients from treatment. This paper focuses exclusively on indemnities and deductibles— demand-side tools that induce low-benefit patients to voluntarily opt out of treatment—and on the **Z** and **F** modes that do not discriminate between high- and low-benefit patients.

1.2 Additional Problem: Technological Progress

Graboyes (2000a) derived criteria for choosing among the four modes. The current paper adds the additional complication of technological progress (defined by rising cure rates), which can lead to some surprising results, which fall into two general categories: (1) Progress causes expected welfare to decline. (2) Progress causes the optimal mode to change in counterintuitive ways. Table 2 expands on the previous hypothetical example, again with made-up data:

Table 2
Hypothetical Case with Technological Progress Added

	Before	After	No insurance
annual premium per capita	$2,250	$2,050	$0
lives saved	5	5	0
deductible paid by Hs	$50,000	$90,000	$0
additional lives that could have been saved by treating all sick people	1	2	7

In Table 2, the cure rate for Ls rises, so the deductible needed to deter Ls also rises (because probabilistically, the benefit is higher). In the "before" column, everyone pays the $2,250 premium. Hs pay an additional $50,000 out-of-pocket for treatment—just enough to deter Ls from treatment. In the "after" column, the cure rate for Ls has risen from 5% to 10%. Now, it takes $90,000 to deter Ls. The number of lives saved is unchanged at 5, but the *ex post* wealth distribution changes. Everyone's premium is lower, but Hs pay a $90,000 deductible instead of $50,000. In the model developed in this paper, the welfare improvement from the lower premium is more than offset by the welfare loss due to less-complete risk sharing represented by the higher deductible.

Second, the model suggests that progress can change the optimal insurance environment in some surprising ways. Referring again to Table 2, the rising cure rates for Ls might eliminate insurance and treatment altogether (as in the "zero insurance" column); the welfare cost of losing a marginal .05% chance at survival might be less the utility gain from the more desirable wealth distribution. [2] The conclusion of this paper will itemize some of the empirical questions that this model raises, including:

- Do rising cure rates actually drive away insurance coverage? If not, why not?
- Do rising cure rates induce higher deductibles or indemnities? If not, why not?
- Do costs associated with switching modes induce markets to select insurance which is not generally optimal at the moment but which is optimal over time?
- Is behavior seen in this model responsible for the lack of indemnity health policies?
- Can effects described by the model skew the directions of medical research?
- How does the structure of the optimal contract affect the diffusion of medical technology?

---

[2] Perhaps the cure rate rises because JAMA publishes new information that improves treatment results without raising costs—a societal learning curve. Thus an external development could render insurance unsustainable.

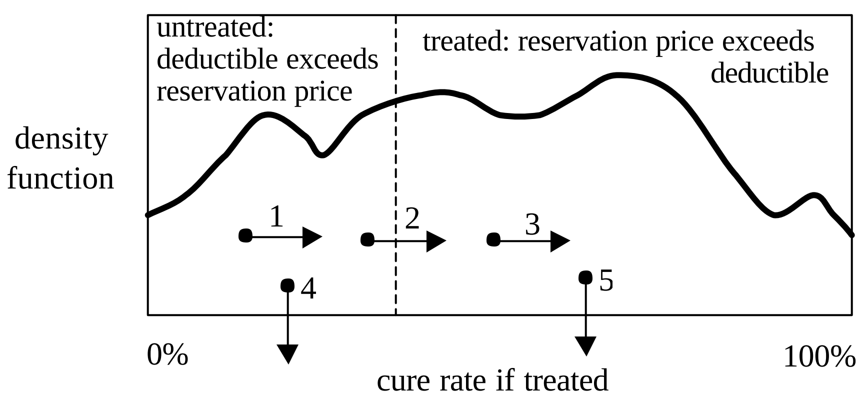
1.3 Digression: Generalized Heterogeneity and Technological Progress

It is important to note that a restricted form of heterogeneity is a central feature of the three essays of this dissertation. People are either sick (Ss) or well (Ws). Ss have either a high cure rate (Hs) or a low cure rate (Ls). Technological progress is also restricted. In the current paper, progress consists of an improved cure rate for either all Ls or all Hs in lockstep. In Graboyes (2000c), progress occurs when population parameters change so that some Ls become Ws, some Hs become Ws, or some Ls become Hs (denoted L→W, H→W, or L→H).

In a more general specification, sick people could be arrayed across a continuous distribution of cure rates (if treated). Technological progress, too could be generalized by allowing each sick person's cure rate to change uniquely, as suggested by Figure 1.

Figure 1 assumes a continuum of cure rates. The arrows indicate how technological

Figure 1
Distribution of sick population across continuum of cure rates



progress affects five individuals, and we presume that this market chooses the optimal deductible policy. (Analyzing the optimal indemnity policy would yield similar findings.)

#1, #2, and #3 show three individuals whose cure rates rise. #4 and #5 show two individuals who now never become sick at all. Each arrow provokes different effects on the two components of utility—*ex post* wealth distribution and *ex post* health.

#4 and #5 are equivalent to L→W and H→W from Graboyes (2000c); they involve a shift in the ratio of sick to well people. In contrast, progress in the current paper and L→H progress in Graboyes (2000c) involve no change in the ratio of sick people to well people and, thus, are in no way analogous to arrows #4 or #5.

In Figure 1, shifts #1 and #3 do not affect anyone's *ex post* wealth because both agents are inframarginal. #3 does improve utility because that agent's cure rate is now higher. In the current paper and in Graboyes (2000c), there are no untreated inframarginal agents like #1. Ls are just on the margin, and Hs are inframarginal but treated.

Most of the interesting results in the current paper and in Graboyes (2000c) arise from agents crossing the threshold from left to right, as does #2. In the current paper, *all* Ls shift marginally to the right. In Graboyes (2000c) (L→H), *some* Ls move a discrete distance to the H reservation price. Only when an agent crosses the dotted line do contract parameters change, thus affecting *ex post* wealth distribution.

Continuous cure rates and unrestricted forms of technological progress leave us with a great number of possible results from technological progress. This dissertation presents some reasons why the two special cases presented in the current paper and in Graboyes (2000c) might be fairly realistic. Demonstrating this empirically will be the subject of future papers. We note that in the more general construction (as in Figure 1), the difference between shifting cure rates of the current paper and shifting population distribution of Graboyes (2000c) become blurred; the separate identities of the κs and the πs arise from a discrete distribution.

The remainder of the paper is structured as follows: Section 2 reviews literature related to this paper. Section 3 reviews the assumptions, notation, and results from Graboyes (2000a) in order to set up the problem addressed in the current paper. Section 4 develops the model of technological progress and reports its results. (The results are proven in the Appendix to avoid disrupting the narrative flow.) Section 5 examines the bimodal grouping of cure rates postulated in the model and speculates on how technological improvements for the two groups might be related. Section 6 draws the paper's conclusions, emphasizing the practical and empirical implications of the results.

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 "Our Money or Your Life"

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.

3.1 Assumptions

Graboyes (2000a) begins with the following assumptions:

(I) *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.

(II) 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.

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

(IV) 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.

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

(VI) 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.

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

(VIII) 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] $$\hat{U}_{d*} = (1-\pi_H)U(y_0 - p_{d*}; w) + \pi_H U(y_0 - p_{d*} - d*; w) - (\pi_S - \pi_H \kappa_H)k$$ $$= U(y_0 - p_{d*}; w) - (\pi_S - \pi_H \kappa_H + \pi_H \kappa_L)k$$, 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.[3] In results (4)-(9), U(·) can be replaced by ln(·). The logarithmic specification also yields the following results:

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

[11] $$d* = \frac{(f-1)(y_0 - p_H x)}{f - (f-1)p_H}$$, 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 $$\frac{1 + (\phi - 1)\pi_L}{[\phi - (\phi - 1)\pi_H]\phi^{\pi_H + \pi_L - 1}} \gtreqless 1$$ [ch.2, (P.8)]

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

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

4. The Model: Effects of Technological Progress on Optimal Insurance

The current paper extends the results from section 3 to ask how the optimal insurance policy changes as medical science makes progress in curing disease. As in results (10) through (15) above, we limit the analysis to a logarithmic utility function. This section describes how deductibles, indemnities, premiums, welfare, and choice of mode respond to changes in the cure rates for Ls and Hs (κ_L and κ_H, respectively). Under certain conditions, medical progress causes welfare to decline and may even render health insurance and treatment undesirable altogether. The effects of increasing κ_H prove to be less paradoxical than does increasing κ_L. All results described in this section are proven in the appendix.

4.1 k_L rises, mode unchanged

This section asks how the optimal insurance contract and welfare change if the insurance mode does not change. So, this section does not consider the case, for example, where an indemnity policy is optimal before κ_L rises, and a deductible policy is optimal afterward. Table 3 shows the following results:

---

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

Table 3

Changes in optimal contract parameters and welfare resulting from changes in κ_L, assuming the optimal mode remains unchanged

Mode	indemnity	deductible	premium	welfare
I: Indemnity policy	1: $\P i^*/\P k_L>0$	———	2: $\P p_{i*}/\P k_L>0$	3: $\P \hat{U}_{i*}/\P k_L<0$
D: Deductible policy	———	4: $\P d^*/\P k_L>0$	5: $\P p_d^*/\P k_L<0$	6: $\P \hat{U}_{d*}/\P k_L<0$
Z: Zero coverage	———	———	———	7: $\P \hat{U}_z/\P k_L=0$
F: Full coverage for Hs and Ls	———	———	8: $\P p_f/\P k_L=0$	9: $\P \hat{U}_f/\P k_L>0$

The numbers 1-9 preceding derivatives and statements correspond to the numbers of propositions in the Appendix.

In mode **I**, a rise κ_L increases the minimum deterrent indemnity i^* because, probabilistically, the value of treatment rises for Ls. The premium also rises because it must now cover a larger deterrent indemnity. Most importantly, welfare declines. Ls end up better off because, while they still do not benefit directly from the medical treatment, they now receive a larger indemnity to deter them. However, the increase in the Ls' welfare is more than offset by a decrease in welfare among Ws (well people) and Hs who now have to pay a larger premium to cover the larger indemnity. Like Ls, Ws and Hs receive no health benefits from the new technology. In other words, the only effect of scientific progress here is to drive individuals farther from the ideal of equal marginal utility of wealth in all health states.

In mode **D**, the deductible rises because, again, the rise in κ_L makes treatment more valuable to Ls; deterring them now requires a larger marginal cost. The premium drops this time, because Hs bear more of the cost out-of-pocket. As in the previous case, welfare declines. Ws and Ls are better off because the premiums are lower (Ls still go untreated); but these gains are more than offset by the welfare loss of the Hs, who must pay a larger share of the cost out-of-pocket.

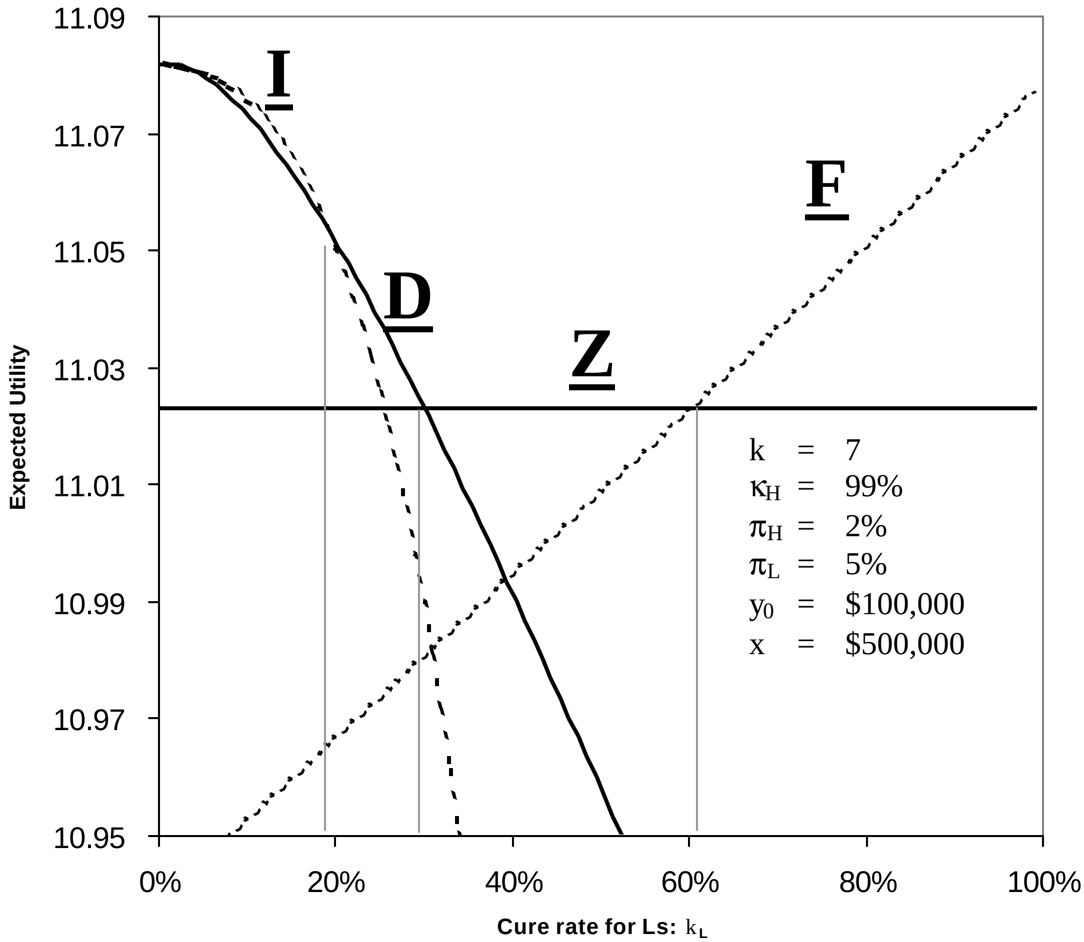
In mode **Z**, welfare is unaffected by a change in κ_L. Ls do not benefit from the new technology, and with no indemnities, deductibles, or premiums, no wealth shifts result. Finally, under **F**, the premium is unaffected because the contract includes no deterrent mechanisms. Welfare increases because, while there are no *ex post* wealth shifts, some Ls now recover because of the increased cure rate.

A real-world analog of this phenomenon might be found in rising out-of-pocket shares of pharmaceutical costs. If the efficacy of drugs rises for low-benefit patients, insurers could raise the out-of-pocket costs. In the end, the higher price continues to deter low-benefit patients, and high-benefit patients are now displeased by the higher deductible, despite the lower premium.

4.2 k_L rises, mode changes

The preceding section examined how changes in κ_L might affect welfare when there is no mode change. This section looks at how changes in κ_L might affect the optimal insurance mode (e.g., shifting the optimal class of policy from an indemnity policy to a deductible policy). Figure 2 illustrates how the optimal mode is determined. The optimal mode is always the uppermost of the four curves, so $\hat{U}$=MAX[$\hat{U}_{i*}$, $\hat{U}_{d*}$, $\hat{U}_z$, $\hat{U}_f$]. In this example (parameters at right), **I** is the preferred mode when κ_L is near 0. As κ_L rises toward 100%, **I** gives way to **D**, then **Z**, and then **F**. As κ_L rises, welfare diminishes steadily until the point that insurance is abandoned altogether. Then, eventually, full insurance becomes optimal and welfare rises once again. Note that with these particular parameters, welfare is at its maximum when κ_L is near 0. Any improvement in Ls' prospects diminishes welfare from that point.

Figure 2
Determining the Optimal Mode



Moving through the modes, welfare is a continuous function of κ_L. Marginal utility is discontinuous at mode shifts, however. Propositions 10 through 13 in the Appendix demonstrate that as κ_L rises, only certain sequences of optimal modes are possible. Specifically:

Prop. 10: *Ceteris paribus*, as κ_L rises:
Z cannot precede **D**, **Z** cannot precede **I**, **F** cannot precede **Z**, **F** cannot precede **D**, and **F** cannot precede **I**

Prop. 11: *Ceteris paribus*, as κ_L rises, **D** cannot precede **I**

Prop. 12: As $\kappa_L \rightarrow 0$, F cannot be preferred to **I** or **D**

Prop. 13: As $\kappa_L \rightarrow \kappa_H$, **D** can't be optimal

So as κ_L climbs from 0 toward κ_H, the optimal mode can only shift in a rightward manner through the modes shown in Figure 3 (e.g., **I**® **Z**® **F** is possible, but not **F**® **D** or **D**® **Z**® **I**).

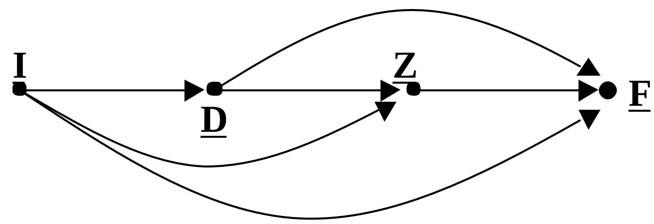


Figure 3
Succession of optimal modes as κ_L rises

I: Full coverage for Hs. Indemnity in lieu of treatment for Ls.
D: Coverage for Hs after paying deductible. No coverage for Ls.
Z: No coverage for Hs or Ls.
F: Full coverage of Hs and Ls.

Table 4 shows eleven possible optimal mode sequences. Each row contains a sequence and an arbitrary set of parameters that will generate that particular sequence. So, for example, in the seventh row of data, parameters (k, $\kappa_{H,}$ $\pi_{L,}$ π_H, y_0, x)=(6, 99%, 10%, 20%, $60,000, $180,000) will give rise to the sequence **I**® **D**® **Z** as κ_L rises from 0% to 99%:

Table 4
Possible modal sequences as κ_L rises from 0 to κ_H
(values are arbitrary examples of each sequence)

Modal sequence				κ	κ_H	p_L	p_H	y_0	x
		Z		5	99%	3%	5%	$100,000	$500,000
I		**Z**		7	83%	4%	5%	$100,000	$500,000
I			**F**	3	99%	1%	10%	$100,000	$200,000
	D	**Z**		7	99%	6%	5%	$100,000	$500,000
	D		**F**	6	99%	10%	1%	$100,000	$200,000
		Z	**F**	3.7	40%	2%	4%	$100,000	$150,000
I	**D**	**Z**		6	99%	10%	20%	$60,000	$180,000
I	**D**		**F**	3	99%	1%	4%	$60,000	$150,000
I		**Z**	**F**	4	90%	1%	9%	$100,000	$300,000
	D	**Z**	**F**	7	99%	6%	3%	$100,000	$500,000
I	**D**	**Z**	**F**	7	99%	3%	5%	$100,000	$500,000

From various entries in Tables 3 and 4, we can see that technological progress can either leave welfare higher, lower, or unchanged. In Table 4, moving through the sequence **D** ® **Z**, welfare will end lower than it began. The sequence **Z** will leave welfare unchanged. And **Z**® **F** will leave things unambiguously better off.

Any number of real-world analogs come to mind. Providers may begin offering an alternative treatment that improves the prospects for low-benefit patients. If Ls greatly outnumber Hs, then the insurer (or employer in the case of employer-based plans), may seek to exclude the alternative treatment as "experimental." Hence, we could hypothesize that the "experimental" designation is sought more often when the number of low-benefit patients is large relative to the number of high-benefit patients. Or, the insurer or employer may seek to redefine the treatment as medically unnecessary. As gastric bypasses have become effective weight-loss treatments, some insurers have sought to define some cases as "not medically necessary" and, thereby, excluded from treatment. Other procedures that insurers have recently sought to exclude include reductive mammoplasty and orthognathic surgery, which some insurers have sought to redefine as "cosmetic."[4]

4.3 κ_H changes, mode does not

This section examines how the optimal health insurance contract is affected by a change in κ_H—the probability of cure for Hs who receive treatment. Unlike the κ_L results, the κ_H results are mostly free of paradoxes—i.e., progress is usually a good thing. Table 5 summarizes the results that are proven in the appendix. The numbers 14 through 22 preceding the derivatives correspond to the numbered propositions.

Table 5

Changes in optimal contract parameters and welfare resulting from changes in κ_H, assuming the optimal mode remains unchanged

Mode	indemnity	deductible	premium	welfare
I:Indemnity policy	14: $\partial i^*/\partial \kappa_H=0$	———	15: $\partial p_{i*}/\partial \kappa_H=0$	16: $\partial \hat{U}_{i*}/\partial \kappa_H>0$
D:Deductible policy	———	17: $\partial d^*/\partial \kappa_H=0$	18: $\partial p^*/\partial \kappa_H=0$	19: $\partial \hat{U}_{d*}/\partial \kappa_H>0$
Z:Zero coverage	———	———	———	20: $\partial \hat{U}_z/\partial \kappa_H=0$
F:Full coverage for Hs and Ls	———	———	21: $\partial p_f/\partial \kappa_H=0$	22: $\partial \hat{U}_f/\partial \kappa_H>0$

The number preceding each derivative refers to the proposition that proves that cell.

[4] This paragraph's three examples were suggested by Dr. Douglas Hadley, regional medical director of an HMO.

In no case does a change in κ_H alter the contract parameters (i.e., indemnity, deductible, or premium). This is intuitive, since these parameters are set to alter behavior in Ls, not Hs. In no case does a rise in κ_H cause welfare to decline, as sometimes occurs with a rise in κ_L.

4.4 k_H changes, mode changes

Potential mode shifts from a rise in κ_H are few and simple. Proposition 23 shows that the mode can shift from **Z** to **I**, **D**, or **F**. *Ceteris paribus*, no other shifts are possible. Table 6 gives arbitrary examples in which preferences do shift from **Z** to **I**, **D**, and **F**. For example, in the final row, κ_H rises from 80% to 90%; at the former value, zero insurance is optimal, while at 90%, full insurance for Ls and Hs is optimal.

Table 6

Possible modal sequences as κ_H rises from its initial value to a new value
(values are arbitrary examples of each sequence)

Modal Sequence	initial k_H	new k_H	k	k_L	p_L	p_H	y_0	x
Z	20%	90%	3	10%	4%	5%	\$100,000	\$500,000
I	75%	90%	3	10%	4%	5%	\$100,000	\$200,000
D	30%	90%	10	10%	4%	5%	\$100,000	\$200,000
F	30%	90%	10	20%	4%	5%	\$100,000	\$200,000
Z **I**	80%	90%	7	5%	4%	5%	\$100,000	\$500,000
Z **D**	80%	90%	7	10%	4%	5%	\$100,000	\$500,000
Z **F**	80%	90%	8	75%	4%	5%	\$100,000	\$500,000

5. Bimodal Heterogeneity with Independent Changes in k_L and k_H

This paper rests on two quite restrictive postulates: (1) Sick people come in exactly two flavors, Ls and Hs, each of which shares a common cure rate; and (2) The cure rates

κ_L and κ_H vary independently. This section discusses whether these postulates are reasonable and how they might be relaxed in future papers.

The first postulate has both a qualitative and quantitative aspect. Qualitatively, it assumes that the distribution of cure rates is discrete rather than continuous. Quantitatively, it assumes that there are only two of these discrete cure rates. One future line of research will be to ask whether physicians see a continuum of cure rates *ex ante* or whether they only perceive a small number of discrete groups, as modeled here.

Assuming two cure rates, or at least a small number of discrete cure rates, may be a reasonable assumption. One reason for having discrete cure rate groups could be that diagnostic categories bundle together a countable number of different conditions. In other words, health insurance might define "lung cancer" and specify levels of financial coverage; but medically, "lung cancer" might actually refer to several distinct conditions, treated similarly, but with different rates of success. Alternatively, a countable number of physiological factors (e.g. age, gender, comorbidities) might account for the differences in cure rates. Two cancer cases may be identical in terms of indicators that comprise the diagnosis and trigger the insurance company's obligation; treatment regimes and costs may also be identical. But, for instance, the cure rates for acute lymphatic leukemia are dramatically different in juveniles and in adults. *De facto*, one could argue that acute lymphatic leukemia in a 45-year old is a different disease than acute lymphatic leukemia in a 5-year old, though the manifestation and treatment may be similar. (Alternatively, differing cure rates may be functions of differing etiologies.) This is really an incomplete contracts problem; the insurer cannot specify all possible combinations of physiological conditions.

A change in medical science's ability to treat comorbidities may change cure rates of the primary illness without affecting either the numbers of sick people or the cost of treating the illness in question. (Dealing with comorbidities might be quite inexpensive in comparison with the primary illness of concern.) Real-world analogs can be found in disease management. For example, the control of diabetes mellitus greatly affects the prognosis for a heart attack patient. Therefore, some disease management companies subsidize their patients' diabetes management program as a way of improving the prospects in the event of a heart attack.[5]

One sidenote: Changes in the incidence of comorbid diseases may yield some interesting paradoxes of their own. In the cardiac/diabetes case above, for example, suppose the deductible is set just high enough to deter Ls from seeking some hypothetical cardiac treatment. Better control over diabetes could raise the cure rate for cardiac patients, thus making the cardiac treatment more valuable to patients. An insurer covering cardiac care would find demand increasing for treatments among Ls—leading to higher deductibles or indemnities and, possibly, to abandonment of cardiac insurance altogether.[6]

As for the quantitative issue: it seems likely that a model with more than two cure rates would not behave too differently from the model developed here, since indemnities and deductibles only work through one marginal population group. It may be that a

---

[5] This example was suggested by my colleague, Dr. Richard Schieken, Chairman of the Department of Pediatric Cardiology, Medical College of Virginia Campus, Virginia Commonwealth University.

[6] It is interesting, though not necessarily realistic, to carry this argument to conspiratorial lengths. In a world of multi-illness policies, a company could subsidize diabetes control as a way of financially destabilizing the cardiac care insurer (who might be a competitor overall). We could label this phenomenon "predatory progress."

model with more than two cure rates would behave similarly to the two-rate model, but that the results would be more cumbersome mathematically.

The second postulate employed in this paper holds that κ_H and κ_L change independently. This brings us to the question of how changes in κ_L and κ_H might be related. Should we assume that they change together additively, multiplicatively, or otherwise? There are two good reasons for looking separately at the changes in κ_L and κ_H. The first reason is because we can. The effects of changes in the two are entirely separable, thus rendering the model more general. We can easily impose restrictions on the relationship between the two changes—$\Delta\kappa_L=\Delta\kappa_H$ for instance. In this case, we could simply sum the derivatives in Tables 3 and 5 to obtain the changes in contract parameters and welfare. More complex κ_L-to-κ_H relationships simply require more complex functions of the Table 3 and Table 5 data.

It is not clear, though, that real-world medical advances affect low-benefit and high-benefit patients equally, or even similarly. Should the cure rates move together additively $[(\kappa_L, \kappa_H)\rightarrow(\kappa_L+\xi, \kappa_H+\xi)]$ or multiplicatively $[(\kappa_L, \kappa_H)\rightarrow(\xi\kappa_L, \xi\kappa_H)]$? Do the rates converge over time or do they diverge? Most likely, the answers vary from disease to disease, and our model allows us to experiment with all possibilities.

One reason for avoiding the notion of a continuum of cure rates is that it is not clear how we should assume these rates change with technological innovation. Would a technological improvement raise the cure rates uniformly across the distribution? If so would it do so additively $[(\kappa_n)\rightarrow(\kappa_n+\xi)\forall n]$, multiplicatively $[(\kappa_n)\rightarrow(\xi\kappa_n)\forall n]$, or by some other rule? These are all empirical questions to be addressed in later papers.

6. Conclusion

This paper has explored ways in which one type of technological progress can affect health insurance markets. We found that improvements in medical science's ability to cure illness can, under certain theoretical conditions, lead to reduced welfare and, in some instances, to abandonment of health insurance and medical treatment altogether. In this section, we review the findings of the paper and outline some further questions that the paper raises.

6.1 Review of the Results

This paper began with a fairly general set of assumptions and then derived underlying conditions rules for specifying the optimal insurance contract. A central feature of the model is a bimodal population distribution with respect to curability. The model examines how the optimal contract changes as the cure rates for both groups change. (The two groups are Ls, who experience low cure rates when treated, and Hs, who experience high cure rates when treated.)

As long as treatment is socially undesirable for Ls (because the cost of treatment outweighs the expected benefit of treatment), higher cure rates for Ls will either diminish welfare or, at best, leave it unchanged. Rising cure rates for Hs, in contrast, never diminish welfare. A rise in either cure rate can change the mode of the optimal health insurance contract, but only in specific directions. So, for example, a rising cure rate for Ls can change the optimal contract from an indemnity contract to a deductible contract, but not the reverse. In the model, the effects of changes in κ_L and κ_H are entirely separable.

In sum the paper provides a set of rules enumerating how the optimal insurance contract can and cannot change, depending on parameters describing wealth, medical technology, population characteristics, and the welfare effects of disease.

6.2 Implications and Further Research

The findings of this paper suggest a number of directions for continued research. The results outlined above have implications with respect to the behavior of insurance markets, research and development and public policy, and health care delivery and overall expenditures. Following are some of the questions that the paper suggests, each of which can be the subject of future research:

Does medical progress actually drive away insurance? Insurers don't generally announce that they are dropping coverage of a particular disease. The market may do so, however, by subtle means. Utilization review can exclude low-benefit patients, though that may not be the explicitly stated purpose. And excluded-by-means-of-UR may be only imperfectly correlated with low-benefit. In a legal environment where laws and courts frown on certain kinds of discrimination, less-than-perfect correlation between the determinants and the exclusion may be good.

Refining the definitions of diagnostic categories is one way of excluding patients whose conditions might previously have been covered by way of “DRG creep” (stretching a diagnosis so that a patient falls within a covered category). In Table 2, rather than raising the deductible so high, it might be better to simply redefine Ls as having an altogether different diagnosis—one whose treatments are not covered, assuming that redefinitions are feasible. (This solution may entail costly monitoring, which we have omitted from this analysis.)

Restricted supply of health care may be a means of discouraging low-benefit patients from receiving treatments. Anecdotally, there are stories of insurers who routinely refuse claims, knowing that some patients (presumably low-benefit ones) will be driven away by the nuisance costs of pursuing a rightful claim. In the case of end-stage care, queuing with a non-random position in line would obviously be a way to tilt treatments toward high-benefit benefits. While a particular insurer may not explicitly drop coverage for illnesses, employers may shift their group policies to insurance plans with exclusionary clauses.

If medical progress doesn't drive away insurance, then why not? One explanation may lie in the public choice domain; people who share a given illness may be especially good at lobbying political authorities for continued coverage. This might help explain why health insurance tends to bundle together the risks of many diseases. (Adverse selection is another factor that may induce bundling.) Some treatments will inevitably become too popular (and hence financially burdensome on subscribers and insurers). If we know in advance that insurers can't drop coverage, then it makes sense to bundle risks of many diseases, knowing that some illnesses will turn out to be financial winners and others financial losers for the insurer (and, indirectly, for the subscribers). This does raise questions, though, about whether the markets with temporary excess rates of profit are contestible.

In the real world, do rising cure rates actually lead to higher deductibles, higher indemnities or abandonment of health insurance coverage? If κ_L rises so that optimal contracts pass through the sequence <u>**I**</u>→<u>**D**</u>→<u>**Z**</u>→<u>**F**</u>, the model predicts a rising indemnity, followed by a shift to a rising deductible, followed by abandonment of insurance, followed by full insurance. The empirical question is whether anything like this sequence actually occurs in health insurance markets. In fact, we generally do not see deductibles or indemnities continually adjusting, perhaps because there are costs associated with such repricing—or even varying the deductible or indemnity from disease to disease. (Explaining changes or complex schedules to consumers might entail significant costs.) So, costly repricing might suggest deductibles or indemnities set to maximize insurers' profits over time, rather than at the instant. So over time, the deductible or indemnity would initially be set higher than and then lower than levels that are socially desirable in the short run.

Do costs associated with switching modes influence insurers to stick with more stable contract modes, even though they may be less efficient in the short-run? This is a variation on the previous issue. Suppose an indemnity contract is optimal today but insurers anticipate a rise in κ_L that will render a deductible contract superior in the future. Will insurers avoid the indemnity contract today, rather than endure the costs of switching to a deductible policy in the future? The dynamics of such strategies could be quite complex, incorporating discount rates and expected technology growth rates.

Are indemnities seldom seen in actual contracts because of problems suggested by this model? Have rising cure rates led to switches from indemnities to other contract modes? Graboyes (2000a) reviewed some evidence that indemnity contracts were once far more common than they are today. In the model presented in this paper, a narrowing of the cure rates between Ls and Hs tend to render indemnity contracts suboptimal. The question, then, is whether these dynamics are, in fact, responsible for the disappearance of such indemnity contracts. In other words, the model suggests that wide variation in cure rates between different population groups makes indemnities more attractive than deductibles. As cure rates converge and populations become more homogeneous with respect to curability, indemnity contracts should give way to deductible contracts. One question to explore is whether the disappearance of indemnities coincided with a lessening of variability of cure rates for various diseases.

Can the effects described in the model skew the directions of medical research? The model predicts that in some instances, advances in medical technology will not be used. Improved treatments for Ls will benefit no patients (and hence, no providers) if insurance continues to exclude Ls from treatment. In such cases, the developer of the new technology would get nothing from his work. So, knowledge of future market conditions may skew research toward improvements in the treatment of Hs or toward improvements in markets where Ls do get treated. For example, an expected sequence of **D**→**Z** might yield very different direction for R&D than would an expected sequence of **D**→**F**. With **D**→**F** (which is the scenario described earlier in conjunction with Table 2), the researcher who develops the new and improved technology would never gain financially from it, because it is never actually used. Empirically, we might expect to see more progress in markets where the numbers of Ls are small, rather than markets with many Ls.

This triage-via-R&D raises some normative questions related to equity: If R&D is publicly financed, would it be desirable to help Hs and effectively withhold the technology available to Ls? And if not, is some sort of second-best corrective mechanism called for (e.g., a subsidy for Ls wishing treatment)?

How does the structure of the optimal contract (and of optimal contracts anticipated in the future) affect the diffusion of medical technology? Weisbrod (1991) noted that diffusion of medical technology through society depends in part upon the structure of health insurance. Suppose **D** optimal at present. Hs are receiving treatment, but a rising κ_L will soon render **Z** optimal. Once this happens, the model suggests that this medical procedure will be abandoned. This is because *ex ante*, no one will wish to pay the premium necessary to insure against the disease and *ex post* no one will be able to afford the treatment. Here, technological progress will result in a sort of negative diffusion—the disappearance of the procedure. If insurers maximize over time, rather than instantaneously, we might see a tendency to avoid procedures that will one day become uneconomical to provide. The empirical question is whether we can predict such biases from environmental factors (population distributions, treatment costs, etc.).

Additional research questions: Finally, this paper has focused on the effects of progress in one area (curative procedures) of medical technology. We can equivalently ask how optimal contracts change as we perturb other environmental parameters. These would include improvements in preventive medicine (a reduction in π_L or π_H), palliative techniques (a reduction in k), production efficiency (a reduction in x), and overall wealth (a rise in y_0).

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. 2000c. Medicine worse than the malady: Cure rates, population shifts, and health insurance. *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 four sets of proofs. Section A.1 demonstrates how changes in κ_L affect indemnities, deductibles, premiums, and welfare under each of the four relevant modes. Section A.2 demonstrates that changes in κ_L cannot induce certain mode sequences. A.3 demonstrates how changes in κ_H affect indemnities, deductibles, premiums, and welfare under each of the four relevant modes. Section A.4 demonstrates the mode sequences that changes in κ_H cannot induce. 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).

A.1 k_L changes, mode does not

This section proves the nine numbered derivatives from Table 3. The numbers preceding each derivative correspond to Propositions 1 through 9 below.

Proposition 1: If k_L rises, the optimal indemnity i* rises [¶i*/¶k_L>0]. Recall:

[10] $$i^* = \frac{(f-1)(y_0 - p_H x)}{1+(f-1)p_L}, \text{ where } f = e^{k_L k}. \qquad \text{So,}$$

(1.1) $$\frac{¶ i^*}{¶ k_L} = \left[\frac{¶ i^*}{¶ f}\right] \times \left[\frac{¶ f}{¶ k_L}\right] = \left[\frac{(y_0 - p_H x)}{[1+(f-1)p_L]^2}\right] \times \left[k e^{k_L k}\right] > 0$$

This is positive, since both bracketed terms are positive. [Q.E.D.]

Proposition 2: If k_L rises, then the premium p_{i*} rises [¶p_{i*}/¶k_L>0]. The premium covers both the cost x of treating Hs and the indemnity i* paid to Ls in lieu of treatment:

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

κ_L only enters the right-hand side of this expression through i*, so

$$(2.1)\qquad \frac{\partial p_{i*}}{\partial \kappa_L} = \frac{\partial p_{i*}}{\partial i*}\frac{\partial i*}{\partial \kappa_L} = \pi_L \frac{\partial i*}{\partial \kappa_L} > 0$$

This is positive because $\pi_L>0$ and, from Proposition 1, $\partial i*/\partial\kappa_L>0$. [Q.E.D.]

Proposition 3: If k_L rises, welfare under the optimal indemnity contract declines [¶$\hat{U}_{i*}$/¶k_L<0]. In logarithmic form, $\hat{U}_{i*}$, from [5] becomes:

$$(3.1)\qquad \hat{U}_{i*} = (1-\pi_L)\,ln(y_0 - p_{i*}) + \pi_L\, ln(y_0 - p_{i*} + i*) - (\pi_S - \pi_H\kappa_H)k \text{, where}$$

$$p_{i*} = \pi_H x + \pi_L i*.$$

Now, differentiating:

$$(3.2)\qquad \frac{\partial \hat{U}_{i*}}{\partial \kappa_L} = -\frac{1-\pi_L}{y_0 - p_{i*}}\frac{\partial p_{i*}}{\partial i*}\frac{\partial i*}{\partial \kappa_L} - \frac{\pi_L}{y_0 - p_{i*} + i*}\frac{\partial p_{i*}}{\partial i*}\frac{\partial i*}{\partial \kappa_L} + \frac{\pi_L}{y_0 - p_{i*} + i*}\frac{\partial i*}{\partial \kappa_L}$$

Dividing the right-hand side by $\partial i*/\partial\kappa_L$ and $\partial p_{i*}/\partial i*=\pi_L$ leaves the sign unchanged, so:

$$(3.3)\qquad \begin{aligned} sgn\left\{\frac{\partial \hat{U}_{i*}}{\partial \kappa_L}\right\} &= sgn\left\{-\frac{1-\pi_L}{y_0 - p_{i*}} - \frac{\pi_L}{y_0 - p_{i*} + i*} + \frac{1}{y_0 - p_{i*} + i*}\right\} \\ &= sgn\left\{-\frac{1-\pi_L}{y_0 - p_{i*}} + \frac{1-\pi_L}{y_0 - p_{i*} + i*}\right\} < 0 \end{aligned}$$

Since $y_0-p_{i*}<y_0-p_{i*}+i*$, this is negative. A rising cure rate for Ls *decreases* welfare.

[Q.E.D.]

Proposition 4: If k_L rises, the optimal deductible d* rises [¶d*/¶k_L<0]. Recall:

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

Differentiating:

$$(4.1)\qquad \frac{\partial d*}{\partial \kappa_L} = \left[\frac{\partial d*}{\partial \phi}\right]\times\left[\frac{\partial \phi}{\partial \kappa_L}\right] = \left[\frac{(y_0 - \pi_H x)}{\phi-(\phi-1)\pi_H}\right]\times\left[ke^{\kappa_L k}\right] > 0$$

This is positive because both bracketed terms are positive, as can be seen by inspection.

[Q.E.D.]

Proposition 5: If k_L rises, the premium p_{d*} falls [$\P p_{d*}/\P k_L<0$]. The premium covers the cost x of treating Hs, minus the deductible d* paid out-of-pocket by patients:

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

κ_L only enters the right-hand side of this expression through d*, so

$$\frac{\partial p_{d*}}{\partial \kappa_L}=\frac{\partial p_{d*}}{\partial d^*}\frac{\partial d^*}{\partial \kappa_L}=-\pi_H\frac{\partial d^*}{\partial \kappa_L}<0 \tag{5.1}$$

This is negative because $\pi_H>0$ and, from Proposition 4, $\partial d^*/\partial \kappa_L>0$. The premium falls because more of the treatment cost is being borne out-of-pocket by patients. [Q.E.D.]

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

$$\hat{U}_{d*}=(1-\pi_H)\ln(y_0-p_{d*})+\pi_H\ln(y_0-p_{d*}-d^*)-(\pi_S-\pi_H\kappa_H)k, \tag{6.1}$$

where $p_{d*}=\pi_H x-\pi_H d^*$. Differentiating:

$$\frac{\P\hat{U}_{d*}}{\P k_L}=-\frac{1-p_H}{y_0-p_{d*}}\frac{\P p_{d*}}{\P d^*}\frac{\P d^*}{\P k_L}-\frac{p_H}{y_0-p_{d*}-d^*}\frac{\P p_{d*}}{\P d^*}\frac{\P d^*}{\P k_L}-\frac{p_H}{y_0-p_{d*}-d^*}\frac{\P d^*}{\P k_L} \tag{6.2}$$

Dividing the right-hand side by $\partial d^*/\partial \kappa_L$ and $-\partial p_{d*}/\partial d^*=\pi_H$ leaves the sign unchanged, so:

$$\begin{aligned} sgn\left\{\frac{\partial \hat{U}_{d*}}{\partial \kappa_L}\right\}&=sgn\left\{\frac{1-\pi_H}{y_0-p_{d*}}+\frac{\pi_H}{y_0-p_{d*}-d^*}-\frac{1}{y_0-p_{d*}-d^*}\right\}\\ &=sgn\left\{\frac{1-\pi_H}{y_0-p_{d*}}-\frac{1-\pi_H}{y_0-p_{d*}-d^*}\right\}<0 \end{aligned} \tag{6.3}$$

Since $y_0-p_{d*}>y_0-p_{d*}-d^*$, this is negative. A rising cure rate for Ls *decreases* welfare. [Q.E.D.]

Proposition 7: If k_L rises, then welfare under the zero-insurance mode is unchanged [$\P \hat{U}_z/\P k_L=0$]. By assumption, treatment is prohibitively expensive without insurance. So, in zero-insurance mode, changes in κ_L will have no effects on *ex post* income or on welfare.

[7] $\hat{U}_z = \pi_W U(y_0;w) + (\pi_H + \pi_L)[U(y_0;w) - k] = U(y_0;w) - \pi_S k$

κ_L does not enter into this equation, so $\partial\hat{U}_z/\partial\kappa_L=0$. Welfare is unaffected by changes in κ_L. Also, with no insurance, there are no indemnities, deductibles, or premiums to consider. [Q.E.D.]

Proposition 8: If k_L rises, the premium under the full-insurance mode is unchanged [¶p_f/¶k_L=0]. From [8] the full-insurance premium equals to cost of treating all Hs and Ls:

[8] $p_f = p_S x$

κ_L does not enter into this equation, so changes in κ_L leave the premium unaffected. [Q.E.D.]

Proposition 9: If k_L rises, then welfare under the full-insurance mode rises [¶$\hat{U}_f$/¶k_L>0]. Here, all Hs and Ls are treated, and all subscribers pay an equal premium $p_f=\pi_S x$. Since Ls are treated, a higher cure rate κ_L among that group improves welfare by increasing the number of Ls who recover from the disease. Proposition 8 tells us that there are no changes in premiums; thus, there are no changes in *ex post* financial position of any agents. Recall:

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

When κ_L rises, the only change in welfare comes from the increased cure rate among Ls, so:

(9.1) $\partial\hat{U}_f/\partial\kappa_L=\pi_L k>0$

We are treating all sick people, so there is no need to compensate anyone for forgoing treatment. Thus, *ex post*, marginal utility of wealth is uniform across individuals. The full-insurance mode is the only one where a rising cure rate for Ls *increases* welfare. This makes intuitive sense, since it is the only mode in which Ls actually get to take advantage of medical science's greater capacity to cure their illnesses. [Q.E.D.]

A.2 k_L changes, mode changes

Section A.1 demonstrated how changes in κ_L affect contract parameters and welfare if the mode remains unchanged. In contrast, this section explores how changes in κ_L can affect which mode is optimal. Given a set of primitive assumptions, changes in κ_L can lead the market through a sequence of modes, but only in particular orders. Figure 2 illustrates how modes change, and Table 4 shows the set of possible (and impossible) mode sequences. Propositions 10-13 here demonstrate why certain sequences are impossible, as follows:

10: *Ceteris paribus*, as κ_L rises:
- **Z** cannot precede **D**
- **Z** cannot precede **I**
- **F** cannot precede **Z**
- **F** cannot precede **D**
- **F** cannot precede **I**

11: *Ceteris paribus*, as κ_L rises, **D** cannot precede **I**
12: As $\kappa_L \rightarrow 0$, F cannot be preferred to **I** or **D**
13: As $\kappa_L \rightarrow \kappa_H$, **D** can't be optimal

Figure 2 shows how the optimal insurance contract is determined for different κ_L. The optimal contract has utility of $\hat{U}=MAX[\hat{U}_{i*}, \hat{U}_{d*}, \hat{U}_z, \hat{U}_f]$. In Figure 2, for very small κ_L, $\hat{U}_{i*}$ is higher than welfare under the other the other three contract modes so we begin in mode **I**. Then, after the **I**-curve cuts the **D**-curve from above, the deductible contract is preferable to the indemnity contract. Then the **D**-curve cuts the **Z**-curve from above, and zero insurance becomes optimal. Finally, the **Z**-curve cuts the **F**-curve from above, and full insurance is optimal.

Each curve is a continuous function of κ_L, so $\hat{U}$ is also a continuous function of κ_L. Thus, a change of mode is not marked by a discrete change in welfare. And if there is a switch from one mode to a second, it must be true that the slope of the second curve is greater than the slope of the first curve at the point where the modes switch. Knowing this allows us to rule out five possible mode switches in Proposition 10 and another switch in Proposition 11.

Proposition 10: A rise in k_L cannot lead to a shift from Z (zero-insurance) to D (deductible policy) or I (indemnity policy); nor can a rise in k_L lead to a shift from F (full insurance) to Z (zero insurance), D (deductible policy) or I (indemnity policy). The proofs are evident from Table 3 and illustrated by Figure 2. The **Z**-curve is always flat, while the **D**- and **I**-curves are always downward-sloping. Thus the **Z**-curve can never cut the **D**- and **I**-curves from above, so the mode will never switch from **Z** to **D** or to **I**. Similarly, the **F**-curve is upward-sloping and, so, can never cut the **Z**-, **D**-, or **I**-curves from above; a rise in κ_L can never lead to an abandonment of full insurance. [Q.E.D]

Proposition 11: A rise in k_L cannot lead to a shift from D (deductible policy) to I (indemnity policy). This proposition requires a lengthy proof consisting of 5 lemmae and is derived from a slightly rearranged version of the deductible/indemnity boundary:

[14] $$\frac{[1+\pi_L\phi-\pi_L]}{[\phi(1-\pi_H)+\pi_H]\times[\phi^{\pi_H+\pi_L-1}]} \gtreqless 1 \quad \text{iff} \quad \hat{U}_{d^*} \gtreqless \hat{U}_{i^*}$$

If $\pi_L \geq \pi_H$, the best indemnity policy will never be preferred to the best deductible policy. So, we limit our analysis here to the case where $\pi_L < \pi_H$. The indemnity policy is preferred to a deductible policy if κ_L is small, and the deductible policy may or may not be preferred to the indemnity policy at higher levels of κ_L.

Notation

For ease of notation, we redefine the bracketed expressions in [14] as u, v, and w:

(11.1) $$\frac{[1+p_L f - p_L]}{[f(1-p_H)+p_H]\times[f^{p_H+p_L-1}]} = \frac{u}{v \times w}$$

Using this expression, we prove the five lemmae that comprise this proposition.

Throughout this proof, we adopt the convention that q'= $\partial q / \partial f$ and q?= $\partial^2 q / \partial f^2$ for any function q(ϕ).

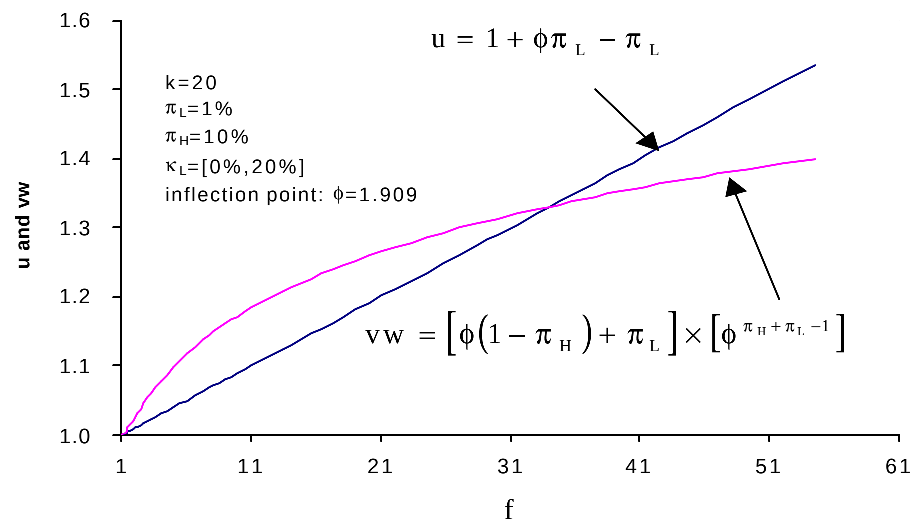
Graphic Description

Figure 4 represents the five lemmae in the case where $\pi_L < \pi_H$.

Lemma 1: If k_L=0, then u=vw=1.
Lemma 2: u is linear and upward-sloping in f, with a slope of p_L.
Lemma 3: vw is upward-sloping in f.
Lemma 4: $\lim_{\phi \to 1} (vw)' = u' = \pi_L$.
Lemma 5: Over the relevant range f Î [1,¥), there exists at most one value $\tilde{\phi}$ where (vw)?=0. If such a point exists, (vw)?>0 for $\phi < \tilde{\phi}$ and (vw)?<0 for $\phi > \tilde{\phi}$.

In Figure 4, the inflection point occurs at $\tilde{\phi}$=1.909. So, as ϕ rises from its lowest value of 1, vw initially rises at an increasing rate, while u rises at a constant rate. Thus, an epsilon beyond ϕ=1, vw>u—meaning that an indemnity policy is welfare-superior to a deductible policy. After ϕ passes the inflection point at ϕ=1.909, the slope of vw begins to decline. In this particular case, vw eventually cuts u from above; to the right of the intersection, a deductible policy is preferred to an indemnity policy.

Figure 4
$p_L < p_H$



Algebraic Proof

To prove the propositions below here, it is useful to know the following relationships:

Table 7
Relationships between u, v, w, and their derivatives

$u = 1+\phi\pi_L - \pi_L$ $= 1+\phi u' - u'$	$u' = p_L$	$u'' = 0$
$v = \phi(1-\pi_H)+\pi_H$ $= \phi v' - (1-v')$	$v' = 1-p_H$	$v'' = 0$
$w = f^{p_H+p_L-1}$	$w' = (p_H+p_L-1)f^{p_H+p_L-2}$ $= (u'-v')wf^{-1}$	$w'' = (p_H+p_L-2)(p_H+p_L-1)f^{p_H+p_L-3}$ $= (u'-v'-1)w'f^{-1}$ $= (u'-v'-1)(u'-v')w'f^{-2}$

Now we demonstrate the five lemmae:

Lemma 1: If k_L=0, then u=vw=1. $\phi=\exp(\kappa_L k)$, so if $\kappa_L=0$, $\phi^n=1$ for all n. If $\phi=1$,

(11.2) $u=[1+\pi_L\phi-\pi_L]=[1+\pi_L(1)-\pi_L]=1$, and

(11.3) $vw = \left[\phi(1-\pi_H)+\pi_H\right]\times\left[\phi^{\pi_H+\pi_L-1}\right] = \left[1(1-\pi_H)+\pi_H\right]\times[1] = 1$

Lemma 2: u is linear and upward-sloping in f, with a slope of p $_L$. This is evident from inspection, since, as the table above shows, u'=π_L and u?=0.

Lemma 3: vw is upward-sloping in f. The slope of vw is:

(11.4) $(vw)' = v'w + vw'$

Into this, substitute $w' = (u'-v')w\phi^{-1}$ from Table 7:

(11.5)
$$\begin{aligned}(vw)' &= v'w + v(u'-v')w\phi^{-1} \\ &= [v'w]\left[1-v\phi^{-1}\right]+\left[vu'w\phi^{-1}\right] > 0\end{aligned}$$

This is positive because both bracketed terms are positive. vw slopes upward. [Q.E.D.]

Lemma 4: $\lim_{\phi\to 1}(vw)' = u' = \pi_L$. From Table 7, restate the previous equation as:

(11.6)
$$\begin{aligned}(vw)' &= v'w + v(u'-v')w\phi^{-1} \\ &= (1-\pi_H)\phi^{\pi_H+\pi_L-1} + \left[\phi(1-\pi_H)+\pi_H\right](\pi_H+\pi_L-1)\phi^{\pi_H+\pi_L-2},\end{aligned}$$ and

(11.7)
$$\begin{aligned}\lim_{\phi\to 1}(vw)' &= (1-\pi_H)\phi^{\pi_H+\pi_L-1} + \left[\phi(1-\pi_H)+\pi_H\right](\pi_H+\pi_L-1)\phi^{\pi_H+\pi_L-2} \\ &= (1-\pi_H)+\left[1-\pi_H+\pi_H\right](\pi_H+\pi_L-1) \\ &= \pi_L\end{aligned}$$

So, the slope of vw is equal to the slope of u at ϕ=1.

Lemma 5: Over the relevant range f Î [1,¥), there exists at most one value $\tilde{\phi}$ where (vw)?=0. If such a point exists, (vw)?>0 for $\phi < \tilde{\phi}$ and (vw)?<0 for $\phi > \tilde{\phi}$. In general:

(11.8) (vw)?=v?w+v'w'+v'w'+vw?

From Table 7, we know that v?=0, so:

(vw)?=2v'w'+vw?, so

(11.9) $(vw)'' \overset{>}{\underset{<}{=}} 0 \text{ iff } vw'' \overset{>}{\underset{<}{=}} -2v'w'$ [Q.E.D.]

Proposition 12: As k_L ® **0 (or** f ® **1),** $\hat{U}_{i*} > \hat{U}_f$ **and** $\hat{U}_{d*} > \hat{U}_f$**. Thus, as** k_L **climbs from 0 to** p_H**, the corresponding sequence of modes cannot begin with F̲.**

[5] $\hat{U}_{i*} = U(y_0 - p_{i*}; w) - (\pi_S - \pi_H \kappa_H - \pi_L \kappa_L)k$, where $p_{i*} = p_H x + p_L i*$, and

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

Subtracting,

(12.1) $\hat{U}_{i*} - \hat{U}_f = U(y_0 - \pi_H x - \pi_H i*; w) - U(y_0 - \pi_H x - \pi_H x; w)$

If i*<x, this equation is positive, so **I̲** is preferred to **F̲**. We can make κ_L arbitrarily close enough to 0 (and, therefore, ϕ close enough to 1) that i*<x. So,

(12.2) $\lim_{\kappa_L \to 0} \hat{U}_{i*} - \hat{U}_f = U(y_0 - \pi_H x; w) - U(y_0 - \pi_H x - \pi_L x; w) > 0$

At very low cure rates for Ls, it must be true that an indemnity policy is preferred to full insurance [or $\hat{U}_{i*} > \hat{U}_f$]. Of course, parameters may be such that the range of κ_L where this is so is trivially small. A similar proof would show that **D̲** is also preferred to **F̲** as κ_L approaches 0, so full insurance is not better than the third in the preference ranking among modes. We leave this proof to the reader. [Q.E.D.]

Proposition 13: As k_L ® k_H**,** $\hat{U}_{d*} < \hat{U}_z$**. Thus, as** k_L **approaches** k_H**, at least one mode (Z̲) will be preferred to D̲.** Recall:

[7] $\hat{U}_{d*} = U(y_0 - p_{d*}; w) - (p_S - p_H k_H + p_H k_L)k$, where $p_{d*} = \pi_H x - \pi_H d*$,

and

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

From these equations, we can deduce that:

(13.1) $$\lim_{\kappa_L \to \kappa_H} \hat{U}_z - \hat{U}_{d*} = \left[U(y_0; w) - \pi_S k\right] - \left[U(y_0 - p_{d*}; w) - \pi_S k\right] > 0$$

So, as κ_L approaches κ_H, it must be true that $\hat{U}_z > \hat{U}_{d*}$. Thus, the sequence will not conclude with a deductible policy as κ_L rises toward κ_H. Experimentation with numbers suggests that a sequence cannot conclude with **I**, but we leave that proof for a later paper.

A.3 k_H changes, mode does not

This section proves the nine numbered cells from Table 5, restated here in equation form. Each cell is a derivative showing how a change in κ_L affects the optimal indemnity, deductible, premium, and welfare under each mode The numbers preceding each derivative correspond to Propositions 14 through 22 proven below.

Proposition 14: If k_H rises, the optimal indemnity i* is unchanged [$\partial i^*/\partial k_H=0$]. Recall:

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

Since κ_H does not enter into this expression, therefore, changes in κ_H do not affect i*. [Q.E.D]

Proposition 15: If k_H rises, the premium p_{i*} is unchanged [$\partial p_{i*}/\partial k_H=0$]. Recall:

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

Again, κ_H does not enter into this expression, so the premium is unchanged. [Q.E.D]

Proposition 16: If κ_H rises, welfare under the optimal indemnity contract rises $[\partial \hat{U}_{i*}/\partial \kappa_H > 0]$. Recall:

[5] $$\hat{U}_{i*} = U(y_0 - p_{i*}; w) - (\pi_S - \pi_H\kappa_H - \pi_L\kappa_L)k$$, where $p_{i*} = \pi_H x + \pi_L i*$,

so

(16.1) $$\frac{\partial \hat{U}_{i*}}{\partial \kappa_H} = \pi_H k > 0$$

Thus, under **I**, an improved ability to cure Hs *increases* welfare. [Q.E.D]

Proposition 17: If κ_H rises, the optimal deductible d* is unchanged $[\partial d*/\partial \kappa_H = 0]$. Recall:

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

Since κ_H does not enter into this expression, therefore, changes in κ_H do not affect d*. [Q.E.D]

Proposition 18: If κ_H rises, then the premium p_{d*} is unchanged $[\partial p_{d*}/\partial \kappa_H = 0]$. The premium covers the cost x of treating Hs minus the deductible d* paid out-of-pocket:

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

Again, κ_H does not enter into this expression, so the premium is unchanged . [Q.E.D]

Proposition 19: If κ_H rises, welfare under the optimal indemnity contract rises $[\partial \hat{U}_{d*}/\partial \kappa_H > 0]$. Recall:

[6] $$\hat{U}_{d*} = U(y_0 - p_{d*}; w) - (\pi_S - \pi_H\kappa_H + \pi_H\kappa_L)k$$, where $p_{d*} = \pi_H x - \pi_H d*$,

so

(19.1) $$\frac{\partial \hat{U}_{d*}}{\partial \kappa_H} = \pi_H k > 0$$

Thus, under **D** an improved ability to cure Ls *increases* welfare. Note, that this derivative is identical to that in the case of an indemnity contract. [Q.E.D]

Proposition 20: If k_H rises, then welfare under the zero-insurance mode remains unchanged [¶$\hat{U}_z$/¶k_H=0]. With zero insurance, expected utility is:

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

κ_H does not appear anywhere in this equation, so $\partial\hat{U}_z/\partial\kappa_H=0$. [Q.E.D]

Proposition 21: If k_H rises, the premium under the full-insurance mode rises [¶$\hat{U}_f$/¶k_H>0]. From [8] the full-insurance premium is the cost of treating all Hs and Ls:

[8] $$p_f = p_S x$$

κ_H does not enter into this equation, so changes in κ_H leave the premium unaffected. [Q.E.D]

Proposition 22: If k_H rises, then welfare under the full-insurance mode rises [¶$\hat{U}_f$/¶k_H>0]. Recall:

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

From Proposition 21, a rise in κ_H leaves the premium unaffected. So the only change in welfare comes from the increased cure rate among Hs:

(22.1) $\partial\hat{U}_f/\partial\kappa_H=\pi_H k>0$

All sick people are treated, so there is no need to compensate anyone for forgoing treatment. *Ex post*, marginal utility of wealth is uniform across individuals. Note that this derivative is identical to those in Propositions 16 and 19. [Q.E.D]

A.4 κ_H changes, mode changes

This section specifies conditions under which a change in κ_H can give rise to a change in the mode of the optimal contract. As in section A.2, a change in mode requires that the first curve cut the second curve from above.

Proposition 23: A rise κ_H can shift the optimal contract mode from Z to I, D, or F. *Ceteris paribus*, no other shifts are possible. We will prove the first part of this proposition by providing numerical examples of shifts from Z to I, D, and F. First, however, we will easily dispense with the second part of the proposition.

We know from Propositions (17), (20), and (23) that:

(23.1) $$\frac{\partial \hat{U}_{i*}}{\partial \kappa_H} = \frac{\partial \hat{U}_d}{\partial \kappa_H} = \frac{\partial \hat{U}_f}{\partial \kappa_H} = p_H k > 0$$

From Proposition (21), we know that:

(23.2) $$\frac{\partial \hat{U}_z}{\partial \kappa_H} = 0$$

Limitations on possible shifts can easily be seen from these two equations. These equations do not allow us to rule out shifts from Z to I, D, and F as κ_H rises; under these circumstances, the relative attractiveness of I, D, and F rises while that of Z does not. On the other hand, changes in I, D, and F will be identical, since the derivatives of all three utility functions are identical. In other words, a rise in κ_H will improve the expected welfare of I, D, and F identically, thus leaving the preference ordering among these three modes unchanged. [Q.E.D]