National Association of Insurance Commissioners

Journal of Insurance Regulation

Cassandra Cole and Kathleen McCullough
Co-Editors

Vol. 39, No. 6

Guaranteed Renewability in Health Insurance: Taking into Account Changes in Risk Status and the Cost of Dying

Annette Hofmann
Patrick Eugster

JIR-ZA-39-06

National Association *of* Insurance Commissioners

The NAIC is the authoritative source for insurance industry information. Our expert solutions support the efforts of regulators, insurers and researchers by providing detailed and comprehensive insurance information. The NAIC offers a wide range of publications in the following categories:

Accounting & Reporting
Information about statutory accounting principles and the procedures necessary for filing financial annual statements and conducting risk-based capital calculations.

Consumer Information
Important answers to common questions about auto, home, health and life insurance — as well as buyer's guides on annuities, long-term care insurance and Medicare supplement plans.

Financial Regulation
Useful handbooks, compliance guides and reports on financial analysis, company licensing, state audit requirements and receiverships.

Legal
Comprehensive collection of NAIC model laws, regulations and guidelines; state laws on insurance topics; and other regulatory guidance on antifraud and consumer privacy.

Market Regulation
Regulatory and industry guidance on market-related issues, including antifraud, product fi ling requirements, producer licensing and market analysis.

NAIC Activities
NAIC member directories, in-depth reporting of state regulatory activities and official historical records of NAIC national meetings and other activities.

Special Studies
Studies, reports, handbooks and regulatory research conducted by NAIC members on a variety of insurance related topics.

Statistical Reports
Valuable and in-demand insurance industry-wide statistical data for various lines of business, including auto, home, health and life insurance.

Supplementary Products
Guidance manuals, handbooks, surveys and research on a wide variety of issues.

Capital Markets & Investment Analysis
Information regarding portfolio values and procedures for complying with NAIC reporting requirements.

White Papers
Relevant studies, guidance and NAIC policy positions on a variety of insurance topics.

For more information about NAIC publications, visit us at:

http://www.naic.org//prod_serv_home.htm

Printed in the United States of America

NAIC Executive Office
444 North Capitol Street, NW
Suite 700
Washington, DC 20001
202.471.3990

NAIC Central Office
1100 Walnut Street
Suite 1500
Kansas City, MO 64106
816.842.3600

NAIC Capital Markets
& Investment Analysis Office
One New York Plaza, Suite 4210
New York, NY 10004
212.398.9000

Purpose

The *Journal of Insurance Regulation* is sponsored by the National Association of Insurance Commissioners. The objectives of the NAIC in sponsoring the *Journal of Insurance Regulation* are:

1. To provide a forum for opinion and discussion on major insurance regulatory issues;

2. To provide wide distribution of rigorous, high-quality research regarding insurance regulatory issues;

3. To make state insurance departments more aware of insurance regulatory research efforts;

4. To increase the rigor, quality and quantity of the research efforts on insurance regulatory issues; and

5. To be an important force for the overall improvement of insurance regulation.

To meet these objectives, the NAIC will provide an open forum for the discussion of a broad spectrum of ideas. However, the ideas expressed in the *Journal* are not endorsed by the NAIC, the *Journal's* editorial staff, or the *Journal's* board.

Guaranteed Renewability in Health Insurance: Taking into Account Changes in Risk Status and the Cost of Dying

Annette Hofmann*
Patrick Eugster**

Abstract

Guaranteed renewability protects policyholders from reclassification risk. Being an important characteristic of social health insurance, the potential for private insurance markets is high given its property of competing with risk selection. Without regulation—because health care expenditures increase strongly near death—it seems questionable whether insurers will be able to sustain guaranteed renewability in the long run, rather than investing in risk-selection activity. Extending the seminal model of Pauly et al. (1995) to include policyholders with improving risk status and the high cost of dying, we show that the actuarially fair guaranteed renewable premium in realistic conditions becomes lower, suggesting that prior studies have overestimated the economic cost of guaranteed renewability, making it more affordable and accessible in practice. Our findings illustrate the potential to overcome the common market failure associated with risk selection by introducing guaranteed renewability into an existing risk-based system.

JEL Classification: G22; I13.
Keywords: health insurance, guaranteed renewability, long-term contracts.

* MetLife Associate Professor, School of Risk Management, Insurance, and Actuarial Science, Peter J. Tobin College of Business, St. John's University, 101 Astor Place, New York, NY; *hofmanna@stjohns.edu.*

** CEO, Dextra Insurance Ltd., Zurich, Switzerland; *Patrick.Eugster@dextra.ch*.

Introduction

Guaranteed renewability of health insurance contracts is an interesting and important feature. It offers stability of premiums in the face of unexpected deterioration of health status, which is a significant risk faced by policyholders. Although guaranteed renewability is traditionally a characteristic of social health insurance, it has recently been written voluntarily into policies by private insurers. In particular, private health insurers in the U.S. have been writing guaranteed renewability into their individual policies in response to pressure from the regulator. [1] In Europe, private health insurers often provide the guaranteed renewability feature which improves their ability to compete with social health insurance (where contributions vary with labor income but not with health status).[2] Evidently, guaranteed renewability amounts to a commitment of the insurer to abstain from future risk selection, which is generally done by replacing (typically older) clients who have become unfavorable risks by (typically younger) clients who are favorable risks.

Several studies (e.g., Handel et al., 2015), have explained and demonstrated for the U.S. that reclassification risk is probably the most severe market failure of health insurance markets today. Guaranteed renewability protects consumers against reclassification risk and may, therefore, prevent risk selection (Pauly, 2012; Pauly et al., 2011). In practice, rate regulation implies that insurance premium rates must be adequate so that insurers remain solvent, which works as a lower limit on prices. When premiums are risk-based and do not protect against reclassification risk, younger and healthier individuals may not purchase coverage, while older and sicker individuals have no choice and are left with excessively high premiums as they grow older. This article suggests that prior studies have overestimated the economic cost of guaranteed renewability, making it more affordable and accessible in practice. As a result, younger individuals may have a higher incentive to "lock in" their health insurance premiums, and a market failure may more easily be overcome by introducing guaranteed renewability as a contract feature into an existing risk-based system.

Because welfare cost may exceed that of risk selection under asymmetric information on health status, recent empirical research has identified an important challenge to guaranteed renewability insurance: Health care expenditures have been found to increase sharply with closeness to death (Zweifel et al., 1999), leading

1. See *https://www.cms.gov/CCIIO/Resources/Files/Downloads/market-rules-technical-summary-2-27-2013.pdf.*

2. The premiums of guaranteed renewability policies decrease if policyholders change from high-risk to low-risk status. For instance, private health insurance in Germany comes with a system where each policyholder pays an individual premium, mainly based on their entry age. Insurers apply surcharges for preexisting medical conditions, such as asthma or diabetes, at the time when the insurance contract is signed; then, when policyholders can demonstrate over time that they no longer suffer from the condition, premiums are adjusted downward accordingly. See *https://www.howtogermany.com/pages/fastfacts2.html*.

Steinmann et al. (2007) to distinguish between expenditure for restoring health and the cost of dying.[3] Moreover, Felder et al. (2007) found that the increase in health care expenditures started five years before death. These findings imply both a strengthened incentive and enhanced capability of health insurers to invest in risk selection, potentially modifying the conditions stated by Pauly et al. (1995) and Cochrane (1995) for the sustainability of guaranteed renewability. Pauly et al. implicitly assume that there is only one possible transition—which moreover is permanent, from low-risk status to high-risk status—neglecting two features that are added here. First, high-risk individuals may become low-risk types again, reflecting pertinent findings by Beck et al. (2010); second, death is introduced as the "absorbing state" (in Markovian language). While the first modification facilitates guaranteed renewability in health insurance, the second may well obviate it, being the ultimate deterioration of health status associated with a lot of extra health care expenditures, especially of high risks.[4] This raises the question regarding the combined effect of returning to low-risk status and death. In other words, can guaranteed renewability survive in the presence of death? The main finding is that the combination of the two modifications may enhance the viability of guaranteed renewability because it causes the upfront guaranteed renewability surcharge to be lower (given realistic parameter values). This finding is of importance to private health insurers outside the U.S., most of whom cannot transfer their enrollees to a special scheme for retirees, such as Medicare. Death and the cost of dying are of high relevance to U.S. private health insurance, as well, given that mortality of age group 15–64 is about 334 out of 100,000 people (0.334%) per year, as compared to 2,011 out of 100,000 people (2.01%) per year for the age group 65–74 .[5]

The remainder of this paper is structured as follows. Section 2 includes a review of the pertinent literature. In Section 3, the Markov process specified by Pauly et al. (1995) is complemented by a positive transition probability from high- to low-risk status, as well as a separate cost of dying. The sections also discuss limitations and extensions of our model modification. Concluding remarks follow in the last section.

3. Following this literature, the "cost of dying" refers to the mortality component of health care expenditures, which needs to be distinguished from the morbidity component. In has been shown that failure to make this distinction results in excessive estimates of future growth of health care expenditures. See Steinmann et al. (2007).

4. In Steinmann et al. (2007, Table 1), health care expenditures increase with age, especially among women, *ceteris paribus*, as long as proximity to death is not controlled for (suggesting that women become relatively high-risk with age). The same effect is present among those dying within two years, which points to an extra cost of dying for female high risks.

5. See U.S. Centers for Disease Control and Prevention (CDC), National Vital Statistics, 2007, Worktable 23.R.

Related Literature

In long-term health insurance, there are two basic solutions for dealing with risk selection, possibly causing high-risk individuals to remain uninsured. One solution is to impose open enrollment combined with community rating by law, as is the case in employer-contracted health insurance in the U.S. and individually contracted health insurance in the Netherlands and Switzerland.

The other solution is to make contracts incentive-compatible. The latter, market-based solution has been investigated by Pauly et al. and Cochrane (1995).

Pauly et al. develop long-term insurance contracts that exhibit time consistency. In Pauly et al., a sequence of incentive-compatible short-term contracts results in guaranteed renewability. In other words, the guaranteed renewability insurance market model is based on the following intuition: A sequence of premiums is offered such that insurers can break even and policies are chosen by both low- and high-risk buyers, regardless of whether they have suffered a loss.[6] In order to be sustainable, the premium schedule continually declines over time to determine who the low-risk buyers are. The highest premiums are thus charged at the beginning in order to protect the insurer from the effect of low-risk individuals leaving for the spot market. As a result, by applying backward induction (as in game theory), one can identify a contract sequence with declining premiums that ensures that insurers can break even in equilibrium. As pointed out by Cochrane (1995, p. 447), these contracts are "renegotiation-proof" and satisfy participation constraints in that both parties are willing to sign the next-period contract in all future circumstances.

Before the federal Affordable Care Act (ACA) was enacted in the U.S., market-based individual private health insurance contracts represented the common system. In contrast, many European countries and Chile have a public community rating-based health insurance system In Germany, for instance, a risk-based private, as well as a community-based public, health insurance system coexist side by side. As stated by Herring and Pauly (2003) for the U.S., "guaranteed renewable health insurance policies do exist in (and dominate) the market, even in the absence of regulation, and appear to be stable" (p. 4). [7] Herring and Pauly (2003) provide direct comparisons of the extent of front-loading of actual health insurance premiums paid in the U.S. with an estimate of the optimal age-path of premiums, using data from

6. Remember that guaranteed renewability insurance contracts are defined here as a sequence of short-term contracts that are incentive-compatible, satisfy participation constraints, and are renegotiation-proof.

7. Note that this description of private health insurance in the U.S. always refers to the situation before the ACA. Under the ACA, private health insurance companies (at least on the exchanges) cannot base their premiums on individual health status and medical history. It became law March 23, 2010, and represents the most significant transformation of the U.S. health care system since Medicare and Medicaid, which were introduced in 1965. See Manchikanti et al. (2011). However, depending on the outcome of legislation under the current President Donald J. Trump, the U.S. individual insurance market might revert to something closer to the pre-ACA state.

the Medical Expenditure Panel Survey (MEPS).[8] They conclude that, "despite the moderately low cost of the guaranteed renewability feature, younger individuals who place a high value on current levels of consumption may still not be willing to pay this low cost unless the breakeven guaranteed renewability premium is subsidized. Regardless, it does seem that existing breakeven schedules come reasonably close to optimal incentive-compatible patterns" (p. 27).

Guaranteed renewability contracts achieve time-consistency by frontloading premiums. As shown by Frick (1998), frontloading may be excessive for individuals who have limited capital endowment in the first periods of the guaranteed renewability contract sequence, which may lead individuals to not buy guaranteed renewability insurance. Yet, the empirical evidence presented by Hendel and Lizzeri (2003), as well as Herring and Pauly (2006), suggests that this may not be a problem in actual practice. This paper addresses this discrepancy by showing that the empirical evidence seems in line with a setting where more realistic model parameters are taken into account.

Introducing guaranteed renewability may be a way to ensure long-term health insurance coverage without risk-selection issues for an entire population, assuming individuals are capable and willing to prepay the frontloading.[9] Patel and Pauly (2002, p. 283) describe the policy choice in the following way, "The alternative to guaranteed renewability, for people concerned about adverse selection, risk rating, or cream skimming, is, as Paul Ginsburg notes, 'setting strong (regulatory) rules for this market.' Those rules usually entail some kind of community rating or limits on risk rating. Such rules themselves cause adverse selection and cream skimming, so they tend to beget still more restrictions on the kinds of policies that can be offered." As a consequence, guaranteed renewability can alleviate market failures due to selection issues, and at the same time may require *less regulation* than in a purely risk-rated system.

Regarding the existence of a trade-off between competition and adverse selection, this article argues that guaranteed renewability may be seen as an efficient alternative to the more severe premium regulation in health insurance necessary under community rating; this is because community rating creates incentives for adverse selection that need to be addressed by insurance pricing and risk selection. Cutler and Reber (1998) compare the benefits of insurance competition with the costs of adverse selection, using health plan choice data from Harvard employees.

8. The MEPS is a survey of families and individuals, their medical providers and employers across the U.S. It is the most complete source of data on the cost and use of health care and health insurance coverage in the U.S.

9. Note that Cochrane's (1995) solution is different, being derived from a multi-period utility function in discrete time. A separate account needs to be created for so-called bidirectional severance payments, which are equal to the excess of the present value (PV) of premiums over the PV of future expected health care expenditures or the excess of the PV of future expected health care expenditures over the PV of future premiums, respectively. This account is designed to avoid ex-post defection by one party (who can also be an individual who gets healthier unexpectedly). This seems to work in many cases, given that risk-averse consumers buy health insurance voluntarily, with premium development ensuring time consistency.

They show that while a voucher-type system creates significant welfare losses due to adverse selection, increased competition reduces premiums significantly; then adverse selection can be minimized by adjusting voucher amounts for individual risk types.

Although it is not clear whether guaranteed renewability health insurance will always be more efficient than the current system, it may require less regulation. The improved efficiency does not necessarily lie in the reduced effort required for risk selection activities by insurance companies alone, but rather in the overall amount of regulation that is necessary in order to prevent market failures.

However, guaranteed renewability is unlikely to work perfectly either,[10] and some adverse side effects of imperfect risk-rating remain, in particular a lock-in of high risks who opt for a guaranteed renewability contract.[11] Still, guaranteed renewability is a market-based mechanism that at least partially overcomes the problem of uncertainty surrounding long-term health status. Assuming that the number of high-loss periods is fixed and the same for everyone, Pauly et al. (1998) derive a level premium schedule for group (e.g., employer-contracted) health insurance. In fact, a double pooling mechanism can be created both at the group level and the insurer level, through having one insurer enroll several groups. Such a double pooling serves to reduce frontloading for guaranteed renewability.

Only three empirical contributions deal with guaranteed renewability contracts. Brown and Connelly (2005) evaluate the Australian government's initiative to foster long-term private health insurance using a guaranteed renewability model with the probabilities of Herring and Pauly (2006). Although they find Australia's lifetime cover to be subject to adverse selection, guaranteed renewability may constitute a voluntary alternative because it avoids loading hikes at higher ages in return for an upfront surcharge. In a second contribution, Shelton Brown and Connelly (2005) extend the Pauly et al. model to 35 periods, allowing for age-dependent loss probabilities. They hypothesize that the existence of large cross-subsidies from healthy, younger individuals to less healthy, older ones is one of the key factors preventing the young from voluntarily buying health insurance in Australia. Substituting this cross-subsidization by risk-rated guaranteed renewability that protects consumers against future deterioration of their health status is a way to overcome this market failure. Finally, Herring and Pauly (2006) estimate the amount of frontloading in existing guaranteed renewability health insurance in the U.S. They construct a risk-rated guaranteed renewability premium profile and compare it with the observed development of premiums during the life of the contract. Indeed, they find a remarkable degree of similarity between predicted and actual time paths.

10. Interestingly, Herring and Pauly (2001) show that, in fact, premiums vary more strongly with risk in community-rated areas than in risk-rated areas of the U.S. This is of particular interest as premiums play a major role in determining health insurance choice (Cameron and Trivedi, 1991).

11. See Patel and Pauly (2002) as well as empirical evidence provided by Hofmann and Browne (2013).

The Model

The starting point of this article is the optimal contract as in Pauly et al., who deduce a time-consistent premium schedule from expected cost development over time. The more the planning period is extended into the future, the larger becomes the initial prepayment to ensure guaranteed renewability. At least in the U.S., the longest observable time horizon is defined by eligibility for Medicare; i.e., up to age 65. In countries without a scheme similar to Medicare, the planning period is limited by the expected time of death. In these countries, transition between health states must include death for deriving a time-consistent optimal health insurance contract.[12]

In the Pauly et al. model, all individuals have the same initial low loss probability in the first period. Individuals who have suffered a loss are then assumed (by themselves and by all insurers) to have a high loss probability in all subsequent periods, whereas those who have not suffered a loss are assumed to remain low risks. This model implies that the premiums for feasible and efficient guaranteed renewability coverage in a full-information world are represented by a sequence of continually declining premiums.[13] Premiums start out (well) above expected health care expenditures of a low risk and gradually decrease to that level (in the final period). The initially high premiums protect the insurer against the event that low risks leave for the spot market.

Assume there is a probability p_{12} of an insured turning from a low risk (with expected future health care expenditures of $p_l L$, where p_l is the common loss probability and L the size of the loss) into a high risk (characterized by $p_h > p_l$ and $H \geq L$, respectively). Accordingly, let p_{11} be the probability that the individual retains low-risk status. Now let high risks return to low-risk status with probability $p_{21} > 0$. Accordingly, they also have a probability p_{22} of remaining high risks. In addition, they may die with probability $p_{24} > 0$, which is also true of low risks with probability $p_{13} > 0$.[14] These probabilities give rise to a Markov process as illustrated in Figure 1.[15]

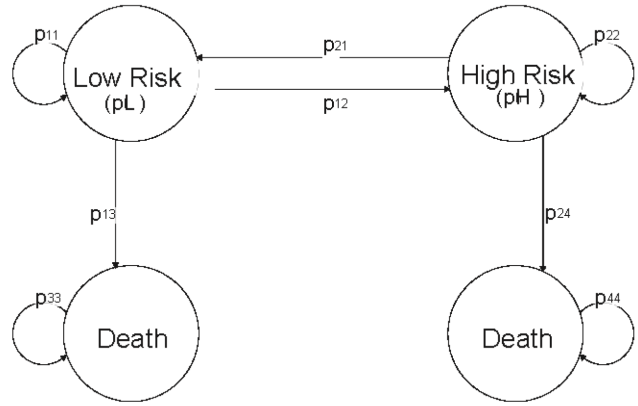
12. Death should be taken into account even for the case of U.S. Medicare; for an attempt outside the U.S., see Zweifel et al. (1999).

13. We follow the Pauly et al. approach of a full-information model here. For an analysis of private information on insurance market equilibrium, see Doherty and Thistle (1996).

14. Herring and Pauly (2006) present empirical evidence suggesting that the probability of death differs significantly between high and low risks. As confirmed by Felder et al. (2010), remaining life expectancy reflecting the possibility of death is important in determining health care expenditures.

15. Transition probabilities in a Markov process are assumed to be constant. In particular, they do not depend on state probabilities of preceding periods, hence $p_{ij} = P(X_{t+1} = j | X_t = i)$.

Figure 1
Markov Process of Risk States



The transition probabilities from Figure 1 are given in the transition matrix A,

$$A = \begin{bmatrix} p_{11} & p_{12} & p_{13} & 0 \\ p_{21} & p_{22} & 0 & p_{24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Pauly et al. deal with a special case of A, namely the case $p_{22} = 1$, $p_{21} = 0$ (no return from high-risk status) and $p_{13} = p_{24} = 0$ (no transition to death). These restrictions will be relaxed step by step.

1.1 Modification No. 1: Positive Probability of Returning to Low Risk (No Absorbing Death)

Neglecting the two absorbing states of death in Figure 1 but assuming a positive probability of returning to the favorable risk status $p_{21} > 0$, the transition matrix A reduces to matrix B,

$$B = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{2}$$

Including a positive probability of returning to low-risk status changes the optimal premium schedule substantially, because the share of high risks in the population develops in a more moderate way. Following Pauly et al., we assume an initial distribution comprising of only low risks (the row vector $b^0 = [1,0]$ below), the t^{th}-period state probabilities are then

$$\underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{=b'} \cdot \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{(t-1)} \tag{3}$$

For example, the state probabilities of Period 3 are given by

$$\underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{=b'} \cdot \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{2} = \begin{bmatrix} \underbrace{(p_{11}^2 + p_{12}p_{21})}_{\text{Prob. low risk}} & \underbrace{(p_{11}p_{12} + p_{12}p_{22})}_{\text{Prob. high risk}} \end{bmatrix}$$

To illustrate with a numerical example and to compare with Pauly et al., we retain Pauly et al. values $p_{11} = 0.9$ and $p_{12} = 0.1$, respectively. Moreover, $p_{21} = 0.25$ (and hence $p_{22} = 0.75$, again in accordance with Pauly et al.).[16] Table 1 displays the development of the shares of high and low risks over five periods in the Pauly et al. model and in our first modification. The overall loss probability is derived by weighting the state-dependent loss probabilities (p_l and p_h) with their respective population shares. Intuitively, a positive probability of returning to the favorable risk status should lower long-run premiums, causing the amount of frontloading to be lower (see Table 1). The lifetime actuarially fair premium, calculated for $L = H = 100$, is $P_{Paulyetal.} = 68.098$ according to the Pauly et al. model and $P_{mod.} = 64.139$ after modification, respectively.

Table 1
Pauly et al. Model (Left) and Modification No. 1 (Right)

Period	PKH :	$p_{21} = 0$	Loss Pr.	Modified :	$p_{21} = 0.25$	Loss Pr.
	High Risks	Low Risks		High Risks	Low Risks	
1	0	1	0.1	0	1	0.1
2	0.1	0.9	0.12	0.1	0.9	0.12
3	0.1900	0.8100	0.13800	0.1650	0.8350	0.133
4	0.2710	0.7290	0.15420	0.20725	0.79275	0.14145
5	0.3439	0.6561	0.16878	0.23471	0.76529	0.14694
	$P_{PKH} = 68.098^*$			$P_{Mod.} = 64.139^*$		

* Following PKH, there is no discounting (zero interest rate). The parameter values used are $p_{12} = p_l = 0.1$ ($= p_L$ in PKH) and $p_{21} = 1 - p_h = 0.25$ ($p_H = 0.3$ in PKH). Loss is $L = H = 100$.

As can be seen from Table 1, the modification $p_{21} = 0.25 > 0$ does not have any effect during the first two periods. However, from then on, the share of high risks

16. In a binary distribution with probabilities π, $(1-\pi)$, the mean waiting period for transition from state 1 to state 2 is $D = 1/\pi$ (see, e.g., Bhattacharyya and Johnson, 1977). Pauly et al. (1998) assume $D = 4$ in their guaranteed renewability group insurance model, implying $\pi = p_{21} = 0.25$.

approaches 0.23471 rather than 0.3439. The overall loss probability begins to change in Period 3 already, converging to 0.14694 rather than 0.16878. The lifetime premium drops by 5.8%, from 68.098 to 64.139. Note that a two-period model would fail to indicate this change, being subject to the restriction that the premium of the last period must be equal to the actuarially fair premium for a low-risk individual. Otherwise, low risks would not take out insurance. However, because the last-period premium and the probability of transition from low- to high-risk status are given while $p_{21} > 0$ is not relevant yet, the premium in the second-to-last period is determined, as well, and cannot differ from the value calculated by Pauly et al. In order to see the effect of $p_{21} > 0$, the Markov process needs to go on for at least three periods. Conversely, extending to more than five periods would increase complexity without adding further insights.

Expected health care expenditure (EHCE) in period t is given by

$$EHCE_t = Prob_{t,low} \cdot EHCE_{low} + Prob_{t,high} \cdot EHCE_{high} \tag{4}$$

where $Prob_{t,low}$ is the probability of being in the low-risk status in period t (and defined analogously for the high-risk status). Expected health care expenditures pertaining to a low risk is $EHCE_{low} = p_l L$, with p_l denoting the constant probability of loss. Expected health care expenditures for a high risk is defined in the same way.

Using (3), one obtains EHCE for each period,

$$EHCE_1 = p_l L \tag{5}$$

$$EHCE_2 = p_{11} p_l L + p_{12} p_h H \tag{6}$$

$$EHCE_3 = (p_{11}^2 + p_{12}p_{21}) p_l L + (p_{11}p_{12} + p_{12}p_{22}) p_h H \tag{7}$$

$$\begin{aligned} EHCE_4 =& (p_{11}^3 + 2p_{11}p_{12}p_{21} + p_{12}p_{21}p_{22}) p_l L \\ &+ (p_{11}^2 p_{12} + p_{12}^2 p_{21} + p_{12}p_{22}^2 + p_{11}p_{12}p_{22}) p_h H \end{aligned} \tag{8}$$

$$\begin{aligned} EHCE_5 =& (p_{11}^4 + 3p_{11}^2 p_{12}p_{21} + 2p_{11}p_{12}p_{21}p_{22} + p_{12}^2 p_{21}^2 + p_{12}p_{21}p_{22}^2) p_l L \\ &+ (p_{11}^3 p_{12} + p_{11}^2 p_{12}p_{22} + 2p_{11}p_{12}^2 p_{21} + p_{11}p_{12}p_{22}^2 + 2p_{12}^2 p_{21}p_{22} + p_{12}p_{22}^3) p_h H \end{aligned} \tag{9}$$

Guaranteed renewability premiums must reflect expected health care expenditures over remaining lifetime; e.g., the premium in Period 5 equals expected health care expenditures in the Period 1 ($P_5 = EHCE_1$; see Pauly et al.), and similarly for premiums in Period 1 through Period 4. This yields the following guaranteed.

$$\begin{aligned} P_1 =& p_{11}^3 p_l L \\ &+ p_{12}([p_{11}^2 p_{21} + p_{11}p_{21}p_{22} + p_{12}p_{21}^2 + p_{21}p_{22}^2] p_l L + [p_{11}p_{12}p_{21} + 2p_{12}p_{21}p_{22} + p_{22}^3] p_h H) \\ &+ p_{11}p_{12}([p_{11}p_{21} + p_{21}p_{22}] p_l L + [p_{12}p_{21} + p_{22}^2] p_h H) \\ &+ p_{11}^2 p_{12}(p_{21} p_l L + p_{22} p_h H) + p_{11}^3 p_{12}(p_h H - p_l L) \end{aligned} \tag{10}$$

$$P_2 = p_{11}^2 p_l L + p_{12}([p_{11}p_{21} + p_{21}p_{22}]p_l L + [p_{12}p_{21} + p_{22}^2]p_h H) + p_{11}p_{12}(p_{21}p_l L + p_{22}p_h H) + p_{11}^2 p_{12}(p_h H - p_l L) \tag{11}$$

$$P_3 = p_{11}p_l L + p_{12}(p_{21}p_l L + p_{22}p_h H) + p_{11}p_{12}(p_h H - p_l L) \tag{12}$$

$$P_4 = p_l L + p_{12}(p_h H - p_l L) \tag{13}$$

$$P_5 = p_l L \tag{14}$$

The guaranteed renewability premium schedule above is a competitive equilibrium (see Pauly et al., Proposition 1). The proof is based on backward induction. All individuals would be willing to pay at least $p_l L$ in the last period. Those individuals who did not suffer a loss in Period 4 would be willing to pay P_4 in Period 4, because the premium for Period 4 plus the premium in Period 5 equals their expected loss over both periods. P_4 has two parts, one that covers the expected loss for a low-risk person in Period 4, and another that covers the expected loss in Period 5 in excess of the premium P_5 $(p_h H - p_l L)$ for individuals who were still low-risk in Period 4 and turn high-risk in Period 5 with probability p_{12}.

Guaranteed renewability premiums importantly depend on weighted averages rather than the mere difference between expected health care expenditures pertaining to high and low risks. This is shown for Period 3 as follows. Adding $(+p_{12}p_l L - p_{12}p_l L) = 0$ to the RHS of (12) yields

$$P_3 = p_{11}p_l L + p_{12}(p_{21}p_l L + p_{22}p_h H) + p_{11}p_{12}(p_h H - p_l L) + p_{12}p_l L - p_{12}p_l L \tag{15}$$

Rearranging terms on the RHS, one obtains

$$\begin{aligned} P_3 &= (p_{11} + p_{12})p_l L + p_{12}(p_{21}p_l L + p_{22}p_h H - p_l L) + p_{11}p_{12}(p_h H - p_l L) \\ &= p_l L + p_{12}(p_{21}p_l L + p_{22}p_h H - p_l L) + p_{11}p_{12}(p_h H - p_l L) \end{aligned} \tag{16}$$

The first term on the RHS of (16) covers expected health care expenditures of a low risk during the current period. The second term covers the risk of an insured turning into a high risk in Period 2, the second-to-last period of the contract. The fact that she or he may become a low risk again in Period 3 is accounted for by taking a weighted average of the cost pertaining to low and high risk, respectively. The last term covers the risk of a person turning into a high risk in Period 3, in full analogy with Pauly et al.

Taking again the Period 3 premium as the example, one can now compare the situation with a positive probability of returning to low-risk status (i.e., P_3 above) with the one in Pauly et al., where a high risk always remains a high risk, and therefore

$$P_{3,Paulyetal.} = p_l L + p_{12}(p_h H - p_l L) + p_{11}p_{12}(p_h H - p_l L) \tag{17}$$

which is essentially P_3 above with $p_{21} = 0$ and $p_{22} = 1$. Now comparing the two Period 3 premiums, one obtains from (16) and (17)

$$\begin{aligned} P_{3,Paulyetal.} - P_3 &= p_{12}(p_hH - p_lL) - p_{12}(p_{21}p_lL + p_{22}p_hH - p_lL) \\ &= p_{12}[p_hH - p_{21}p_lL - p_{22}p_hH] \\ &= p_{12}[(1 - p_{22})p_hH - p_{21}p_lL] \\ &= p_{12}[(1 - p_{22})p_hH - (1 - p_{22})p_lL] \\ &= p_{12}[(1 - p_{22})(p_hH - p_lL)] > 0 \end{aligned} \tag{18}$$

because $H \geq L$ and $p_h > p_l$. As a result, in this more realistic setting of a positive probability of returning to low-risk status, the guaranteed renewability premium is always lower than the one calculated by Pauly et al.

1.2 The Steady-State Probability Distribution

In Table 1, only five periods were considered. To find the long-term steady-state probability distribution (assuming ergodicity of the Markov process), matrix B is used and the long-run steady-state probabilities p^*_l and p^*_h are calculated. The steady-state distribution vector v of an ergodic Markov process satisfies

$$v' \times B = v' \tag{19}$$

Therefore, the following system of equations solves for the steady-state distribution vector v,

$$\begin{aligned} p_l^* + p_h^* &= 1 \\ \begin{bmatrix} p_l^* & p_h^* \end{bmatrix} \times \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} &= \begin{bmatrix} p_l^* & p_h^* \end{bmatrix} \end{aligned} \tag{20}$$

Solving for p_l^* and p_h^* yields

$$\begin{aligned} p_l^* &= \frac{p_{21}}{1 + p_{21} - p_{11}} = 0.714 \\ p_h^* &= \frac{p_{12}}{1 + p_{12} - p_{22}} = 0.286 \end{aligned} \tag{21}$$

In the long run, 71.4% of the insureds belong to the favorable low-risk and 28.6% to the unfavorable high-risk category, respectively, representing the insurance company's long-term health risk profile. This result is consistent with Pauly et al. (1998).

1.3 Modification No. 2: The Impact of Death and Extra Cost of Dying

During the past few years, there has been a growing literature revolving around the red herring hypothesis which claims that the influence of age on health care expenditures is dwarfed by that of closeness to death (e.g., see Lubitz and Riley, 1993; Zweifel et al., 1999; Werblow et al., 2007; and Steinmann et al., 2007). The red herring hypothesis suggests an extra cost of dying ($C_h, C_l > 0$) which occurs during the last year of people's lives. It differs from regular health care expenditures in two ways: First, it almost always fails to restore health; second, it is roughly twice as high as regular health care expenditures (see, e.g., Zweifel et al., 2004). For example, the extra cost of dying as a low risk, C_l, may be the cost associated with death that arises after a car accident (short stay in the hospital), while the extra cost of dying as a high risk, $C_h > C_l$, may be the cost associated with a heart attack or stroke (calling for a longer and more expensive stay in the intensive care unit of the hospital before death occurs).

In the Pauly et al. model, the extra cost of dying ($C_h, C_l > 0$) is neglected, implying biased estimates of future health care expenditures. Remember that under modification no. 1 (assuming a positive probability of returning to low-risk status), the guaranteed renewability premium is always below (or equal to) the premium calculated by Pauly et al. The distribution of health care expenditures across the periods depends on the probabilities and the number of periods taken into account in the model. Now, two effects are at work: 1) the probability of returning to low-risk status (reducing health care expenditures); and 2) the extra cost of dying (increasing health care expenditures).

In the model (and reflecting reality), the premium is paid before health care expenditures occurs; next, risk status may change before the start of Period 2, and so on. In analogy to Pauly et al., our argument always assumes that the individual starts in the low-risk state at time $t = 1$ with loss probability p_l, while the analysis stops at time $t + 1$; i.e., in Period 2. In a two-period model, premiums must satisfy the following restriction,

$$P_1 + P_2 = p_l L + p_{11} p_l L + p_{12} p_h H \tag{22}$$

A one-period contract signed at the start of Period 2 requires the premium for low risks to be no higher than their expected cost because everyone starts as a low risk by assumption. Therefore $P_2 = p_l L$. This in turn implies that the premium in Period 1 is given by

$$P_{1,Paulyetal.} = p_l L + p_{12}(p_h H - p_l L) \tag{23}$$

where $P1,Paulyetal.$ is the premium in Period 1 and $p11 = 1 - p12$ because $p13 = 0$. $P1,Paulyetal.$ is the premium derived in Pauly et al.

Now introduce death as an absorbing state. For singling out this fact, suppose that high risks cannot become low risks again ($p_{21} = 0$).

Assuming death occurs after H has been spent on the high risk and L on the low risk, respectively, death takes place before the premium P_2 is paid, so only survivors pay it in Period 2, but everyone is offered a contract in Period 1 (because the change in risk status occurs after the premium is paid). This implies the following restriction on premium income,

$$P_1 + (1 - p_{13})P_2 = p_l L + p_{11} p_l L + p_{12} p_h H + p_{13} C_l + p_{11} p_{13} C_l + p_{12} p_{24} C_h \tag{24}$$

where C_h is the extra cost of dying as a high risk (see footnote 5 again). Moreover, P_2 must now also cover the extra cost of dying as a low risk,

$$P_2 = p_l L + p_{13} C_l \tag{25}$$

Substituting, one obtains for the guaranteed renewability premium P_1,

$$\begin{aligned} P_1 &= p_{11} p_l L + p_{12} p_h H + p_{13}(p_l L + p_{13} C_l) + p_{11} p_{13} C_l + p_{12} p_{24} C_h \\ &= p_l L + p_{12}(p_h H - p_l L) + p_{13} C_l + p_{12}(p_{24} C_h - p_{13} C_l) \end{aligned} \tag{26}$$

Note that the premium P_1 is lower than or equal to $P_{1,PKH}$ given in (23) as long as there is no excess cost of dying ($C_l = C_h = 0$). The first two terms of (26) serve to cover expected health loss, with the first term referring to a low risk during Period 1 and the second, to the possible increase in expected cost weighted by the probability of becoming a high risk in Period 2. With $C_h, C_l > 0$, however, there are two additional terms reflecting death as the terminal state. The first is for the expected extra cost of dying in Period 1 as a low risk, the second for the possible increase or decrease in the extra cost of dying due to a transition from low- to high-risk status at the end of Period 1.

Due to the positive probability of death ($p_{13} > 0$), the guaranteed renewability premium with death will be lower in the Period 1 than calculated by Pauly et al. ($P_{1,Paulyetal.}$), generating a higher demand for guaranteed renewability insurance contracts than in Pauly et al. Indeed, comparison of (26) with (23) reveals that

$$P_{1,Paulyetal.} - P_1 = (p_h H - p_l L)(p'_{12} - p_{12}) - p_{13} C_l - p_{12}(p_{24} C_h - p_{13} C_l) \tag{27}$$

where p'_{12} denotes the probability of switching health status in the Pauly et al. model and p_{12} the probability of switching health status in the present model. Note that (27) is positive, resulting in a higher demand for guaranteed renewability insurance contracts than in Pauly et al.:

1. The expected cost differential ($p_hH - p_lL$) is large, positive probabilities of death ($p_{13} > 0, p_{24} > 0$) strongly reduce the probability of transition from low- to high-risk status ($p'_{12} - p_{12}$is large), C_l is small, and the expected cost differential ($p_{24}C_h - p_{13}C_l$) is small;
2. The expected cost differential ($p_hH - p_lL$) is large, positive probabilities of death ($p_{13} > 0, p_{24} > 0$) strongly reduce the probability of transition from low- to high-risk status ($p'_{12} - p_{12}$is large), C_l is small, and $C_h/C_l < p_{13}/p_{24}$.

The second situation is unlikely to be satisfied because $p_{13}/p_{24} < 1$ in general and $C_h/C_l > 1$. Focusing on the first situation, the crucial condition is the force of competing risks. They must be effective enough to make the transition from low- to high-risk status unlikely, because they cause people who are low risks to die rather frequently. But provided this crucial condition is satisfied, the requirement ($p_hH - p_lL$) to be large is rather realistic. For instance, the cost weight attributed by U.S. Medicare to "heart transplant or heart assisting system" [Code 104 of diagnosis-related group (DRG)] is 19.55. Even for a given condition, a high risk can cost more than the double of a low risk, as exemplified by "pancreas, liver, and shunt procedures with complications" (Code 192) with cost weight 4.05, compared to "pancreas, liver, and shunt procedures without complications" (Code 193) with cost weight 1.63.[17]

1.4 Modifications Combined: Positive Probability of Returning to Low Risk, Death and Extra Cost of Dying

While a positive probability of returning to the low-risk state ($p_{21} > 0$) decreases the guaranteed renewability premium, the introduction of death as an absorbing state in combination with a high cost of dying increases it. The final question to be addressed is whether the *net effect* of the two modifications of the Pauly et al. model is to increase or decrease the guaranteed renewability premium and, hence, to undermine or facilitate the viability of guaranteed renewability in health insurance. The analysis above has revealed that for $p_{21} > 0$ to make a difference, at least three periods need to be analyzed (see Table 1). Therefore, a three-period model is examined in this section. As before, only survivors pay premiums. Total premium income over three periods amounts to

$$P_1 + (1 - p_{13})P_2 + [p_{11}(1 - p_{13}) + p_{12}(1 - p_{24})]P_3 \tag{28}$$

Compared to (24), the extra third term states that Period 3 premium income comes from two sources: 1) from low risks who survived both periods; and 2) from

17 . See *https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/MS-DRG-Classifications-and-Software.*

low risks turned high risks who survived (recall that initially everyone is a low risk by assumption).

Consider a policyholder who signs a health insurance contract in the Period 1. On the one hand, the upfront loading is highest in this period; on the other hand, guaranteed renewability requires each premium component of (28) to cover its pertinent expected health care expenditures. When the expected cost of dying is accounted for, the Period 3 premium becomes

$$P_3 = p_l L + p_{13} C_l \tag{29}$$

Adjusting (26) to include the possibility of returning to low-risk status implies

$$P_1 + (1-p_{13})P_2 = p_l L + p_{11} p_l L + p_{12}(1-p_{24}) p_h H + p_{12} p_{24} C_h + p_{13} C_l + p_{12} p_{21} p_l L + p_{12} p_{21} p_{13} C_l \tag{30}$$

The guaranteed renewability premium $P_{comb.}$ that combines both modifications then amounts to starting from (28), using (40) for the first two terms, and adding (29),

$$\begin{aligned} P_{comb.} \quad = \quad & p_l L + p_{11} p_l L + p_{12}(1-p_{24}) p_h H + p_{12} p_{24} C_h + p_{13} C_l + p_{12} p_{21} p_l L + p_{12} p_{21} p_{13} C_l \\ & + [p_{11}(1-p_{13}) + p_{12}(1-p_{24})][p_l L + p_{13} C_l] \end{aligned} \tag{31}$$

Collecting terms, one obtains

$$\begin{aligned} P_{comb.} \quad = \quad & \left[1 + p_{11} + p_{12} p_{21} + p_{11}(1-p_{13}) + p_{12}(1-p_{24})\right] p_l L + \left[p_{12}(1-p_{24})\right] p_h H \\ & + p_{13}\left[1 + p_{12} p_{21} + p_{11}(1-p_{13}) + p_{12}(1-p_{24})\right] C_l + p_{12} p_{24} C_h \end{aligned} \tag{32}$$

In the following, matrix $A_{comb.}$ represents realistic values assigned to transition probabilities. Using values similar to the matrices above, one may set $p_{13} = 0.05$, a conservative value because the mortality rate of the subpopulation age 45 or younger is about 0.002 p.a. in industrial countries (OECD health statistics, 2015). At the same time, p_{12} needs to be increased (from 0.2 to 0.3) to reflect the fact that there is no transition to Medicare at age 65 here. This implies $p_{11} = 0.65$. As to p_{21}, it is set to 0.25 as before, while $p_{24} = 0.2$, implying $p_{22} = 0.55$.

$$A_{comb.} = \begin{bmatrix} 0.65 & 0.3 & 0.05 & 0 \\ 0.25 & 0.55 & 0 & 0.2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Inserting the values from matrix $A_{comb.}$ into (32) results in

$$P_{comb.} = 2.5825 p_l L + 0.24 p_h H + 0.096625 C_l + 0.06 C_h \tag{33}$$

Using the guaranteed renewability premium determined by Pauly et al. as stated in (17) above and with p_{12} defined above (27), one obtains

$$\begin{aligned} P_{3,Paulyetal.} &= p_l L + p_{12}'(p_h H - p_l L) + p_{11} p_{12}'(p_h H - p_l L) \\ &= 0.4225 p_l L + 0.58 p_h H \end{aligned} \tag{34}$$

Collecting terms results in

$$P_{comb.} - P_{3,Paulyetal.} \approx 2.16 \cdot p_l L - 0.34 \cdot p_h H + 0.0967 \cdot C_l + 0.06 \cdot C_h \tag{35}$$

Inserting $C_l = 2p_l L$ and $C_h = 2p_h H$ as approximations from Felder et al. (2004), this amounts to

$$P_{comb.} - P_{3,Paulyetal.} \approx 2.3534 \cdot C_l - 0.22\, C_h \tag{36}$$

Therefore, for this estimated difference to become positive, the cost of dying as a high-risk patient C_h would have to be 10.7 (= 2.3534/0.22) times as high as that of treating a low-risk patient C_l. Remember from above that even for a given condition, a high risk can cost significantly more than the double of a low risk, as exemplified by "pancreas, liver, and shunt procedures with complications" (Code 192) with cost weight 4.05, compared to "pancreas, liver, and shunt procedures without complications" (Code 193) with cost weight 1.63.[18] However, a cost weight of 10 times more is rather rare; thus, a realistic set of parameter values suggests that a positive probability of returning to low-risk status combined with death and its associated extra cost may indeed very well lower the guaranteed renewability premium compared to the Pauly et al. benchmark.

1.5 Guaranteed Renewability Insurance and Capital Market Imperfections

In the Pauly et al. model, perfect capital markets allow consumers to obtain their most preferred consumption stream, and the premium sequence is optimal in the sense of Pareto. Period 1 equals expected losses in that period plus some front-loading, which solves problems associated with renegotiation after a later change in

18 . See *https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/MS-DRG-Classifications-and-Software.*

risk status. However, capital market constraints may prevent consumers from borrowing sufficient funds to pay this premium, a concern noted by Frick (1998), who argues that the Pauly et al. premium contains a front-loading that may be too large to be affordable in practice. He shows that imperfect capital markets combined with a high subjective rate of time preference may result in consumers' failure to purchase guaranteed renewability contracts.[19] This section reevaluates Frick (1998) by deriving a critical rate of time preference in the extended model where consumers stop buying the guaranteed renewability contract.

In the presence of capital market imperfections, consumers cannot use future income as a collateral to obtain credit. Therefore, the Period 1 guaranteed renewability premium must be entirely financed using Period 1 income. Let y denote the consumers' constant per-period income to finance premium payments. Assume further that consumers discount the future at the rate $0 < \beta \leq 1$.

The consumers' optimization problem over two periods reads[20]

$$\max_{P_1;P_2;\beta} \quad U(y - P_1) + \beta(1 - p_{13})U(y - P_2) \tag{37}$$

subject to

$$P_1 + P_2 = p_l L + p_{11} p_l L + p_{12} p_h H;\ p_{13} = 0, or \tag{38a}$$

$$\begin{aligned} P_1 + (1 - p_{13})P_2 = {} & p_l L + p_{11} p_l L + p_{12} p_h H \\ & + p_{13} C_l + p_{11} p_{13} C_l + p_{12} p_{24} C_h;\ p_{13} > 0 \end{aligned} \tag{38b}$$

where (38a) is the standard Pauly et al. constraint as discussed by Frick (1998), and (38b) includes a positive probability of death and extra cost of dying, as in (30).

With restriction (38a), the first-order conditions imply

$$\begin{aligned} U^0[y - P_1] &= \beta \cdot U^0[y - P_2] \\ P_1 + P_2 &= p_l L + p_{11} p_l L + p_{12} p_h H \end{aligned} \tag{39}$$

Remember that for low risks to participate in the guaranteed renewability series of contracts, their Period 2 contract must be the single-period contract that is actuarially fair. From (39), one can see that for the special case of $\beta = 1$, the Period 1 and Period 2 premiums would have to be identical in the optimum, which is not

19. Note that Pauly et al. mention capital market imperfections in their motivation, but their proof of Pareto optimality of the guaranteed renewability contract sequence explicitly assumes perfect capital markets.

20 As insurers can initiate a new guaranteed renewability sequence at the beginning of each new period, longer time horizons are variations of the two-period model, assuming $p_{21} = 0$.

compatible with guaranteed renewability. For a guaranteed renewability premium schedule to be viable, income in Period 1 needs to be significantly higher than Period 2 in order to capture the front-loading contained in the Period 1 premium (i.e., $y_1 > y_2$). Inserting the values from Pauly et al., $P_{2,Paulyetal.} = p_l L$ and $P_{1,Paulyetal.} = p_l L + p_{12}(p_h H - p_l L)$ into (39), one finds a critical subjective rate of time preference

$$\frac{U'[y_1 - p_l L - p_{12}(p_h H - p_l L)]}{U'[y_2 - p_l L]} = \bar{\beta}_{Paulyetal.} \tag{40}$$

Provided y_1 is sufficiently high, (40) indicates a threshold $\bar{\beta} < 1$, separating those low risks who choose a guaranteed renewability sequence from those who do not.[21]

With restriction (38b), the first-order conditions imply

$$\begin{aligned} U^0[y_1 - P_1] &= \beta \cdot U^0[y_2 - P_2] \\ P_1 + (1 - p_{13})P_2 &= p_l L + p_{11} p_l L + p_{12} p_h H + p_{13} C_l + p_{11} p_{13} C_l + p_{12} p_{24} C_h \end{aligned} \tag{41}$$

Inserting (25) and (26) for GR, one finds a critical subjective rate of time preference

$$\frac{U'[y_1 - p_l L - p_{12}(p_h H - p_l L) - p_{13} C_l - p_{12}(p_{24} C_h + p_{13} C_l)]}{U'[y_2 - p_l L - p_{13} C_l]} = \bar{\beta}_{comb.} \tag{42}$$

Remember that the premium P_1 is lower or equal to $P_{1,Paulyetal.}$ as long as there is no extra cost of dying ($C_l = C_h = 0$). As long as the extra cost of dying is not excessively high, the critical time preference $\bar{\beta}_{comb.}$ is lower than $\bar{\beta}_{Paulyetal.}$ in (41) because in (42), the numerator is somewhat higher but the denominator clearly higher than in (41). However, a lower critical time preference implies sustainability of guaranteed renewability. As a result, guaranteed renewability contracts are likely to be sustainable in the presence of death.

1.6 Impacts of Increasing Health Care Expenditures and Interest

An issue confronting insurers and policy makers is the unrelenting increase in health care expenditures. To address this issue, let health care expenditures increase with a rate $g > 1$ per period (assumed to be the same for high and low risks). However, this increase is balanced to some extent if a rate of interest $i > 0$ is paid on the Period 1 premium, rendering the guaranteed renewability premium more

21. Note that $y1$ may be the result of some state-financed subsidy for purchasing health insurance or a tax exemption.

affordable. When these two additional modifications are combined, the budget constraints in the Pauly et al. and extended model become, respectively,

$$(1+i)P_1 + P_2 = (1+i)p_l L + p_{11} p_l L(1+g) + p_{12} p_h H(1+g);\ p_{13} = 0$$

$$(1+i)P_1 + (1-p_{13})P_2 = (1+i)p_l L + (1+i)p_{13} C_l + p_{11} p_l L(1+g) \tag{43a}$$

$$+ p_{12} p_h H(1+g) + p_{11} p_{13} C_l(1+g) + p_{12} p_{24} C_h(1+g);\ p_{13} > 0 \tag{43b}$$

The optimization problem composed of (37) and the restrictions (43a) or (43b) are now examined using a numerical example (a closed solution for the critical value $\bar{\beta}$ is not available).

Table 2
Parameter Values Assumed in the Numerical Example*

No death (Pauly et al.)							With death					
p_{11}	=	0.9	L	=	20		p_{11}	=	0.9	L	=	20
p_{12}	=	0.1	H	=	20		p_{12}	=	0.07	H	=	20
p_{13}	=	0	y_1	=	100		p_{13}	=	0.03	y_1	=	100
p_{21}	=	0.75	i	=	0.1		p_{21}	=	0.65	i	=	0.1
p_{22}	=	0.25	g	=	0.1		p_{22}	=	0.25	g	=	0.1
p_{24}	=	0					p_{24}	=	0.1	C_l	=	5
p_l	=	0.1					p_l	=	0.1	C_h	=	10
p_h	=	0.25					p_h	=	0.25	p_{death}	=	0.047

* Following Frick (1998), logarithmic utility is assumed.

The calculation proceeds as follows (see the Appendix for details). Suppose $\beta = 1$; then unconstrained individuals equate marginal utilities across periods. In order to check whether an unconstrained solution is possible ($\beta = 1$), the maximum premium that a low-risk consumer is willing to pay in Period 2 is substituted into the restrictions, thus ensuring that premiums cover expected cost.

The lowest possible value of time preference compatible with unconstrained optimization can be derived by dividing Period 1 by Period 2 marginal utility, with both premiums set to their actuarially fair per-period values (the fair per-period premium in Period 2 is expected health care expenditures of low risks, while any remaining health care expenditures is paid in Period 1),

$$\frac{U'[y - P_{1,fair}]}{U'[y - P_{2,fair}]} = \bar{\beta}. \tag{44}$$

To keep the results comparable to Pauly et al. (1995) and Frick (1998), Table 2 contains their parameter values in the left column. Inserting the values given in

Table 2 into (43a), one obtains[22] $(1 + i)P1_{,Paulyetal.} + P2_{,Paulyetal.} = (1 + i)plL + p11plL(1 + g) + p12phH(1 + g)$ $P1_{,Paulyetal.} = 5.78$.

Every consumer is constrained in the Pauly et al. case because $P_{1,Paulyetal.} = 5.78 > P_{2,Paulyetal.} = p_lL(1 + g) = 2.2$.

Generally, there is a critical combination of cost growth (g) and interest rate (i) that is compatible with a guaranteed renewability premium over two periods.[23]

One can differentiate between three groups of consumer types by identifying critical β-values for consumer types:

- Consumers with strong preference for current consumption ($\beta < 0.526$) will not buy guaranteed renewability insurance because the guaranteed renewability surcharge contained in their Period 1 premium exceeds their expected utility gain from a guaranteed renewability contract.
- Consumers with a moderate to low preference for current consumption ($0.526 \leq \beta < 0.999$) do buy guaranteed renewability insurance but would prefer shifting more of the premium burden to the later period.
- Very patient consumers $0.999 \leq \beta \leq 1$ also opt for the guaranteed renewability contract. They would not even want to shift more of the premium burden to the later period.

Note that this third set would be empty if death had not been considered; therefore, guaranteed renewability not only survives but may actually thrive in the presence of death.

Concluding Remarks

This article extends the seminal work by Pauly et al. (1995) on guaranteed renewability of health insurance contracts. Pauly et al. implicitly assume that there is only one possible transition, which moreover is permanent: from low-risk status to high-risk status. This neglects two important facts. First, several individuals may return to low-risk status, as has been documented by Beck et al. (2010) for an observation period of five years (taken into account by Pauly et al. (1998) in their calibration of guaranteed renewability contracts in U.S. group health insurance). Second, there is always the transition to death, associated with an extra cost of dying,

22. An alternative way of arriving at the same result is to set premiums in both periods equal and solve for a uniform premium across periods using the budget constraint. If the result exceeds the maximum possible premium in the later period, then even the most patient individuals (i.e., those with $\beta = 1$) are restricted in their optimization.

23. Any individual with $\beta = 1$ prefers uniform premiums, while those with $\beta < 1$ prefer $P1 < P2$. If guaranteed renewability uniform premiums are not viable because if $P2 < P1$ always holds, every individual will be constrained.

especially among the high risks (Steinmann et al., 2007). Because U.S. mortality in the relevant age groups is about 0.0034 p.a., the probability of surviving until age 65 is approximately 84% (0.8434 = 0.9966),[50] amounting to an overall mortality risk of at least 15%. Therefore, even though private health insurance contracts expire at age 65 with the transition to Medicare in the U.S., insurers have to cover the cost of dying with a non-negligible probability. This is, of course, even more true for private health insurers in other countries with their lifetime contracts. While the possible return to low-risk status has a high potential to lower the guaranteed renewability premium, the extra cost of dying is likely to increase it. This article shows that, on balance, these two facts reduce the actuarially fair guaranteed renewable premium in realistic conditions, suggesting that prior studies have overestimated the economic cost of guaranteed renewability, making it more affordable and accessible in practice.

Several more insights can be derived from the analysis performed in this paper. A first insight is that it takes at least three periods for a positive probability of returning from high- to low-risk status ($p_{21} > 0$) to make a difference. The reason is two restrictions on the guaranteed renewability premium. First, it must be actuarially fair to low risks during the last period of their lives in order to be attractive to them. Second, $p_{21} > 0$ cannot be relevant anymore in the last period. Therefore, if the last period is Period 2 as in a two-period model, the two components of the guaranteed renewability premiums fail to reflect $p_{21} > 0$. For three and more periods, however, the guaranteed renewability premium is indeed lower than the one determined by Pauly et al. Conversely, the introduction of death as a terminal state does increase the guaranteed renewability premium compared to Pauly et al., in particular if the extra cost of dying is high due to heroic efforts to prolong the lives of high-risk patients. An exception is possible if the probability of transition from low-risk to high-risk status is lowered by the possibility of death acting as a competing risk.

When health insurance premiums are risk-based and do not protect against reclassification risk, younger and healthier individuals may not purchase coverage, while older and sicker individuals have no choice and are left with excessively high premiums as they grow older. If guaranteed renewability becomes more affordable and accessible, younger individuals may have a higher incentive to "lock in" their health insurance premiums, and such a market failure may more easily be overcome by introducing guaranteed renewability as a contract feature into an existing risk-based system. Hence, as a result of the analysis in this paper, introducing guaranteed renewability may be a way to ensure long-term health insurance coverage without risk selection issues for an entire population. For regulators concerned about adverse selection and "cream skimming," the alternative to guaranteed renewability would be to maintain strong regulatory rules, which entail community rating and limits on risk rating. Guaranteed renewability can alleviate market failures due to adverse selection and, at the same time, requires less regulation than in a purely risk-rated system.

Our extensions are confirmed to hold in practice by Herring and Pauly (2006), acknowledging the fact that people may recover from being a high risk. They write in their conclusion: "We found that the amount of front-loading necessary to effectively fund guaranteed renewability insurance is mitigated by several factors: low-risk expected expenses increase with age, the likelihood of becoming high risk increases with age, and high-risk people either recover or die." Regarding the third point, Herring and Pauly use data from the MEPS to estimate differences in health care expenditures immediately after becoming a high risk versus five years after becoming a high risk, which addresses both changes in risk status and cost of death.[24]

The findings presented in this study are subject to several qualifications. First, long-term sustainability is not always guaranteed in real-world health insurance contracts. Second, the effect of death is two-fold in the extended Pauly et al. model; i.e., the higher health care expenditures during the period of death is explicitly modeled, but the lowered total health care expenditures (after all, the dead do not cause health care expenditures) is implicitly modeled. Third, actuarially fair premiums do not exist in health insurance; proportional loadings are common. Proportional loadings are known to curtail the demand for insurance generally; however, risk aversion has been found to increase with age, in particular after retirement (see, e.g., Halek and Eisenhauer, 2001), rendering guaranteed renewability especially attractive to older individuals. Yet, the upfront loading contained in the guaranteed renewability premium of older consumers is very high because they stand to benefit from a small number of periods where they return from high-risk to low-risk status. The net effect of these influences on the demand for guaranteed renewability is ambiguous, calling for additional investigation. Furthermore, demand is also curtailed in case of a high time preference of consumers, making them unwilling to bear the frontloading of guaranteed renewability premiums. This consideration is of particular importance when insurers have little scope for crediting upfront premium payments with accrued interest. At present, this is the case due to the quantitative easing policy of major central banks, which therefore serves to undermine the economic viability of guaranteed renewability. Yet, the basic finding that guaranteed renewability in health insurance can "survive death" when the possibility of a high-risk returning to low-risk status is taken into account is likely to be robust.

24. Although Herring and Pauly (2006) also account for the probability of death altering the length of time covering one's higher spending, the authors only implicitly illustrate these various effects by using real-world data to estimate the resulting premium path, without disentangling the relative contributions, while this study explicitly specifies the formulae and considers them separately.

Appendix

The calculation of the net gain in expected utility due to guaranteed renewability health insurance is considered here. As in Frick (1998), a logarithmic risk utility function is assumed. For the utility gained from insurance (4*EU*) in Period 1 and using the parameter values of Table 2, one obtains

$$
\begin{aligned}
\mathrm{M}\,EU_1 &= (1-p_{13})\{\ln[p_l(y-L)+(1-p_l)y]-p_l\ln[y-L]-(1-p_l)\ln[y]\} \\
&= 0.0020 \qquad \text{(A.1)}
\end{aligned}
$$

With cost growth g and interest i, one obtains for Period 2,

$$
\begin{aligned}
\mathrm{M}\,EU_2 = &\ \{(1-p_{13})p_{11}+(1-p_{24})p_{12}\} \\
&\ \{\ln[(p_{11}p_l+p_{12}p_h)(y-L(1+g))+(p_{11}(1-p_l)+p_{12}(1-p_h))y]\} \\
&-\{(1-p_{13})p_{11}+(1-p_{24})p_{12}\}\{(p_{11}p_l+p_{12}p_h)\ln[y-L(1+g)]\} \\
&-\{(1-p_{13})p_{11}+(1-p_{24})p_{12}\}\{(p_{11}(1-p_l)+p_{12}(1-p_h))\ln[y]\} \\
= &\ 0.1027 \qquad \text{(A.2)}
\end{aligned}
$$

$\mathrm{M}\,EU_2$ is much larger than $\mathrm{M}\,EU_1$ because every insured is a low risk in the Period 1, whereas the probability of loss is much higher in the next period. In total, the utility premium obtained is,

$$
\begin{aligned}
\mathrm{M}\,EU &= \mathrm{M}\,EU_1+\beta(1-p_{13})\,\mathrm{M}\,EU_2 \\
&= 0.0020+0.0996\beta>0 \qquad \text{(A.3)}
\end{aligned}
$$

The loss in expected utility caused by the binding restriction $P_2 = p_l L$ can be calculated by determining the optimal Period 1 premium P_1^* when the Lagrangian multiplier μ is set to zero in the optimization problem (37). This value is used to calculate the increase in expected utility $(\triangle\ EU_1^*>0)$ in Period 1, which is the difference between the optimized utility and the utility with
a binding restriction,

$$
P_1^* = \frac{106.7\beta-92.1211}{1.1+1.067\beta}; \qquad \text{(A.4)}
$$

$$
\mathrm{M}EU_1^* = \ln\left[100-\frac{106.7\beta-92.1211}{1.1+1.067\beta}\right]-4.5797. \qquad \text{(A.5)}
$$

However, in Period 2 the binding restriction causes a reduction in expected utility (M $EU_2^* < 0)$ compared to the value associated with the optimized value of P_2^*. This reduction is given by

$$P_2^* \quad = \quad 100 - 110\beta + \frac{1.1\beta(106.7\beta - 92.1211)}{1.1 + 1.067\beta}; \tag{A.6}$$

$$\mathrm{M}EU_2^* \quad = \quad \ln\left[110\beta - \frac{1.1\beta(106.7\beta - 92.1211)}{1.1 + 1.067\beta}\right] - 4.5812 \tag{A.7}$$

The total reduction of expected utility due to the binding constraint therefore amounts to

$$\mathrm{M}EU^* = \underbrace{\triangle\ EU_1^*}_{>0} + \beta(1 - p_{13})\underbrace{\triangle\ EU_2^*}_{<0} > 0 \tag{A.8}$$

The difference between (A.3) and (A.8) is the net gain in expected utility from having guaranteed renewability insurance,

$$\mathrm{M}_n EU = \mathrm{M}\, EU - \mathrm{M}\, EU^* \tag{A.9}$$

With the parameter values taken from Table 2, the critical value of β for M*n* $EU = 0$ is $\hat{\beta} = 0.526$.

References

Beck, K., M. Trottmann, and P. Zweifel, 2010. "Risk Adjustment in Health insurance and Its Long-Term Effectiveness," *Journal of Health Economics*, 29(4): 489–498.

Bhattacharyya, G.K., and R.A. Johnson, 1977. *Statistical Concepts and Methods*, New York, NY: Wiley.

Brown, H., and L.B. Connelly, 2005. "Lifetime Cover in Private Insurance Markets," *International Journal of Health Care Finance and Economics*, 5(1): 75–88.

Cameron, A.C., and P.K. Trivedi, 1991. "The Role of Income and Health Risk in the Choice of Health Insurance: Evidence from Australia," *Journal of Public Economics*, 45(1): 1–28.

Cochrane, J.H., 1995. "Time-Consistent Health Insurance," *Journal of Political Economy*, 103(3): 445– 473.

Cutler, D.M., and S.J. Reber, 1998. "Paying for Health Insurance: The Trade-Off Between Competition and Adverse Selection," *Quarterly Journal of Economics*, 113(2): 433–466.

Doherty, N.A., and P.D. Thistle, 1996. "Adverse Selection with Endogenous Information in Insurance Markets," *Journal of Public Economics*, 63(1): 83–102.

Felder, S., A. Werblow, and P. Zweifel, 2010. "Do Red Herrings Swim in Circles? Controlling for the Endogeneity of Time to Death," *Journal of Health Economics*, 29(2): 205–212.

Frick, K., 1998. "Consumer Capital Market Constraints and Guaranteed Renewable Insurance," *Journal of Risk and Uncertainty*, 16(3): 271–278.

Halek, M., and J.G. Eisenhauer, 2001. "Demography of Risk Aversion," *Journal of Risk and Insurance*, 68(1): 1–24.

Handel, B., I. Hendel, and M.D. Whinston, 2015. "Equilibria in Health Exchanges: Adverse Selection Versus Reclassification Risk," *Econometrica*, 83(4): 1261–1313.

Hendel, I., and A. Lizzeri, 2003. "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *Quarterly Journal of Economics*, 118(1): 299–328.

Herring, B., and M. Pauly, 2003. "Incentive-Compatible Guaranteed Renewable Health Insurance," National Bureau of Economic Research, NBER Working Paper No. 9888.

Herring, B., and M.V. Pauly, 2001. "Premium Variation in the Individual Health Insurance Market," *International Journal of Health Care Finance and Economics*, 1(1): 43–58.

Herring, B., and M.V. Pauly, 2006. "Incentive-Compatible Guaranteed Renewable Health Insurance Premiums," *Journal of Health Economics*, 25(3): 395–417.

Hofmann, A., and M. Browne, 2013. "One-Sided Commitment in Dynamic Insurance Contracts: Evidence from Private Health Insurance in Germany," *Journal of Risk and Uncertainty*, 46(1): 81– 112.

Manchikanti, L., D. Caraway, A.T. Parr, B. Fellows, and J.A. Hirsch, 2011. "Patient Protection and Affordable Care Act of 2010: Reforming the Health Care Reform for the New Decade," *Pain Physician*, 14(1): E35–E67. Accessed online at *https://www.painphysicianjournal.com/current/pdf?article=MTQyMw%3D%3D&journal=59*.

Lubitz, J.D., and G.F. Riley, 1993. "Trends in Medicare Payments in the Last Year of Life," *New England Journal of Medicine*, 328(15): 1092–1096.

Panthoefer, S., 2016. "Risk Selection Under Public Health Insurance with Opt-Out," *Health Economics*, 25: 1163–1181.

Patel, V., and M.V. Pauly, 2002. "Guaranteed Renewability and the Problem of Risk Variation in Individual Health Insurance Markets," *Health Affairs*, 22: W280–W289.

Pauly, M., 2012. "Wussinomics: The State of Competitive Efficiency in Private Health Insurance," *International Journal of Health Care Finance and Economics*, 12(3): 235–245.

Pauly, M., A. Nickel, A., and H. Kunreuther, 1998. "Guaranteed Renewability with Group Insurance," *Journal of Risk and Uncertainty*, 16(3): 211–221.

Pauly, M.V., H. Kunreuther, and R. Hirth, 1995. "Guaranteed Renewability in Insurance," *Journal of Risk and Uncertainty*, 10(2): 143–156.

Pauly, M.V., K. Menzel, H. Kunreuther, and R.A. Hirth, 2011. "Guaranteed Renewability Uniquely Prevents Adverse Selection in Individual Health Insurance," *Journal of Risk and Uncertainty*, 43(2): 127.

Schoder, J., M. Sennhauser, and P. Zweifel, 2010. "Fine-Tuning of Health Insurance Regulation – Unhealthy Consequences for an Individual Insurer," *International Journal of the Economics of Business*, 17(3): 313–327.

Shelton Brown, H., and L.B. Connelly, 2005. "Market Failure in Long-Term Private Health Insurance Markets: A Proposed Solution," *Applied Economics Letters*, 12(5): 281–284.

Steinmann, L., H. Telser, and P.S. Zweifel, 2007. "Aging and Future Healthcare Expenditure: A Consistent Approach," *Forum for Health Economics & Policy*, 10(2).

Werblow, A., S. Felder, and P. Zweifel, 2007. "Population Ageing and Health Care Expenditure: A School of 'Red Herrings'?" *Health Economics*, 16(10): 1109–1126.

Zweifel, P., and M. Breuer, 2006. "The Case for Risk-Based Premiums in Public Health Insurance," *Health Economics, Policy and Law*, 1(02): 171–188.

Zweifel, P., S. Felder, and M. Meier, 1999. "Ageing of Population and Health Care Expenditure: A Red Herring?" *Health Economics*, 8(6): 485–496.

Zweifel, P., S. Felder, and A. Werblow, 2004. "Population Ageing and Health Care Expenditure: New Evidence on the 'Red Herring,'" *Geneva Papers on Risk and Insurance: Issues and Practice*, 29(4): 652–666.

Journal of Insurance Regulation

Guidelines for Authors

Submissions should relate to the regulation of insurance. They may include empirical work, theory, and institutional or policy analysis. We seek papers that advance research or analytical techniques, particularly papers that make new research more understandable to regulators.

Submissions must be original work and not being considered for publication elsewhere; papers from presentations should note the meeting. Discussion, opinions, and controversial matters are welcome, provided the paper clearly documents the sources of information and distinguishes opinions or judgment from empirical or factual information. The paper should recognize contrary views, rebuttals, and opposing positions.

References to published literature should be inserted into the text using the "author, date" format. Examples are: (1) "Manders et al. (1994) have shown. . ." and (2) "Interstate compacts have been researched extensively (Manders et al., 1994)." Cited literature should be shown in a "References" section, containing an alphabetical list of authors as shown below.

Cummins, J. David and Richard A. Derrig, eds., 1989. *Financial Models of Insurance Solvency*, Norwell, Mass.: Kluwer Academic Publishers.

Manders, John M., Therese M. Vaughan and Robert H. Myers, Jr., 1994. "Insurance Regulation in the Public Interest: Where Do We Go from Here?" *Journal of Insurance Regulation*, 12: 285.

National Association of Insurance Commissioners, 1992. *An Update of the NAIC Solvency Agenda*, Jan. 7, Kansas City, Mo.: NAIC.

"Spreading Disaster Risk," 1994. *Business Insurance,* Feb. 28, p. 1.

Footnotes should be used to supply useful background or technical information that might distract or disinterest the general readership of insurance professionals. Footnotes should not simply cite published literature — use instead the "author, date" format above.

Tables and charts should be used only if needed to *directly support* the thesis of the paper. They should have descriptive titles and helpful explanatory notes included at the foot of the exhibit.

Papers, including exhibits and appendices, should be limited to 45 double-spaced pages. Manuscripts are sent to reviewers anonymously; author(s) and affiliation(s) should appear only on a separate title page. The first page should include an abstract of no more than 200 words. Manuscripts should be sent by email in a Microsoft Word file to:

Cassandra Cole and Kathleen McCullough
jireditor@gmail.com

The first named author will receive acknowledgement of receipt and the editor's decision on whether the document will be accepted for further review. If declined for review, the manuscript will be destroyed. For reviewed manuscripts, the process will generally be completed and the first named author notified in eight to 10 weeks of receipt.

Published papers will become the copyrighted property of the *Journal of Insurance Regulation*. It is the author's responsibility to secure permission to reprint copyrighted material contained in the manuscript and make the proper acknowledgement.