Guaranteed Renewability in Insurance

MARK V. PAULY
Health Care Systems Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

HOWARD KUNREUTHER
Operations & Information Management Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

RICHARD HIRTH
Department of Health Management and Policy, University of Michigan, Ann Arbor, MI 48109-2029

Abstract

We propose a guaranteed renewability (GR) insurance in which a sequence of premiums would enable insurers to break even and would be chosen by both low- and high-risk buyers, whether or not they had suffered a loss. The premium schedule would continually decline over time, as the insurer collects more information to determine who the low-risk buyers are. The highest premiums are charged initially to protect the insurer if low-risk individuals leave for the spot market. The concluding portion of the article discusses the limitations of a GR policy in the health and environmental liability area, the most serious being instability in estimates of underlying loss trends.

Key words: uncertainty, insurance, renewability, competition

Introduction

Insurances of many types typically cover some risk for a limited time period. For instance, health insurance contracts conventionally provide coverage against loss for a year, although longer and shorter terms are possible. Over that time period, the premium is known with certainty. At the beginning of the next period, however, the factors which determine the level of the premium that an insurance purchaser is charged may change. The probability of loss may change, or the amount of the loss may also change. In particular, the person is subject to the risk that the premium at the beginning of the second year may be higher or lower than in the previous one.

In some cases, the change in premium can be dramatic. In the case of individual health insurance, for example, the discovery of an indicator of chronic disease can cause a person's insurance premiums to jump substantially. Sudden jumps in premiums because of changes in perceived risk levels are less characteristic of other policies, but they do happen for liability and automobile insurance. In fact, in some situations, such as environmental pollution, the next period's premium may be so high that it is equivalent to not

offering coverage. At the most simplistic level, sudden and unexpected changes in insurance premiums represent an additional risk to lifetime or long-term wealth of consumers and owners of businesses. Moreover, for health, auto, or environmental liability, substantial and unexpected surges in premiums can raise public policy questions, especially if higher premiums discourage coverage or have substantial impacts on wealth, and if coverage is necessary to finance health services or repair environmental damage. Indeed, a perception that changes in underwriting practices have seriously affected the small group and individual health insurance markets has caused politicians, insurance regulators, and trade associations to propose new regulations and practices for those markets. In some cases, the possibility of premium changes for high-risk individuals and small groups forms part of the justification for tax-financed national health insurance. These regulations all move toward some form of limited community rating, in which per-period premiums are required to be less dependent on known risk factors, but rather are averaged across a set of distinguishable risk categories.

This article will investigate the question of how idealized competitive markets function in situations in which expected losses change over time and insurance buyers are risk averse. We show that, if consumers and insurers are equally well informed, the competitive market can deal optimally with changes in risk levels in certain circumstances. We also show that, as usual, imperfect information or asymmetry of information can lead to problems. Finally, we show that, in a world with what we call "market level" uncertainty, competitive insurance markets may nevertheless still have some problems.

1. Why do premiums change?

We model the process of premium determination when loss probabilities change in an exceedingly simple way. We suppose initially that the premium for some amount of coverage of a risk of some type is a simple multiple (markup) of the expected value of expense. This is a plausible assumption if losses are numerous and independent, if there is no asymmetry of information, and if insurance markets are competitive. The premium P_i for person i is then simply

$$P_i = E_i(L)(1 + \lambda), \tag{1}$$

where $E_i(L)$ is the expected value of the loss to be experienced by person i, and λ is the "loading" or administrative cost percentage of expected losses. We further simplify, without loss of severality, by assuming initially that $\lambda = 0$ so that the premium is actuarially fair. The loss expectation can be written as

$$E_i(L) = p_iL, \tag{2}$$

where L is the fixed level of the insured loss, and p_i is the probability that a loss will occur. In this simple formulation, anything that causes p_i to change will cause P_i to change. (This is true regardless of the level of loading.)

How would risk-averse persons view the possibility that P_i could change if p_i changed? When P_i changes unexpectedly, a risk-averse person potentially suffers fluctuations in his lifetime income, since the changed future series of premiums alters his income available for other consumption. If the potential loss from premium increases is substantial enough in terms of wealth, a risk-averse person may seek to insure in some fashion against the loss in lifetime wealth. How is such "coverage" to be provided, and what level will be desired?

A critical factor in answering this question is, not unexpectedly, the degree of asymmetry of information between insurance buyer and insurance seller. Here we initially assume no asymmetry. When the event which indicates a change in risk level occurs, it is simultaneously detected by insured and insurer. There is thus no question of adverse selection. With no loading, no moral hazard, and no adverse selection, the existence of some type of optimal competitive equilibrium should not be in question. The interesting question still remains as to the form that this equilibrium will take, which is the subject of this article.

Since insurers typically only observe claims and premium payments by the insured, how can the insurer have the same information as the insured? One simple way to make this assumption plausible is to further assume that filing a claim in one period perfectly signals a change in the probability for all future periods. A slightly more generalized model might distinguish among types of claims: a claim for hospitalization for a broken arm would not trigger a reestimate of future premiums, since it is unrelated to normal hospitalization expenses, while a claim for hospitalization for diabetes would. Likewise, flooding of a house due to the breakage of a water main would not induce a higher premium on water damage to property, while claims from leakage due to normal rain storms would.

2. A three-period model

We first describe a highly simplified model in which a claim in any one period signals a change in probability for two subsequent periods.[1] In section 5, this is generalized to N periods. For simplicity, we will also assume no discounting (zero interest rate).

Initially, all persons have the loss probability p_L. In each of the first two periods, a person who has suffered a loss is believed, by himself and all insurers, to have a loss probability p_H in any subsequent periods. In contrast, a person who has not (yet) suffered a loss is assumed to have the lower probability p_L. Once a person becomes a high risk, his probability is assumed to stay at p_H; further experiences do not cause the insurer or the insured to change their estimation of the probability.[2]

With a three-period lifetime, it is immediately apparent that one possible equilibrium policy is a lifetime policy, in which the premium is paid in the first period and covers the loss in all three periods. In the first period, the loss probability is p_L. Since the proportion of insureds that become higher risk in the second period is p_L, and the proportion that remains low risk is $(1 - p_L)$, the loss probability in the second period is $(p_L p_H + (1 - p_L)p_L)$; in the third period the proportion of low-risk individuals is $(1 - p_L)^2$, and the

loss probability is $p_L p_H + (1 - p_L)p_L p_H + (1 - p_L)^2 p_L$. This last expression distinguishes among three groups: the proportion p_L who suffered a loss in period 1 and are classified early as high risks, the proportion $p_L(1 - p_L)$ who did not suffer a loss in period 1 but do suffer a later loss in period 2, and the remaining proportion who do not suffer a loss in either period and therefore have loss probability p_L in period 3. The lifetime premium P_{LT} which just covers the expected loss is therefore given by:

$$P_{LT} = [(p_L + p_L p_H + (1 - p_L)p_L + p_L p_H + (1 - p_L)p_L p_H \\ + (1 - p_L)^2 p_L)]L \qquad (3)$$

or

$$P_{LT} = [p_L L][3 + 3(p_H - p_L) - p_L(p_H - p_L)].$$

Such a policy would obviously fully protect each person against any period-to-period change in probability, while being financially feasible for a competitive insurer whose expected profits under this system of premiums would be zero.

To illustrate with a numerical example, suppose that $p_L = .1$ and $p_H = .3$. Table 1 depicts the proportion of individuals who are in the high-and low-risk groups at the end of each period, as well as the loss probability. Note that the probability of a loss increases as one moves to later periods, reflecting the larger proportion of high-risk individuals. If $L = 100$, then the lifetime premium $P_{LT} = 35.8$.

2.1. Voluntary single-period insurance

For a number of reasons, persons might prefer to pay the premiums period-by-period, rather than as a single once-in-a-lifetime payment. "Capital market imperfections" is one label which we attach to such reasons. There has been some work on describing these imperfections (Hahn, 1971), but the theory is not well developed. One possible "imperfection," the inability to predict future prices for medical services, will be discussed later in the article. Other "imperfections" have to do with limitations on individual borrowing early in life, when wealth and income are low. A binding contract to pay a lifetime premium in installments would avoid some of these problems. If dropping coverage and failing to pay future premiums by those with good experience in the early

Table 1. Proportion of individuals in high- and low-risk groups and loss probabilities for a three-period problem

Period	Proportion of individuals		Loss probability
	High-risk group	Low-risk group	
1	0	1	0.1
2	0.1	0.9	0.12
3	0.19	0.81	0.138

periods could be ruled out by contract, then a lifetime policy would still be possible. However, it is difficult or costly to enforce such a contract on the insured.

Thus the question for our model is whether there is an arrangement that avoids or reduces such difficulties with precommitment. That is, we seek a breakeven schedule of period-by-period premiums which guarantees the right for those who become high risks to continue to purchase coverage at the same premium as is offered to low risks, and which at the same time is able to make retaining coverage an optimal strategy in every period for the low risk. Is there a "self-enforcing" or incentive-compatible schedule of period-by-period premiums?

The answer is affirmative. We establish this by first showing that purchasing insurance according to a particular sequence of premiums is a breakeven schedule for insurers and would be chosen by all buyers; it is a competitive equilibrium (Proposition 1). Then we show that this schedule is optimal (Proposition 2).

We first introduce the schedule of premiums in Proposition 1:

Proposition 1: The sequence of guaranteed renewability (GR) premiums
$$P_1 = [p_L L][1 + (p_H - p_L) + (1 - p_L)(p_H - p_L)]; P_2 = [p_L L][1 + (p_H - p_L)];$$
$P_3 = p_L L$ is a competitive equilibrium.

Proof: The proof is based on backwards induction. All persons would be willing to pay at least $p_L L$ in period 3. Those who have already suffered a loss (so that their loss probability in period 3 is p_H) would obviously be willing to pay the low premium $p_L L$, but so would those who have yet to suffer a loss and whose loss probability is p_L. Those who did not suffer a loss in period 1 would be willing to pay P_2 in period 2, since the premium for period 2 plus the premium P_3 in period 3 equals their expected loss over both periods. In effect, the period 2 premium has two parts, one part that covers the expected loss for a low-risk person in period 2 (or $p_L L$), and one part that covers the expected loss in period 3 in excess of the premium P_3 (or $p_L L(p_H - p_L)$) for persons who were still low risk in period 2. The premium for period 1 is constructed in the same way, but now there are three parts: the low-risk loss ($p_L L$), the period 2 expected additional loss of those who suffered a loss in period 1 or ($p_L L(p_H - p_L)$), and the period 3 expected loss of those who did not suffer a loss in period 1 but will suffer a loss in period 2 or $p_L L (1 - p_L)$ $(p_H - p_L)$.

At this schedule of premiums, insurers will break even. The sum of the premiums over the three periods equals $[3p_L + 3p_L(p_H - p_L) - p_L{}^2(p_H - p_L)] L$, or $p_L L[3 - 3(p_H - p_L) - p_L(p_H - p_L)]$, which is the same as the lifetime premium P_{LT} (described in equation (3)) that is just sufficient to cover all expected losses. That is, the total amount of premiums collected according to the sequence in proposition 1 is sufficient to allow the insurer to break even. Not only is this sequence of premiums feasible in a competitive market, but it is also optimal, as shown in Proposition 2.

Proposition 2: The competitive equilibrium sequence of GR premiums is Pareto optimal.

Proof: From the viewpoint of period 1, the person expects to obtain complete insurance coverage in all periods and to pay a known series of certain premiums. If the person is

risk averse, if capital markets are perfect, and if insurance administrative costs are zero, the person can achieve the optimum (certain) sequence of consumption. Marginal utility of income will be equalized across all periods for the buyer, while the insurer will just break even; there is no other breakeven sequence of premiums that can make the buyer better off.

When individual probabilities diverge, the possibility of cream-skimming generally arises. Low-risk persons might locate a spot contract that is preferable to remaining in the GR plan. However, the design of our guaranteed renewability policy makes it im-mune to cream-skimming, even though the GR premiums are independent of changes in risk classification. Because GR premiums in any period are fair for low-risk buyers, there is neither cross-subsidization nor any reason for the low-risk buyer to leave the firm. Thus, the loss of a low-risk conumer to the spot market has no impact on the insurer's profits, while the loss of a high-risk consumer is profitable to the GR insurer. Even in the worst-case scenario, where only high-risk consumers remain in the GR policy, the plan continues to break even. Effectively, in the GR plan, the consumers initially prepay enough premiums to cover the excess losses of everyone who becomes a high risk. Since the viability of the policy is independent of the continued presence of low risks, possible reasons why consumers might leave, such as divergent expectations, affect only the at-tractiveness of GR to consumers, not the economic viability of the policy.

   Any *lower* level of payment than what we envision would leave the policy open to rational defection by good risks. This scheme also envisions an arrangement in which all insureds pay the same net premium in any period. For a fascinating model that derives an alternative (though similar) method of achieving multi-period consistency through the use of refunds see Cochrane and Meltzer (1992).

3. The time path of premiums

It is obvious that this sequence of premiums displays a declining path. The premium in the first period has to be high enough to permit persons who become high risks in that period to continue to pay the same sequence of premiums in future periods as those who have yet to suffer a loss. Not all of the premium in excess of p_L is paid in the first period, however, since the persons who experienced a loss in that period also pay more than p_L L in the second period. But there is some additional markup.

   This proves that a "level-premium" guaranteed renewability policy, in which one pays a fixed supplement in each period (but the last) to guarantee the right to continue buying coverage at that premium, is not feasible if the time period is finite. Such a policy would cause low risks to drop out in the later periods, since their expected premiums would exceed their expected losses. Thus, a level premium policy would become vulnerable to cream-skimming or defection.

   An implication of our model is that, the longer the time period, the larger the amount by which the guaranteed renewability premium in the first period will exceed the one-period fair premium $p_L L$. Indeed, if p_L is large, the proportion of the population which becomes high risk grows large by the last period.

4. Generalization to N periods

This simple model easily generalizes to N periods. The key idea is that the premium in excess of $p_L L$ in the first period has to be enough to buy the right in all subsequent periods to pay the same sequence of premiums as those who have yet to suffer a loss.

This means that the premium schedule takes on a "back to front" property in which the premium collected from all persons in the first period is just enough to cover total losses in the Nth period, the premium collected in the second period just covers losses in the $(N - 1)$st period, and so on. Of course, we already know that premiums collected in the Nth period, $p_L L$, are just enough to cover losses in the first period.

Let $P_N(k)$ denote the premium in period k for an $N \geq k$ period problem. In the three-period example, the third-period loss is $p_L L (p_H + (1 - p_L)p_H + (1 - p_L)^2)$. The first period premium $P_3(1) = p_L L [1 + (p_H - p_L) + (1 - p_L)(p_H - p_L)]$, or $p_L L [p_H + (1 - p_L)p_H + (1 - p_L)^2]$, which is the same as the third-period loss. The second-period premium $P_3(2) = p_L L [p_H + (1 - p_L)]$, which is the second-period loss; and the third-period premium is $P_3(3) = p_L L$, which is the first-period loss. For our three-period numerical example, the respective premiums are $P_3(1) = 13.8$, $P_3(2) = 12$, and $P_3(3) = 10$, which, of course, yields a lifetime premium of $P_{LT} = 35.8$.

In the general case, the last (Nth) period loss from a policy that runs from $t = k$ to $t = N$ is $L_k(N) = p_L L [p_H + (1 - p_L)p_H + (1 - p_L)^2 p_H + \cdots + p_L(1 - p_L)^{N-k-2}p_H + p_L(1 - p_L)^{N-k-1}]$. The premium for period k is

$$P_N(k) = p_L L (1 + (p_H - p_L) + (1 - p_L)(p_H - p_L) + (1 - p_L)^2 (p_H - p_L) + \cdots + (1 - p_L)^{N-k-2} (p_H - p_L).$$

Notice that this expression involves the same sequence of terms p_H, $(1 - p_L)p_H$, $(1 - p_L)^2 p_H$, $\ldots$, $(1 - p_L)^{N-k-2}p_H$ as does $L_k(N)$. It can be shown that the remaining terms equal $p_L(1 - p_L)^{N-k-1}$. If $P_N(k) = L_k(N)$, it follows that $P_N(k + 1) = L_k(N - 1)$, and, in general, $P_N(k + j) = L_k(N - j)$, as long as $k + j \neq N - j$.[3]

5. Implications of the model

This model implies that premiums for feasible efficient guaranteed renewability coverage in a full information, non-myopic world will be represented by a sequence of continually decreasing premiums. Premiums begin above $p_L L$ and gradually decline to that level. The initial period premium will be higher (relative to $p_L L$) the larger p_H is relative to p_L, and the longer the total time period over which coverage is to be purchased. (This is the opposite pattern from that for multi-period coverage that emerged in the Kunreuther–Pauly (1983) model with asymmetric information. In this model, it is necessary to charge a higher premium initially to make sure that the insurer breaks even if low-risk persons leave for the spot market. As one moves over time, the insurer is able to discriminate more low-risk buyers, and hence to lower premiums to reflect this. In the Kunreuther–Pauly model, the insurer attracts buyers by charging a low premium and

capitalizes on the increasing amount of inside information which they gather over time by continually raising premiums.)

However, since the expected loss of even healthy individuals increases as they age, this pattern of declining premiums may not be displayed in actual health insurance markets. That is, p_L "naturally" increases with age. If the rate of increase is rapid enough, then a level premium version of guaranteed renewability might by chance become feasible. While precisely level premiums are unlikely to be optimal, the equilibrium schedule for health insurance with an age trend would approximate a more level premium than the one implied by the above model.

If individuals are highly uncertain as to how long they will want coverage, which leads them to be myopic in their behavior, then they will prefer single-period insurance rather than a lifetime contract. In general, the shorter the time horizon, the more unattractive is GR coverage, particularly if the administrative costs are relatively high in relation to the expected losses. Persons who are uncertain as to how long they will be living in a particular home or community may prefer to purchase single-period homeowners insurance. They may also be reluctant to purchase GR coverage on health insurance if the coverage was restricted to their present location.

6. Different initial levels of risk

There are two possible circumstances that describe the transition from a low-risk to a high-risk state. They differ based on the assumption about the underlying prior probability distribution of illness.

The simple model already described imagines that all persons are initially at the same level of risk, but that exogenous events cause their risk level to change. For instance, a person who suffers an accident that causes disability may be at higher risk for the rest of his or her life. However, there is no underlying variation in the prior probability of an accident.

A second model might assume that people are genetically endowed with different probabilities of becoming ill.[4] A medical event may signal the kind of risk a person is, and, thereafter, the person will be regarded as higher risk. One of the issues which we will explore is whether this model leads to different conclusions than the other one. The main difference is that, in such a case, the low probability for those who have yet to suffer a loss will decline over time (effects of age aside). The reason is that the proportion of low risks among those without a loss increases over time.

Formally, let α be the known proportion of high risks in the first period, say, because they are genetically predisposed to some illness. The loss probability is then:

$$p_N^1 = \alpha p_c + (1 - \alpha) p_u,$$

where p_c is the loss probability for a carrier of the higher risk, and p_u is the loss probability for those unaffected. In the next period, the probability for those yet to suffer a loss becomes:

$$p_N^2 = \alpha(1 - p_c)p_c + (1 - \alpha)(1 - p_u)p_u.$$

By the tth period, the loss probability is

$$p_N^t = \alpha(1 - p_c)^t p_c + (1 - \alpha)(1 - p_u)^t p_u.$$

Since p_c is greater than p_u, it is obvious that the weight for high risks is declining, and that the loss probability for persons yet to suffer a loss is declining.

The analysis of the optimal multi-period sequence of premiums in this case is the same as in the previous case, except that p_L is replaced by p_N^t and the no-loss probability varies over time. The primary consequence of assuming that there is a predisposition to the illness which varies across individuals is to steepen the gradient of optimal premiums over time, since (given the same initial level of p_L and p_N) p_N falls over time.

7. Limitations of guaranteed renewability coverage

There are several reasons why GR coverage may be less attractive to individuals than the above models imply. We examine these below.

7.1. Divergent expectations

The assumption that all insurers have the same estimate of L, the level of loss conditional on an adverse event, may not always be realistic. In particular, insurers may well differ on their forecasts of future medical care costs to be incurred once illness strikes, largely based on their guesses about future health care prices and technology. This is especially likely to be the case when there is generalized uncertainty about the rate of growth in losses, as was the case in the late 1980s for medical care costs. Variation across insurers in their forecasts of future costs can cause guaranteed renewability (GR) to fail to be offered in a competitive market. Rather than lock himself into a long-term relationship with a particular insurer, a buyer may prefer to preserve the option of switching to the firm that offers a lower price.

Formally, let there be n insurance firms. Let insurer i's expected value of L (for any level of probability) be represented by $L_i = \bar{L} + \varepsilon_i$, where $\sigma_{\varepsilon i}^2 > 0, \bar{\varepsilon}_i = 0$, and $i = 1$, ..., n. If the insurance buyer has perfect information about every insurer's premium in any period, he can choose to purchase from the firm offering the lowest price (of n prices).

For the present, assume that a risk-neutral insurer i sets its one-period premium for coverage with guaranteed renewability to high or low risks by the rule

$$P_i = g\left(p_H, p_L, (\bar{L} + \varepsilon_i)\right).^5$$

All firms are assumed to agree upon the loss probabilities p_H and p_L as being the same for all firms; however, the ε_i differ across firms. Under this formulation, the firm with the lowest value of ε_i will set the lowest price and attract all of the business in the market.

More generally, one could model the process as a search model, with probability q (P_i) representing the probability that a customer of firm i will discover another firm charging a premium below P_i.

The individual contemplating the continued purchase of GR coverage from any firm i then faces four possible alternatives in any future period: (1) he may be a low risk and find that P_i is the lowest price; (2) he may be a low risk and find that P_i is not the lowest price; (3) he may be a high risk and find that P_i is the lowest price; (4) he may be a high risk and find that P_i is not the lowest price.

It is easy to see that the value of GR coverage is reduced if a high-risk person might be able to pay lower premiums by switching to an insurer with a low value of ε_i, rather than by continuing to pay the sequence of GR premiums to his current insurer. If expectations are the same for all insurers, this cannot happen, since the sequence of spot-market, high-risk premiums charged by firm j to a high risk for N periods, (or NP_{Hj}), will surely be greater than the sequence of GR premiums charged by a particular firm i. However, if the value of E differs for different firms, then the relevant comparison is between $NP_{Hb(t)}$ and the sequence of GR premiums for firm i, where $P_{Hb(t)}$ is the best (lowest) high-risk premium in each period t.

For instance, in the three-period model, there may be a chance that the third (last-period) GR premium P_{3i} ($= p_L L$) is greater than some P_{Hj} in that period, or that P_{2i} ($= P_{3i} + p_L L(p_H - P_L)$) is greater than some P_{Hj} in the second period. Alternatively, there may be no lower price in the last period, but the price in the second period may be sufficiently lower that the sum of the lowest premiums is less than the sequence of GR premiums. This possibility reduces the value of GR coverage to an insured individual.

In general, the chance that there is a better sequence of non-GR ("spot") premiums is greater the longer the period of coverage is, the larger the variance of spot premiums in periods close to the time of initial purchase is, and the lower the transaction costs associated with different premiums offered by insurers are.

7.2. Specifying the level of coverage

Another reason why guaranteed renewability may not be attractive is related to the fact that the insurer must generally fix (or at least limit) the level of coverage in the future. Were this not the case, the high-risk person facing a premium per-dollar of coverage of p_L will buy insurance in excess of $\$L$, since the premium will be actuarially favorable to him.

In the simple model developed thus far, optimal coverage at any time period equals $\$L$. Hence, an ideal guaranteed renewability policy will need to fix in advance the level of coverage to be provided at any premium. With a positive adminstrative loading, the optimal level of coverage will decrease. However, as long as *some* level of insurance coverage will be demanded in any foreseeable set of circumstances, there will be a market for renewability. The buyer will purchase such protection for some low level of coverage, and will purchase supplementable (and nonrenewable) coverage at the margin.

7.3. Risk-averse insurers

A third possible reason for the lack of interest in such coverage is that insurers may not be risk neutral with respect to uncertainty about the value of L in each time period. In contrast to the risk of loss, medical cost inflation is not a poolable risk. If insurers are generally risk averse, they will require a risk premium to be added to the fair premium for their coverage. In response to this extra cost, insurance buyers may choose to retain some of the risk, and therefore may not buy complete lifetime coverage. Cutler (1992) offers a demonstration of this proposition for nursing home insurance. However, this only shows that GR coverage will not be complete; it will still generally be optimal for there to be *some* GR coverage, even if there is systematic risk, and insurers are risk averse. The optimal arrangement is for every agent in society to accept some part of the total portfolio of risks. For example, although no one knows precisely how much will be needed to cover high-risk costs five years from now, delivering some finite amount of benefit in such cases will surely be optimal.

7.4. Lock-in problems

Another deficiency of guaranteed renewability is the opportunity it gives to the insurer to exploit the fact that the buyer is locked into dealing with a particular insurance firm. Having prepaid the right to buy future coverage at lower rates than other insurers would offer, the high-risk buyer would be at the mercy of the insurer if it raised premiums closer to the market levels for that buyer.

The defense against this possibility is, in part, a careful drawing of the initial contract, and, in part, the method for linking the premium under guaranteed renewability to the premium for other insurance which the firm sells. Thus, the future premium will typically be linked to the premium which the firm charges for new standard business. This control is not foolproof; if owners of the insurer would tolerate a disappearance of its new business (or perhaps lodge it in a legally separate firm), it might still pay, at the margin, to exploit renewability coverage to some extent. In much the same form as the rules governing Medicare HMO business, a volume requirement for "regular" business (or new business even with renewability) may be needed to prevent exploitation.

For the linkage strategy to work optimally, the amount of guaranteed renewability coverage must be small relative to the amount sold in the "average risk" market. If the individual insurer's market share is relatively high, at the margin, it will pay to raise the price for standard coverage, lose some sales in that market, but be entitled to charge more to high risks. Rather than link the premium under guaranteed renewability coverage to the firm's own premium, some buyers may prefer a link to some market indicator of costs.

8. Guaranteed renewability and employee mobility

In current insurance markets, the type of underwriting and premium adjustment that leads to the substantial jumps in premiums is usually limited to the individual and small

employment group markets. Larger employment-based groups, which often provide health insurance, are usually not medically underwritten. Though they are almost always experienced rated or self-insured, the large size of the group means that the group's average premium will usually not fluctuate because of the health experience of particular employees or dependents. Since persons who are self-employed or employed in small groups also have or have had employment opportunities with large firms, what does this do to the value of GR health insurance coverage?

To answer this question, we need to model the process by which employees move among employers. In what follows, we assume that switching is endogenous, depending in part on relative wage/fringe packages at different firms. We assume that employees of small firms can choose whether to purchase GR coverage, and, in doing so, take account of the prior availability of implicit GR coverage in the large-firm sector. (That is, if large firms do not or cannot charge higher premiums or pay lower money wages to persons who become high risks, they, in effect, provide GR coverage.) Large firms are assumed to display no behavioral response to a high-risk job seeker, either in the money wages which they offer new workers or their willingness to accept such a worker.

It follows that a worker who would have preferred a small-firm job (because of locational convenience, say) can choose voluntarily to switch to a large-firm job if he becomes high risk, if the value of GR coverage in the large firm exceeds his (dis)utility cost of switching. If the cost of switching is less than the cost of explicit GR coverage for the small group, the small firm will not choose to purchase such coverage. Even if a job in the large-firm sector is not certain, the utility loss from taking a chance on losing coverage may be less than the cost of GR coverage for some small-firm workers.

This simple idea implies that the proportion of small groups buying (explicit or implicit) GR coverage will decline if the market price of (explicit or implicit) GR coverage rises relative to the cost of switching jobs. The substantial increase in medical costs in the 1980s, which raised the cost of GR coverage may have been such a disproportionate increase. This represents the first step in an adverse selection or "lemons" problem. Large firms might eventually be expected to take the health risk status of workers and dependents into account in hiring. While some large employers have allegedly done so, many others have adopted the political strategy of supporting a mandate of GR or community-rated coverage on their small firm competitors.[6]

9. Optimal guaranteed renewability

If there is no administrative cost to adding guaranteed renewability coverage to a one-period premium, and if the future sequence of loss probabilities is entirely outside the insured's control, it will be optimal to buy fully guaranteed renewability coverage, that is, complete protection against changes in risk. Either positive administration loading or moral hazard can make such full coverage undesirable.

The moral hazard argument is easiest to see. If the insured could engage in costly activities to reduce the future probability of becoming a high risk (e.g., health habits, waste disposal practices), full coverage would reduce the level of those activities. To

reduce such moral hazard, sharing of some portion of high-risk premium costs (as a form of coinsurance) will generally be optimal. That is, the insured would pay some proportion of any increase or decrease in premiums that might be related to changes in risk.

If administrative cost is not zero, the optimal form of coverage depends on the way administrative cost varies with the "extent" of protection against risk changes. If there were a positive *marginal* loading cost to greater and greater smoothing of unexpected premium changes, then the optimal solution would involve a kind of deductible. The person would not want protection against small changes in loss probability, but would still seek to be protected against large changes. However, one special feature of fully guaranteed renewability coverage is that it completely avoids the need for determining what an individual's loss probability actually is, after insurance is sold. In other words, GR coverage avoids the need for (and cost of) underwriting. While providing and running the complex pattern of premiums described above will itself be costly, these observations do suggest (absent moral hazard) an all-or-nothing character of guaranteed renewability: either one will choose to be completely protected or not protected at all. The choice will depend on the level of administrative costs, on the extend of prospective variation on premiums, and on the individual's risk averson.

10. Conclusion

Guaranteed renewability appears to be an attractive feature for a number of different types of insurance. Its most serious enemy is instability in estimates of underlying (and industry-wide) loss trends. The apparent failure of such coverage to exist in most of small-group health insurance and environmental liability insurance in the late 1980s may reflect such instability; the instability may itself have contributed to the breakdown of a kind of implicitly guaranteed renewability that had existed between purchasers of small-group insurers and their usual insurers over many years. Instability is not easily treated or cured, though better forecasting would appear to be an attractive option in the small-group and individual markets.

Acknowledgements

Notes

1. We analyze three periods, rather than two, because the interesting properties of multi-period contracts often are displayed in periods other than the initial or terminal periods.

2. A more gradual learning process would display similar properties to this limiting case.
3. John Villiani provided helpful comments on an earlier version of this article that enabled us to obtain this result.
4. We are indebted to Edi Karni for this suggestion.
5. If insurers consider the possibility of "winner's curse" behavior, they will follow a more complicated rule. All we require at this point is that some insurers charge less than others for the same coverage.
6. The Americans with Disabilities Act puts substantial limitation on an employers' ability to refuse to hire a person who happens to be a high risk, or to charge higher applicant premiums. However, the applicability of the law to variation in money wages seems limited.

References

Cochrane, J., and D. Meltzer. (1992). "Time-consistent Health Insurance." Working paper, Department of Economics, University of Chicago.

Cutler, D. (1992). "Why Doesn't the Market Fully Insure Long Term Care Insurance." Working paper, Harvard University and NBER.

Hahn, F. (1971). "Equilibrium With Transactions Costs," *Econometrica* 39, 417–439.

Kunreuther, H. and M. Pauly. (1985). "Market Equilibrium With Private Knowledge: An Insurance Example," *Journal of Public Economics* 26, 269–288.