Can guaranteed renewability survive in the presence of death?

February 2, 2015

Abstract

Guaranteed renewability (GR) is an important feature of health insurance. It offers stability of premiums in the face of unexpected deterioration of health status. Although GR is a characteristic of social health insurance, it has recently been written voluntarily into policies by private insurers. Since death is the ultimate deterioration of health status, it seems questionable whether in the long run, insurers will be able to sustain GR rather than investing in risk selection activity. Extending the work of Pauly, Kunreuther, and Hirth (1995), this contribution finds that anticipation of death may actually cause demand for GR to increase in a number of circumstances, hence contributing to its sustainability.

JEL Classification: G22; I13.

Keywords: Health insurance, Guaranteed renewability, Long-term contracts.

Highlights:

- Guaranteed renewability (GR) disconnects premiums from deteriorations of health status over time
- With death, as the ultimate health loss, it is questionable whether GR is sustainable in the long run
- Risk selection competes with GR; yet it is risky because future health status can also improve: this fact is taken into account in this contribution
- A crucial finding is that anticipation of death enhances rather than undermines demand for GR

1 Introduction

Guaranteed renewability (GR) is an important feature of health insurance. It offers stability of premiums in the face of unexpected deterioration of health status. Although GR is a characteristic of social health insurance, it has recently been written voluntarily into policies by private insurers. In particular, private health insurers in the United States have been writing GR into their individual policies in response to pressure from the regulator.[1] Private insurers in Germany seek to be better able to compete with social health insurance where contributions vary with labor income but not with health status. Evidently, GR amounts to a commitment of the insurer to abstain from risk selection by replacing (typically older) clients who have become unfavorable risks by (typically younger) clients who are favorable risks. More generally, GR can be seen as the efficient alternative to premium regulation which (especially in the guise of community rating) creates an incentive for competitive insurers to engage in risk selection (Patel and Pauly, 2002). Policymakers have sought to neutralize this incentive by adding regulation in the guise of risk adjustment, with both little chance of success and unexpected negative side effects (Zweifel and Breuer, 2006; Schoder et al., 2010).

However, recent empirical research has identified an important challenge to GR: Healthcare expenditure (HCE) has been found to increase sharply with closeness to death (Zweifel et al., 1999), leading Steinmann et al. (2007) to distinguish between expenditure for restoring health and the cost of dying. Moreover, Felder et al. (2007) found that the increase in HCE 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 GR in a way that an empty set results. However, another empirical finding is of relevance in this context. Insureds who are high risk at a given point in time have a positive probability of becoming low risks later (Beck et al., 2010). Therefore, GR has the advantage of letting the insurer avoid the risk of losing a customer who may be a favorable risk in a future period.

The remainder of this paper is structured as follows. After a review of the a literature in Section 2, the Markov process specified by Pauly et al. (1995) is complemented by a positive transition probability from high to low risk status and a separate cost of dying in Section 3. The model is generalized in Section 4 to include a positive rate of interest as well as a positive rate of time preference. Section 5 specifies the new conditions defining the set of sustainable GR policies and comes up with suggestions for policy. A summary and conclusions follow in the last section.

2 Related literature

In long-term health insurance, there are two basic solutions for dealing with risk selection (i.e., the problem that high-risk individuals remain uninsured). One solution is to impose mandatory community rating by law, as it is the case in employer-contracted health insurance in the United States and individually-contracted health insurance in Switzerland. The other solution is to make contracts incentive-compatible. The latter, market-based solution has been investigated by Pauly, Kunreuther, and Hirth (1995, PKH henceforth), and Cochrane (1995). They derive long-term

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

insurance contracts that exhibit time consistency. In PKH, this sequence of incentive-compatible short-term contracts results in guaranteed renewability (GR). As pointed out by Cochrane (1995, p. 447), these contracts must be renegotiation-proof, satisfying a participation constraint in that all parties are always willing to sign the next contract under all future circumstances. Long-term insurance contracts are defined as a sequence of short-term contracts that are incentive compatible, in accordance with the participation constraint and renegotiation-proof. Applying game theory and backward induction, PKH derive a contract sequence with declining premiums.

GR contracts achieve time-consistency by frontloading premiums. As argued by Frick (1998), frontloading may be excessive for individuals who have limited capital endowment. Yet, the empirical evidence presented by Herring and Pauly (2006) suggests that this seems not a problem in practice. GR thus may be a way to ensure long-term health insurance for an entire population, provided however everyone is actually capable and willing to prepay the frontloading in GR insurance. The solution by Cochrane (1995) 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 expenditure ('HCE) or the excess of the PV of future expected HCE 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).

The second solution is prevalent in the U.S. individual health insurance market, where health insurers have the right to risk rate premiums. Risk-averse consumers buy health insurance voluntarily, with premium development ensuring time consistency. 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."

However, GR is unlikely to work perfectly either. In particular, individuals typically pay less than under perfect risk rating, presumably because insurers fear the bad publicity 'excessive premiums' may trigger.[2] Therefore, some adverse side effects of imperfect risk rating remain, e.g. a lock-in of high risks who opt for a GR contract [Patel and Pauly (2002)]. Still, GR is a market-based mechanism that at least partially overcomes the problem of uncertainty surronding long-term health status.

The starting point of this article is the optimal contract as in PKH, 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 GR. At least in the United States, the longest observable time horizon is defined by eligibility for Medicare (age 65). In countries without a scheme similar to Medicare, however, 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.[3]

2 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 United States.

3 Death should be made endogenous even for the case of U.S. medicare; for an attempt outside the U.S., see Zweifel et al. (1999).

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 health insurance (e.g. employer-contracted). In fact, a double pooling mechanism can be created (at the group level and at the insurer level, through having one insurer enrol several groups). Obviously, such a double pooling serves to reduce frontloading for GR.

Four recent empirical contributions deal with individual GR contracts. Brown and Connelly (2005) evaluate the Australian government's initiative to foster long-term private health insurance using a GR model with the probabilities of Herring and Pauly (2006) while they find Australia's lifetime cover to be subject to adverse selection. GR may constitute a voluntary alternative bcause it avoids loading hikes at higher ages in return for an upfront surcharge. In a second contribution, Shelton Brown and Connelly (2005) extend the PKH 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. They see substituting this cross-subsidization by risk-rated GR that protects consumers against future deterioration of their health status as a way to overcome this market failure. Herring and Pauly (2006) estimate the amount of frontloading in existing GR health insurance in the United States. They construct a risk-rated GR premium profile and compare it with the observed development of premiums during the life of the contract. They find an amazing degree of similarity between the predicted and the actual time paths. Hofmann and Browne (2013) analyze frontloading in existing GR private health insurance in Germany. They show that more frontloading is generally associated with lower rates of lapse, suggesting a lock-in of consumers.

In this paper, we add two new features to the PKH model. First, high-risk individuals may become low risk types again, reflecting pertinent findings by Beck et al. (2010); second, we introduce death as the 'absorbing state' (in Markovian language). These important modifications raise the question of whether the GR concept can survive in the face of death. The main finding is that the presence of death actually *enhances* the viability of GR because it causes the upfront GR surcharge to be *lower*.

3 The model and extensions

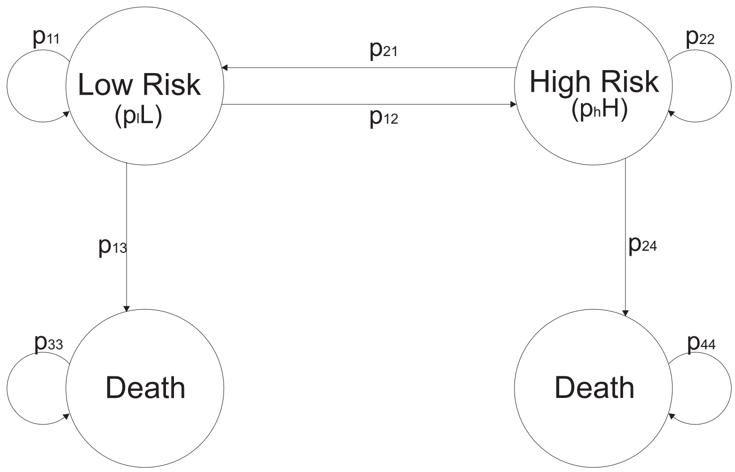
In the PKH model, all individuals have the same low loss probability in a first period. Next, individuals who have suffered a loss are believed (by themselves and all insurers) to have a high loss probability in all subsequent periods, whereas those who have not suffered a loss are believed to stay a low risk. This model implies that the premiums for feasible and efficient GR coverage in a full-information world are represented by a sequence of continually declining premiums. Premiums start out (well) above expected HCE of a low risk and gradually decrease to that level (in the final period).

Following PKH, assume there is a probability p_{12} of an insured turning from a low risk (with expected future HCE of $p_l L$, where p_l is the common loss probability and L the size of the loss) to a high risk (characterized by $p_h \geq p_l$ and/or $H \geq L$, respectively). Accordingly, let p_{11} be the probability that the individual remains in the low-risk status. A lifetime health insurance premium that is compatible with GR is based on a set of transition probabilities,

$$p_{ij} = P(X_{n+1} = j | X_n = i).$$

These transition probabilities give rise to a Markov process (see Figure 1).

Figure 1 Markov process of risk states



As an extension of PKH, high risks are assumed to 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 $p_{13} > 0$.[4] The probabilities are contained in the transition matrix A,

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

PKH deal with a special case of A, namely the case $p_{21} = p_{13} = p_{24} = 0$ and $p_{22} = 1$, while their loss probabilities in the low-risk state are $p_{11} = 1 - p_l$ and $p_{12} = p_l$, respectively. These restrictions will be relaxed, step by step, first, a positive probability of returning to low-risk status, i.e., $p_{21} > 0$ is introduced.

3.1 Positive probability of returning to low risk (no absorbing death)

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

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

4 Herring and Pauly (2006) present empirical evidence suggesting that the probability of death differs significantly between high and low risks.

Including a positive probability of returning to low-risk status changes the optimal premium schedule significantly, since the share of high risks in the population develops in a more moderate way. Assuming an initial distribution comprising only low risks (the row vector $(b) = [1, 0]$ below), the t^{th}-period state probabilities are given by

$$\underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{=b} \cdot \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{(t-1)} \tag{3}$$

In period 3, the probabilities 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 the PKH results, assume their values $p_{11} = 0.9$ and $p_{12} = 0.1$, respectively, moreover, $p_{21} = 0.25$ (thereby fixing p_{22} at 0.75).[5]

Table 1 displays the development of the shares of high and low risks over five periods in the PKH model and in our modification. The overall loss probability is derived by weighting the state-dependent loss probabilities $[(p_l)$ and $(p_h)]$ with the respective share in the population. Intuitively, a positive probability of returning to the favorable risk status should lower long-run premiums, causing the amount of frontloading to be lower. This is illustrated in Table 1 using the same parameter values as PKH. The lifetime premium, calculated for $L = 100$, is $P_{PKH} = 68.098$ and $P_{Mod.} = 64.139$ for $p_{12} = 0.25$.

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 approaches 0.23471 rather than 0.3439. The overall loss probability begins to change in period three already, converging to 0.14694 rather than 0.16878, the lifetime premium drops by 5.8 percent. 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 less or 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 PKH. Therefore, in order to see the effect of $p_{21} > 0$, the Markov process needs to go on for at least three periods. In return, extending to more than five periods would increase complexity without adding further insights.

Expected HCE is given by

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

[5] 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 (1977)). Pauly et al. (1998) assume $D = 4$ in their GR group insurance model, implying $\pi = p_{21} = 0.25$.

Table 1 PKH model (left) and modification (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	0.835	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). Parameter values 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 $L=100$.

where $EHCE$ is expected health care expenditure, $Prob_{low}$ is the probability of being in the low-risk status t periods from now (defined analogously for the high-risk status). Expected HCE is calculated for low and high risks, respectively, as indicated by Equation (4). For instance, the average health care cost pertaining to a low risk is $EHCE_{low}=p_lL$ with p_l denoting the probability of L. Expected cost for a high risk is defined in the same way.

Using (3), one obtains expected HCE for each period,

$$EC_1 = p_lL \tag{5}$$

$$EC_2 = p_{11}p_lL + p_{12}p_hH \tag{6}$$

$$EC_3 = (p_{11}^2 + p_{12}p_{21})p_lL + (p_{11}p_{12} + p_{12}p_{22})p_hH \tag{7}$$

$$EC_4 = (p_{11}^3 + 2p_{11}p_{12}p_{21} + p_{12}p_{21}p_{22})p_lL + (p_{11}^2p_{12} + p_{12}^2p_{21} + p_{12}p_{22}^2 + p_{11}p_{12}p_{22})p_hH \tag{8}$$

$$EC_5 = (p_{11}^4 + 3p_{11}^2p_{12}p_{21} + 2p_{11}p_{12}p_{21}p_{22} + p_{12}^2p_{21}^2 + p_{12}p_{21}p_{22}^2)p_lL + (p_{11}^3p_{12} + p_{11}^2p_{12}p_{22} + 2p_{11}p_{12}^2p_{21} + p_{11}p_{12}p_{22}^2 + 2p_{12}^2p_{21}p_{22} + p_{12}p_{22}^3)p_hH. \tag{9}$$

GR premiums must reflect expected lifetime HCE; e.g. the premium in period five equals expected HCE in the first period ($P_5=EC_1$; see PKH), and similarly for premiums in periods one to four. This yields the following GR premium schedule:[6]

[6] Formulae (5) to (9) yield the values in Table 1. The assumptions are $p_{11}=1-p_l$; $p_{12}=p_l$; $p_{13}=p_{24}=p_{21}=0$; $p_{22}=1$ (PKH; LHS of Table 1) and $p_{11}=1-p_l$; $p_{12}=p_l$; $p_{13}=p_{24}=0$; $p_{21}=1-p_h$; $p_{22}=p_h$, respectively (Modified; RHS of Table 1). The parameter values are $p_l=0.1$, $p_h=0.75$, and $L=H=100$.

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

$$
\begin{aligned}
P_2 =& p_{11}^2 p_l L + p_{12}([p_{11}p_{21} + p_{21}p_{22}] p_l L + [p_{12}p_{21} + p_{22}^2] p_h H) \qquad (11) \\
&+ p_{11}p_{12}(p_{21}p_l L + p_{22}p_h H) + p_{11}^2 p_{12}(p_h H - p_l L)
\end{aligned}
$$

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

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

$$
P_5 = p_l L. \qquad (14)
$$

The GR premium schedule above is a competitive equilibrium (see PKH, Proposition 1). It should be noted that GR premiums depend on weighted averages rather than the difference between expected cost pertaining to high and low risks. This is shown for the third period as follows. Adding $(+p_{12}p_l L - p_{12}p_l L)$ to the RHS of Equation (12) yields

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

Rearranging terms on the RHS, one obtains

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

The first term on the RHS of Equation (16) covers expected HCE of a low risk during the current period. The second term covers the risk of an insured turning into a high risk in period two, the second to last period of the contract. The fact that she or he may become a low risk again in period three 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 three, in full analogy with PKH.

3.2 The steady state probability distribution

In Table 1, only five periods were considered. To find the steady-state probability distribution (assuming ergodicity of the Markov process), one can manipulate Equation (2) to establish a relationship between PKH and the GR group insurance model by Pauly et al. (1998). The steady-state distribution vector v of an ergodic Markov process satisfies

$$
v' \times B = v'. \qquad (17)
$$

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

$$
\begin{aligned}
p_l^* + p_h^* &= 1 && (18)\\
\begin{bmatrix} p_l^* & p_h^* \end{bmatrix} \times \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} &= \begin{bmatrix} p_l^* & p_h^* \end{bmatrix} && (19)
\end{aligned}
$$

Solving for p_l^* and p_h^* yields

$$
\begin{aligned}
p_l^* &= \frac{p_{21}}{1 + p_{21} - p_{11}} = 0.714 && (20)\\
p_h^* &= \frac{p_{12}}{1 + p_{12} - p_{22}} = 0.286. && (21)
\end{aligned}
$$

In the long run, 71.4 percent of the insured belong to the favorable low-risk and 28.6 percent to the unfavorable high-risk category, respectively, representing the insurance company's (basic) low/high health risk profile over the long run. This result is consistent with Pauly et al. (1998) as well as Nickel (2005).

3.3 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 that claims the influence of age on HCE to be 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)]. This suggests an extra cost of dying ($C_h, C_l > 0$) which may have an impact on the GR premium.

In the PKH model, the cost of dying is neglected (see Figure 2 (a)). Once the premium is paid, HCE may accrue. Next, risk status may change before the start of period $t + 1$. Accordingly, premiums must satisfy the following restriction in a two-period model,

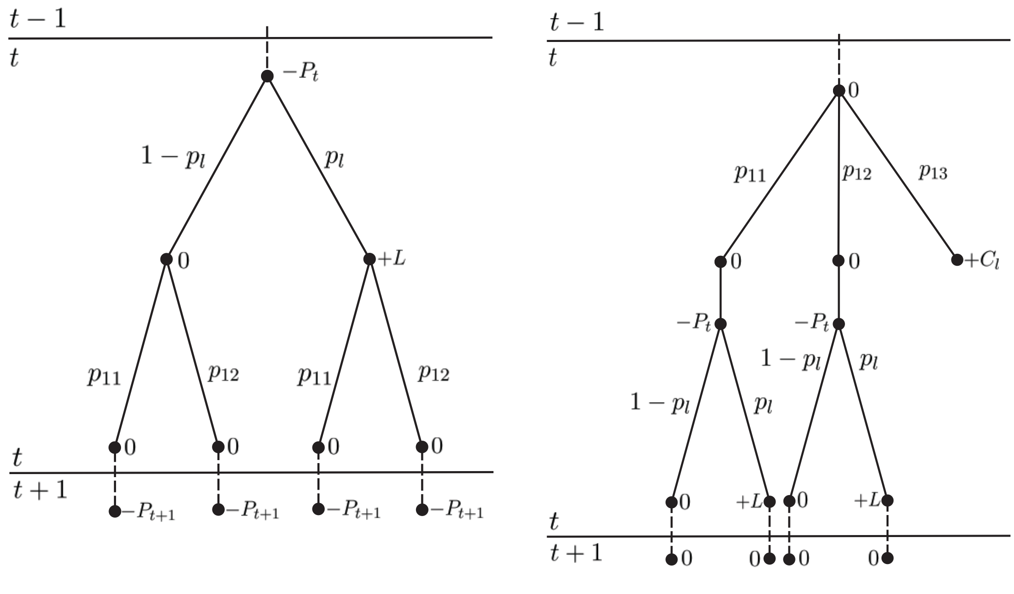
$$P_t + P_{t+1} = p_l L + p_{11} p_l L + p_{12} p_h H. \tag{22}$$

Turning to Figure 3: In the first scenario, $P_{t,c}$ clearly increases compared to $P_{t,a}$. Turning to scenario two and three one can distinguish no less than nine subcases. The upshot is that there are many parameter constellations that cause the GR premium with death to be lower in the first period than calculated by PKH ($P_{t,PKH}$). This in turn implies that critical time preference that is still compatible with GR insurance is lowered (i.e. generating a higher demand for GR insurance contracts as shown in section 4.4).

Now we introduce death as an absorbing state. However, the time of death proves to be very important. In analogy to PKH, the argument always assumes that the individual starts in the low-risk state at time t, while the analysis stops at time $t + 1$.

A one-period contract would require the premium for low risks to be no higher than their expected cost. Therefore P_{t+1} can be set equal to $p_l L$. This in turn implies that the premium in period t is given by

Figure 2 The PKH model and its modification



(a) Timing in the PKH model　　(b) Modified PHK model including death I

$$P_{t,PKH} = p_l L + p_{12}(p_h H - p_l L), \tag{23}$$

where $P_{t,PKH}$ is the premium in period t with the timing illustrated in Figure 2 (a) and $p_{11} = 1-p_{12}$ since $p_{13} = 0$. $P_{t,PKH}$ is the premium derived in PKH.

Now death is introduced. For simplicity, high risks cannot become low risks ($p_{21} = 0$). In variant I (see Figure 2 (b)), the change in risk status takes place before the premium is paid. Because death and change in risk status occur at the beginning of the period, only survivors pay the premium in period $t+1$ and only low risks are offered GR insurance in period t. Premiums must now satisfy

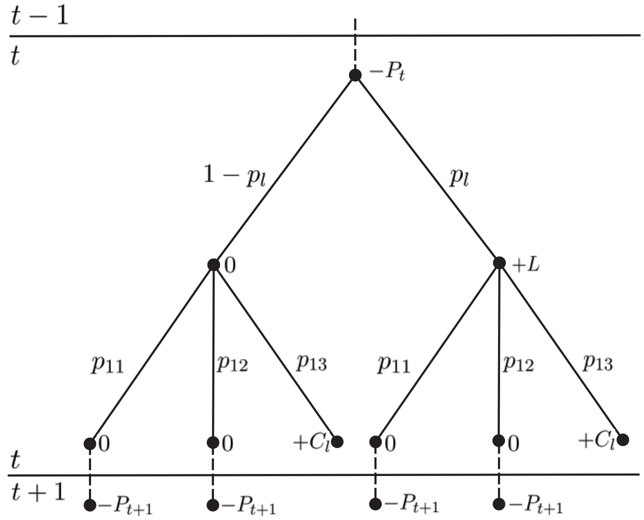
$$P_t + (1-p_{13})P_{t+1} = p_l L + p_{11} p_l L + p_{12} p_h H + p_{13} C_l, \tag{24}$$

where C_l is the extra cost of dying as a low risk. Again substituting $P_{t+1} = p_l L$ and rearranging terms, one has

$$\begin{aligned} P_{t,b} &= p_{11} p_l L + p_{12} p_h H + p_{13}(C_l + p_l L) \\ &= p_l L + p_{12}(p_h H - p_l L) + p_{13} C_l, \end{aligned} \tag{25}$$

since $p_{11} + p_{13} = 1 - p_{12}$. Three possible scenarios emerge from introducing death as a terminal state:

Figure 3 Modified PHK model including death II (c)



- A positive probability of death ($p_{13} > 0$) *reduces* the probability of remaining in the low-risk status but leaves the probability of transition from low- to high-risk status *unchanged* ($p_{11} \downarrow$, $p_{12} \rightarrow$).
- A positive probability of death ($p_{13} > 0$) *reduces* the probability of transition from low- to high-risk status but leaves the probability of remaining in the low-risk status *unchanged* ($p_{11} \rightarrow$, $p_{12} \downarrow$).
- A positive probability of death ($p_{13} > 0$) *reduces* both the probability of transition from low- to high-risk status and the probability of remaining in the low-risk state ($p_{11} \downarrow$, $p_{12} \downarrow$).

However, death could also occur after HCE has been spent on the high risk. This modification leads to variant II (see Figure 3 or scenario (c)). Again, only survivors pay the premium in period $t+1$. However, everyone is offered a contract in period t since the change in risk status occurs at the end of the period after the premium is paid. This implies the following set of restrictions on the premium,

$$P_t + (1 - p_{13})P_{t+1} = p_l L + p_{13} C_l + p_{11} p_l L + p_{12} p_h H + p_{11} p_{13} C_l + p_{12} p_{24} C_h, \tag{26}$$

where C_h is the extra cost of dying as a high risk. However, P_{t+1} must also cover the extra cost of dying as a low risk, calling for the restriction,

$$P_{t+1} = p_l L + p_{13} C_l. \tag{27}$$

Substituting, one obtains for the GR premium

$$\begin{aligned} P_{t,II} &= p_{11} p_l L + p_{12} p_h H + p_{13}(p_l L + p_{13} C_l) + p_{11} p_{13} C_l + p_{12} p_{24} C_h \\ &= p_l L + p_{12}(p_h H - p_l L) + p_{13} C_l + p_{12}(p_{24} C_h - p_{13} C_l). \end{aligned} \tag{28}$$

In full analogy to Figure 2 (b), three scenarios can be distinguished. Note that premiums $P_{t,I}$ and $P_{t,II}$ are both lower than or equal to $P_{t,PKH}$ as long as zero excess cost of dying is assumed ($C_l = C_h = 0$). The first two terms on the RHS of Equation (29) serve to cover expected health loss, with the first term referring to a low risk during period t and the second, to the possible increase in expected cost weighted by the probability of becoming a high risk in period $t+1$. The first two terms of Equation (29) correspond to Equation (23), i.e. the PKH model. With $C_l > 0$ and $C_h > 0$, however, there are two additional terms reflecting death as a terminal state. The first is for the expected extra cost of dying in t 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 t.

4 GR insurance and consumer capital market constraints

Consider a U.S. worker who purchases private health insurance coverage at the age of 20 and keeps it until the age of 65. Even for such a long contract life, prepayment to ensure GR can constitute a heavy financial burden, making GR outright unaffordable for young individuals in the presence of capital market imperfections. Frick (1998) shows that imperfect capital markets in combination with high subjective rates of time preference result in failure to purchase a GR contract. Indeed, a critical value of time preference exists such that consumers do without GR.

4.1 Frick (1998) revisited

In this section, Frick's argument is revisited using the different chronologies developed in Figures 2 and 3 and the corresponding premium structures. Following Frick (1998), it is sufficient to establish a two-period model.[7] The optimization problem then reads (with y denoting constant per period income to finance premiums and with $\beta \leq 1$ being the subjective discount rate):

$$\begin{aligned} \max_{P_1, P_{2H}} \quad & U(y - P_1) + \beta[p_{12}U(y - P_{2H}) + p_{11}U(y - P_{2L})] \\ \text{s.t.} \quad & P_1 + p_{12}P_{2H} + p_{11}P_{2L} = p_l L + p_{11}p_l L + p_{12}p_h H \\ & P_{2L} = p_l L \end{aligned} \tag{29}$$

Here, P_1 is the first-period premium paid by everyone as a low risk, while the second-period premium differs between risks. The second-period premium paid by those who become high risks P_{2H} differs from that paid by low risks P_{2L}. The Lagrangian and its first-order conditions yield

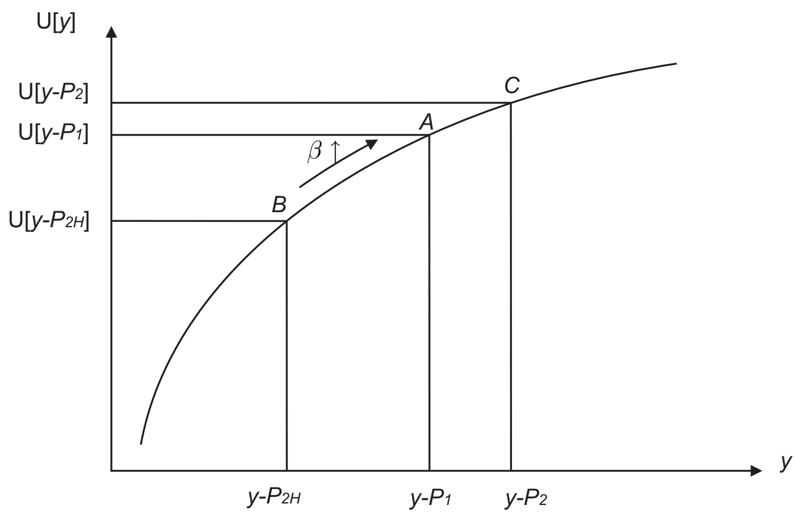
$$U'[y - P_1] = \beta U'[y - P_{2H}] \tag{30}$$

$$p_l L + p_{12}p_h L = P_1 + p_{12}P_{2H}. \tag{31}$$

[7] As insurers can initiate new GR sequences at the beginning of any new period, longer time frames are variations of the two-period model.

Note that consumers can attain these conditions only given the assumption that a new insurer may enter the market in any period. An optimization over just two periods is therefore sufficient. It is revealed that the higher the time preference β of individuals, the higher is the premium P_{2H} that individuals are willing to accept. Figure 4 illustrates.

Figure 4 Consumer's risk utility function



Let the risk utility function of a consumer who has paid P_1 and hence has first-period disposable income $(y - P_1)$ pass through point A. Given that the consumer remains in the contract and pays premium P_{2H} from his second period income, let the risk utility function pass through point B, with higher marginal utility. Now with $\beta = 1$ (no time preference), points A and B would have to coincide to make marginal utilities equal in both periods. However, if marginal utility $U'[y - P_{2H}]$ is multiplied by $\beta < 1$ (indicating positive time preference), point A must lie to the right of point B, indicating that a second-period premium $P_{2H} > P_1$ is acceptable in optimization. As one can glean from Equation (31), P_{2H} only enters the budget constraint of individuals if $p_{12} > 0$, i.e., a positive probability of becoming an unfavorable risk. Therefore, Equations (30) and (31) together imply that the insurers have some scope in skimming off willingness to pay by consumers who have both marked time-preference and a positive probability of becoming a high risk.

The formula derived by Frick (1998) can be generalized to include additional probabilities and manipulated to derive long-term premium paths, permitting to calculate it for any contract length. However, the major contribution by Frick (1998) is to prove the existence of a critical value of time preference $\bar{\beta}$ separating individuals who buy GR insurance from those who do not. Based on logarithmic utility and the actuarially fair premium per period, one obtains these threshold values of $\bar{\beta}$ for all three chronologies in Figures 2 and 3:

$$\bar{\beta}^a = \frac{y - p_h H}{y - p_l L} \qquad \text{(as in PKH);} \tag{32a}$$

$$\bar{\beta}^b = \frac{y - p_h H}{y - p_l L - p_{13} C_l}; \tag{32b}$$

$$\bar{\beta}^c = \frac{y - p_h H - p_{24} C_h}{y - p_l L - p_{13} C_l}. \tag{32c}$$

Compared to PKH [Equation (32a)], chronology (b) results in a higher rather than lower critical value of time preference. A lower value of $\bar{\beta}$ is only possible in chronology (c). However, the outcome depends on the relative magnitude of $p_{24}C_h$ and $p_{13}C_l$. As pointed out in section 3, Herring and Pauly (2006) estimate the probability of death to be significantly larger for high risks. The cost of death with regard to different risks has never been studied econometrically and would be an interesting topic for future research.

4.2 A more general approach

Alternatively, the optimization problem can retain the second-period community-rated premium. A consumer considering a future change in risk status now optimizes

$$\begin{aligned} \max_{P_1, P_2} \quad & U(y - P_1) + \beta(1 - p_{13})U(y - P_2) \\ \text{s.t.} \quad & P_1 + (1 - p_{13})P_2 = p_l L + p_{11} p_l L + p_{12} p_h H \\ & P_2 \leq p_l L \end{aligned} \tag{33}$$

where exemplary the constraint in form of Equation (22) is used. The Lagrangian and its first-order conditions yield

$$U'[y - P_1] = \beta U'[y - P_2] + \frac{\mu}{(1 - p_{13})} \tag{34}$$

$$P_1 + (1 - p_{13})P_2 = p_l L + p_{11} p_l L + p_{12} p_h H \tag{35}$$

Two cases can be distinguished:

- $P_2 = p_l L \;\rightarrow\; \mu \geq 0$. This would imply $U'[y - P_1] \geq \beta U'[y - P_2]$.
- $P_2 < p_l L \;\rightarrow\; \mu = 0$. This would imply $U'[y - P_1] = \beta U'[y - P_2]$.

In the first case, the constrained second-period premium prohibits the financial burden on consumers from increasing later (see Figure 4 again, with $\beta = 1$, the optimal value for consumers could even be a point like C, to the right of point A). The second case is similar to the result derived by Frick (1998), where consumers want to equate first- and second-period marginal utilities. This means that premium P_{t+1} exceeds P_t and would be the same again only in the special case of $\beta = 1$ (A and C coincide in Figure 4).

4.3 Positive rate of interest and increasing health care cost

Up to this point, the main question to be answered has been whether GR contracts are likely to be sustainable in the presence of death. While the answer tends to be positive with regard to the timing presented in Figure 3, there are additional issues confronting policy makers. On the one hand there is the unrelenting increase in the cost of health care which poses a challenge to any health insurer seeking to write GR contracts. On the other hand increases in the cost of health care relax the second-period premium constraint. In addition, financing the upfront GR surcharge would be more affordable if consumers' payments were credited with a positive rate of interest. With $g > 0$ denoting the rate of cost increase and $i > 0$ the rate of interest paid on the GR frontloading, the budget constraints [Equations (22), (24), and (26)] become

$$(1+i)P_t + P_{t+1} = (1+i)p_l L + p_{11}p_l L(1+g) + p_{12}p_h H(1+g) \tag{36a}$$

$$(1+i)P_t + (1-p_{13})P_{t+1} = (1+i)p_l L + p_{11}p_l L(1+g) + p_{12}p_h H(1+g) + p_{13}C_l(1+g) \tag{36b}$$

$$(1+i)P_t + (1-p_{13})P_{t+1} = (1+i)p_l L + (1+i)p_{13}C_l + p_{11}p_l L(1+g) + p_{12}p_h H(1+g) + p_{11}p_{13}C_l(1+g) + p_{12}p_{24}C_h(1+g) \tag{36c}$$

The time of death and its impact on the critical rate of time preference $\bar{\beta}$ is analyzed in the next section. Generally, a positive probability of death continues to lower $\bar{\beta}$, thus causing demand for GR insurance contracts to increase.

4.4 Determining the critical value of time preference

The optimization problem composed of Equation (33) and restrictions (36a)-(36c) are now examined and a numerical example will be used to illustrate the influence of time of death, continuing and comparing the three chronologies introduced in section 3.3.

To keep the results comparable to Pauly et al. (1995) and Frick (1998), Table 2 first contains the same parameter values, with the probability of falling ill (p_l) set equal to the probability of transition from low- to high-risk status (p_{12}) and the probability of remaining in high-risk status set equal to the probability of positive health care cost in the high-risk state ($p_h = p_{22}$).

The calculation proceeds as follows. Suppose $\beta = 1$. Unconstrained individuals equate marginal utilities across periods, as stated below Equation (35). In order to check whether being unconstrained is possible (with $\beta = 1$), the maximum premium that a low-risk status consumer is willing to pay in $t+1$ is substituted into the restrictions ensuring that premiums cover expected cost. Inserting the values given in Table 2 into Equation (36a), one obtains[8],

[8] 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 ($\beta = 1$) are restricted in this optimization.

Table 2 Parameter values assumed*

Chronology (a), no death (PKH)							Chronologies (b) and (c), with death					
p_{11}	=	0.9	p_h	=	0.25		p_{11}	=	0.9	L	=	20
p_{12}	=	0.1	L	=	20		p_{12}	=	0.07	H	=	20
p_{13}	=	0	H	=	20		p_{13}	=	0.03	Y	=	100
p_{21}	=	0.75	Y	=	100		p_{21}	=	0.65	i	=	0.1
p_{22}	=	0.25	i	=	0.1		p_{22}	=	0.25	g	=	0.1
p_{24}	=	0	g	=	0.1		p_{24}	=	0.1	C_l	=	5
p_l	=	0.1					p_l	=	0.1	C_h	=	2
							p_h	=	0.25			

* Following Frick, we also assume logarithmic utility.

$$\begin{aligned}(1+i)P_{t,a} + P_{t+1} &= (1+i)p_l L + p_{11} p_l L(1+g) + p_{12} p_h H(1+g) \\ P_{t,a} &= 2.3.\end{aligned}$$

Every consumer is constrained in chronology (a), because $P_{t,a} > P_{t+1} = p_l L(1+g) = 2.2$.

The lowest possible value of time preference compatible with unconstrained optimization can be derived by dividing first-period by second-period marginal utility, with both premiums set to their actuarially fair per-period values (the fair per-period premium in $t+1$ is the low risks' expected cost and any remaining costs are paid in period t),

$$\frac{U'[Y - P_{t,fair}]}{U'[Y - P_{t+1,fair}]} = \bar{\beta}. \tag{37}$$

The RHS of Table 2 displays the parameter values used for the two chronologies (b) and (c) that deal with death. Assuming one quarter high-risk individuals (approximately the steady-state distribution of section 6.3.1), average death probability in this population is $p_{Death} = 0.047$.[9] For chronology (b) they are plugged into Equation (36b),

$$\begin{aligned}(1+i)P_{t,b} + (1-p_{13})P_{t+1} &= (1+i)p_l L + p_{11} p_l L(1+g) + p_{12} p_h H(1+g) + \\ &\quad p_{13} C_l (1+g) \\ P_{t,b} &= 2.236.\end{aligned}$$

This value again exceeds the maximum premium a low-risk consumer is willing to pay in the second period, viz. 2.2 as before. Turning to chronology (c) one has

[9] Life Expectancy at birth for Swiss men is 77.2 years with average probability of death at age 77 being 0.049 [see Federal Statistical Office (BFS) (2007)]. Since GR may be used as a lifetime contract this end of life setting is assumed to model the last two periods.

$$\begin{aligned}(1+i)P_{t,c}+(1-p_{13})P_{t+1} &= (1+i)p_lL+(1+i)p_{13}C_l+p_{11}p_lL(1+g)+p_{12}p_hH(1+g)+\\ &\quad p_{11}p_{13}C_l(1+g)+p_{12}p_{24}C_h(1+g)\\ P_{t,c} &= 2.363.\end{aligned}$$

This falls short of the maximum acceptable value of $P_{t+1}=2.365$. Table 3 summarizes the effect of each parameter on both premiums [in chronology (c)] and states whether an increase in the respective parameter helps to achieve $P_{t+1}\geq P_t$.

Table 3 Derivatives of both premiums and their effects

Parameter	$\frac{\partial P_t}{\partial \text{Parameter}}$	$\frac{\partial P_{t+1}}{\partial \text{Parameter}}$	Effect
i	$-Rp_{12}((p_hH-p_lL)+(p_{24}C_h-p_{13}C_l))/(1+i)$	0	✓
g	$p_{12}((p_hH-p_lL)+(p_{24}C_h-p_{13}C_l))/(1+i)$	$p_lL+p_{13}C_l$	(✓)
C_l	$p_{13}(1-p_{12}R)$	$p_{13}(1+g)$	✓
C_h	$p_{12}p_{24}R$	0	–
p_lL	$1-p_{12}R$	$1+g$	✓
p_hH	$p_{12}R$	0	–
p_{24}	$p_{12}C_hR$	0	–

Note: (✓) = (contingent on parameter values) increasing this parameter relaxes the constraint; – = increasing this parameter tightens the constraint; $R=(1+g)/(1+i)$.

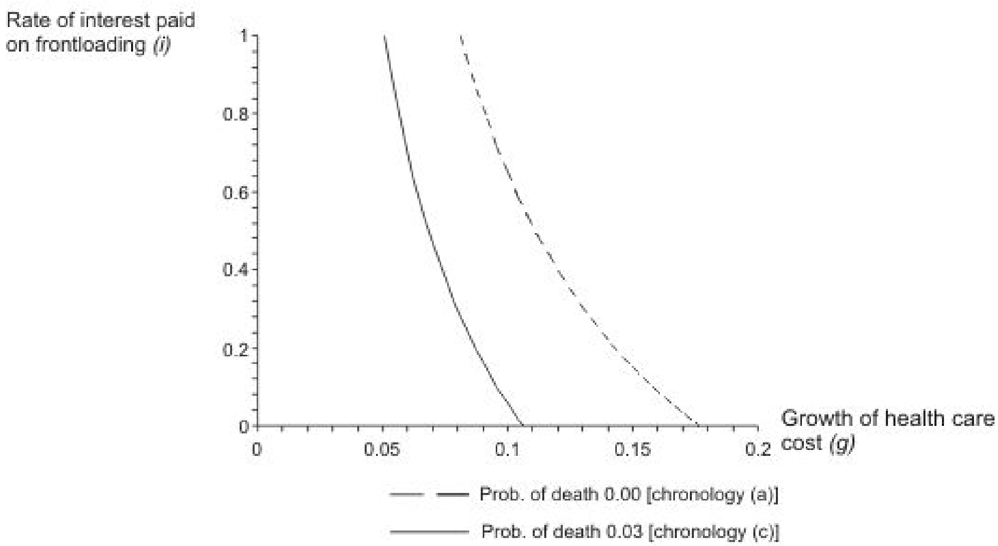
On the one hand this is true whenever P_t is increased less than P_{t+1}. On the other hand this occurs when P_t is decreased more than P_{t+1}. Increasing the rate of interest paid on prepayments, expected health care costs, or the cost of dying as a low risk all relax the second-period premium constraint. Increasing the rate of growth in health care costs just does so if $p_{12}(p_hH+p_{24}C_h)<(1+i+p_{12})(p_lL+p_{13}C_l)$. However, every parameter associated with the high-risk state aggravates the second-period premium constraint. The outcome of changing a transition probability in the low-risk state crucially depends on the remaining parameters and the effect on its composite probabilities.

More generally, there is a critical combination of cost growth (g) and interest rate (i) that can achieve uniform premiums over two periods.[10] The locus of critical values in Figure 5 is derived by assuming that the second- and the first-period premium are the same. Figure 5 first displays the PKH case (zero probability of death). Any combination of (g,i) lying to the origin from this locus cannot achieve uniform premiums. This is only possible whenever growth in the cost of health care exceeds 17.8 percent. If the insurer earns positive interest on the GR frontloading, this critical value drops close to 10 percent. However, the steep slope of the locus makes clear that the easying element is cost growth (g) since it directly affects the restriction on the second-period premium. The second locus of Figure 5 corresponds to chronology (c) (positive probability of death). With the parameter values given, (g) may be below 11 percent p.a. and uniform premiums are possible. Increasing the rate of interest in this case lowers the respective growth rate to almost 5 percent.

Another locus of interest is the net gain in expected utility from a GR contract as a function of time preference β. The expected utility gained from GR insurance less the present value of the associated

[10] Any individual with $\beta=1$ would prefer uniform premiums, while those with $\beta<1$ would prefer $P_t<P_{t+1}$. If uniform premiums are not viable because $P_{t+1}<P_t$ always holds every individual will always be constrained.

Figure 5 Combination of i and g that are compatible with GR



penalty through the restricted second-period premium with parameters again taken from the RHS of Table 2 is displayed in Figure 6 [deriving Figure 6 is illustrated in **the Appendix** using the chronology from Figure 3 or scenario c)]. As could be expected, consumers with a high level of time preference ($\beta < 0.526$) would rather have no insurance than a GR contract since the utility reduction outweighs the benefit from insurance. However, there is a group of (rather patient) consumers who benefit from a GR contract. On the other hand, imposing a GR mandate probably would result in an overall efficiency loss.

Figure 6 Net expected utility of GR insurance as a function of time preference

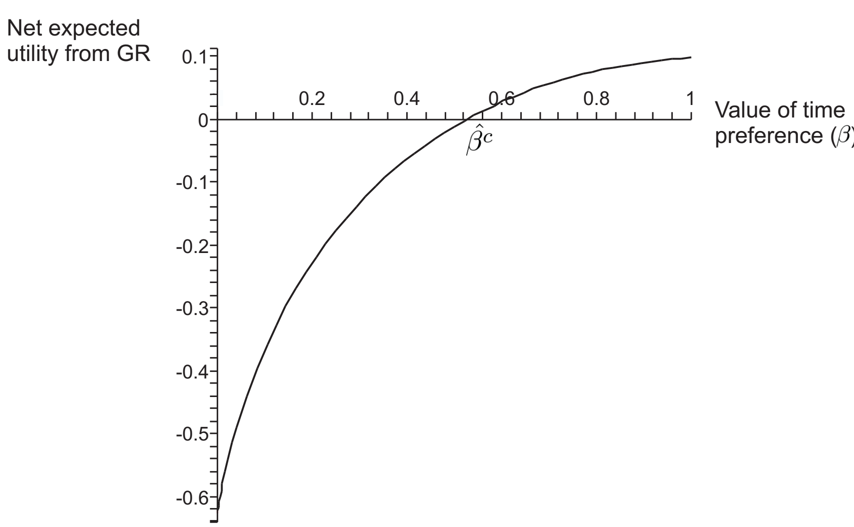
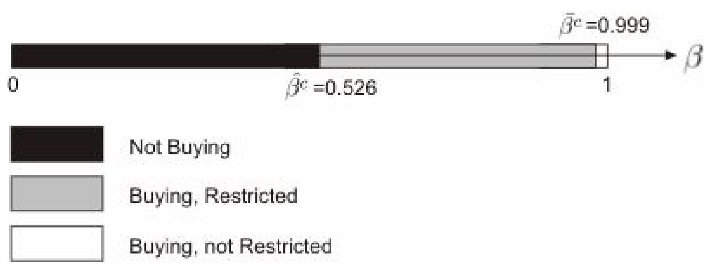


Figure 6 derives the lower critical value ($\hat{\beta}^c$) for chronology (c) and is drawn for consumers with Lagrange multiplier $\mu \geq 0$ in the optimization problem [Equation (33)]. Therefore, the constraint on their second-period premium is binding. If this constraint is relaxed, one obtains a second, upper critical value ($\bar{\beta}^c$) for chronology (c). This value may be compared with the value that would result from Equation (32c). The critical value in the model by Frick (1998) is lower because the community-rating of P_{t+1} is retained here.

Figure 7 displays the critical values of the 'more general approach' that are based on the same calibrations as Figures 5 and 6. One can distinguish the following three types.

- Individuals with strong preference for current consumption $\beta < 0.526$ will not buy GR insurance because the GR surcharge contained in their first-period premium exceeds their expected utility gain from a GR contract.
- Individuals with a moderate to low preference for current consumption $0.526 \leq \beta < 0.999$ do buy GR insurance but would prefer shifting more of the premium burden to the later period.
- Very patient individuals $0.999 \leq \beta \leq 1$ also opt for the GR contract. They would not even have wanted to shift more of the premium burden to the later period.

Figure 7 Critical values of time preference (variant III)



This third set would be empty if death had not been considered (as in PKH and depicted in Figure 5 where the point $i = g = 0.1$ is to the left of the locus without death) because this serves to shift critical $\bar{\beta}$ values up towards 1. Therefore, GR not only survives but may actually thrive in the presence of death.

5 Conclusion

This article extends the work by Pauly, Kunreuther, and Hirth (1995, PKH) on guaranteed renewability (GR). PKH implicitly assume that there is only one possible transition, which moreover is permanent, that from low-risk status to high-risk status. This neglects two important possibilities: First, some insureds do return to low-risk status, as has been documented by Beck et al. (2010) for an observation period of five years and taken into account by Pauly et al. (1998) in their calibration of GR contracts in U.S. group health insurance. Second, there is always the transition to death associated with a high cost of dying. Finally, the upfront loading contained in the GR premium earns interest while healthcare expenditure (HCE) increases over time. Therefore, the question arises of whether an individual GR insurance contract can "survive" these modifications.

One way to answer this question is to determine the value of consumers' time preference that is compatible with GR. Consumers who discount the future heavily do not want to pay the GR frontloading because they are strongly interested in current consumption, while those who are patient opt for GR. If the critical value distinguishing these two types shifts toward the impatient type by the modifications considered, the set of consumers opting for GR shrinks, possibly to an empty set. If, however, the critical value distinguishing these two types shifts toward the patient type, the modifications considered serve to enhance GR in health insurance. Consumers may face even harsher credit rationing than assumed here when having to come up with the GR surcharge early in their life cycle. Nevertheless, this research permits to conclude that GR constitutes a possible solution to the problem of insuring the risk of deteriorating health status without any particular mandate by public policy.

However, this test turns out not to be sufficient in all cases. Notably, including the probability of transition from high- to low-risk status does not affect this critical value of time preference while still serving to reduce lifetime premiums by 5.8 percent, thus making GR more attractive in the example given in section 3.2. When death is accounted for as a terminal state, the critical value of time preference is not affected in all cases either. The outcome strongly depends on the time of death and its respective cost. This is illustrated for the model by Frick (1998) and the alternative approach. Thus, GR is found to survive well in the presence of death. This is also true when the cost of health care increases and interest is being paid on the frontloading provided by consumers. However, the rate of interest turns out to be far less important. Of course, these statements are conditional on the parameter values entering simulations. Analyzing the effect of the remaining parameter values using the more general approach reveals that results are most sensible to changes in low risks' cost of health care, their probability and their cost of death. Future research should therefore also address the empirical question of the ratio between different risks' cost of dying. In addition, the consequence of using alternative utility functions should also be studied.

References

Beck, K., Trottmann, M., and Zweifel, P. (2010). Risk adjustment in health insurance and its long-term effectiveness. *Journal of Health Economics*, 29(4):489–498.

Bhattacharyya, G. K. (1977). *Statistical concepts and methods, Gouri K. Bhattacharyya, Richard A. Johnson*. Wiley. New York. US.

Brown, H. and Connelly, L. B. (2005). Lifetime cover in private insurance markets. *International Journal of Health Care Finance and Economics*, 5(1):75–88.

Cochrane, J. H. (1995). Time-consistent health insurance. *Journal of Political Economy*, 103(3):445–473.

Federal Statistical Office (BFS) (2007). *Mortality tables of Switzerland 1998/2003*. Federal Statistical Office (BFS), Neuchatel.

Frick, K. (1998). Consumer capital market constraints and guaranteed renewable insurance. *Journal of Risk and Uncertainty*, 16(3):271–278.

Herring, B. and Pauly, M. V. (2001). Premium variation in the individual health insurance market. *International Journal of Health Care Finance and Economics*, 1(1):43–58.

Herring, B. and Pauly, M. V. (2006). Incentive-compatible guaranteed renewable health insurance premiums. *Journal of Health Economics*, 25(3):395–417.

Hofmann, A. and Browne, M. (2013). One-sided commitment in dynamic insurance contracts: Evidence from private health insurance in Germany. *Journal of Risk and Uncertainty*, 46(1):81–112.

Lubitz, J. D. and Riley, G. F. (1993). Trends in medicare payments in the last year of life. *New England journal of medicine*, 328(15):1092–1096.

Nickel, A. (2005). Versicherung der veränderung der risikosituation in der pkv - (insuring a change in risk in german private health insurance. *Zeitschrift für Versicherungswesen - Journal of Insurance Business*, 16:505–509.

Patel, V. and Pauly, M. V. (2002). Guaranteed renewability and the problem of risk variation in individual health insurance markets. *HEALTH AFFAIRS-MILLWOOD VA THEN BETHESDA MA-*, 22:W280–W289.

Pauly, M., Nickel, A., and Kunreuther, H. (1998). Guaranteed renewability with group insurance. *Journal of Risk and Uncertainty*, 16(3):211–221.

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

Shelton Brown, H. and Connelly, L. B. (2005). Market failure in long-term private health insurance markets: a proposed solution. *Applied Economics Letters*, 12(5):281–284.

Steinmann, L., Telser, H., and Zweifel, P. S. (2007). Aging and future healthcare expenditure: a consistent approach. *Forum for Health Economics & Policy*, 10(2).

Werblow, A., Felder, S., and Zweifel, P. (2007). Population ageing and health care expenditure: a school of "red herrings"? *Health economics*, 16(10):1109–1126.

Zweifel, P., Felder, S., and Meier, M. (1999). Ageing of population and health care expenditure: A red herring? *Health Economics*, 8(6):485–496.

Appendix

The calculation of net gain in expected utility due to GR health insurance is considered here. As in Frick (1998), a logarithmic risk utility function is assumed. For the utility gained from insurance ($\triangle EU$), in the first period one is left with,

$$\begin{aligned} \triangle EU_1 &= (1-p_{13})\{\ln[p_l(y-L)+(1-p_l)y] - p_l\ln[y-L]-(1-p_l)\ln[y]\} \\ &= 0.0020. \end{aligned} \tag{C.1}$$

With cost growth g and interest i, one obtains for the second period,

$$\begin{aligned} \triangle EU_2 &= \{(1-p_{13})p_{11}+(1-p_{24})p_{12}\} \\ & \quad \{\ln[(p_{11}p_l+p_{12}p_h)(y-L(1+g))+(p_{11}(1-p_l)+p_{12}(1-p_h))y]\} \\ & \quad -\{(1-p_{13})p_{11}+(1-p_{24})p_{12}\}\{(p_{11}p_l+p_{12}p_h)\ln[y-L(1+g)]\} \\ & \quad -\{(1-p_{13})p_{11}+(1-p_{24})p_{12}\}\{(p_{11}(1-p_l)+p_{12}(1-p_h))\ln[y]\} \\ &= 0.1027 \end{aligned} \tag{C.2}$$

EU_2 is much larger than EU_1 because every insured is a low risk in the first period, whereas the probability of loss is much higher in the next period. In total, the utility premium obtained is,

$$\begin{aligned} \triangle EU &= \triangle EU_1 + \beta(1-p_{13}) \triangle EU_2 \\ &= 0.0020 + 0.0996\beta > 0 \end{aligned} \tag{C.3}$$

The loss in expected utility caused by the binding restriction $P_2 = p_l L$ can be calculated by determining the optimal first-period premium P_1^* when the Lagrange-multiplier is $\mu = 0$ in the optimization problem in Equation (33). 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,

$$P_1^* = \frac{106.7\beta - 92.1211}{1.1+1.067\beta}; \tag{C.4}$$

$$\triangle EU_1^* = \ln\left[100 - \frac{106.7\beta - 92.1211}{1.1+1.067\beta}\right] - 4.5797. \tag{C.5}$$

However, in period 2 the binding restriction causes a reduction in expected utility ($\triangle EU_2^* < 0$) compared to the optimized value of P_2^*. The reduction is given by,

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

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

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

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

The difference between (C.3) and (C.8) is the net gain in expected utility from having GR insurance.

$$\triangle_n\ EU = \triangle\ EU - \triangle\ EU^* \tag{C.9}$$

With the parameter values taken from Table 2 the critical value for $\triangle_n\ EU = 0$ of $\hat{\beta}$ is 0.526. Figure 6 illustrates the difference as a function of β.