Lapse-Based Insurance

Daniel Gottlieb and Kent Smetters*

June 6, 2016

Abstract

Most individual life insurance policies lapse before expiration. Insurers sell front-loaded policies, make money on lapsers, and lose money on non-lapsers. We propose and test a simple model where consumers do not fully take into account the likelihood of needing money during the future policy period. Policy data from two major life insurers support the comparative statics of our model but do not support competing theories, including reclassification risk, hyperbolic discounting, or administrative costs. We also conducted a survey with recent customers of a large national insurer that directly supports our mechanism.

JEL No. D03, G22, G02

*Olin Business School, Washington University in St. Louis and Wharton School, University of Pennsylvania. Daniel Gottlieb: dgottlieb@wustl.edu. Kent Smetters: smetters@wharton.upenn.edu. This paper was previously titled “Narrow Framing and Life Insurance.” We thank Nicholas Barberis, Daniel Bauer, Roland Bénabou, Pedro Bordalo, Sylvain Chassang, Keith Crocker, Kfir Eliaz, Erik Eyster, Hanming Fang, Xavier Gabaix, Nicola Gennaioli, Michael Grubb, Paul Heidhues, Botond Kőszegi, Lee Lockwood, Ted O’Donoghue, Matthew Rabin, Andrei Shleifer, Paul Siegert, Justin Sydnor, Jeremy Tobacman, Jean Tirole, Daniel Sacks, Georg Weizsäcker, Richard Zechara, and seminar participants at the Central European University, ESMT, the European Behavioral Economics Meeting, Federal Reserve Board/George Washington University, NBER Insurance, NBER Household Finance, Penn State, Princeton, the Risk Theory Society, Washington University in St. Louis, University of Wisconsin-Madison, and the University of Pennsylvania for comments. We also thank James Finucane for outstanding research assistance.
## Contents

1 Introduction .................................................. 1

2 Key Stylized Facts ........................................... 4
   2.1 Lapsing is the Norm .................................... 5
   2.2 Lapse-Supported Pricing ............................... 6
   2.3 Front Loading ........................................... 9

3 The Model ....................................................... 9
   3.1 Timing ..................................................... 9
   3.2 Consumer Utility ........................................ 11
   3.3 Firm Profits .............................................. 12
   3.4 Equilibrium .............................................. 12
   3.5 Testing the Comparative Statics of the Model ........ 16
   3.6 Testing the Mechanism ................................ 19
   3.7 Inefficiency and the Effect of Secondary Markets ..... 20
   3.8 Heterogeneous Shocks .................................. 21

4 Other Potential Explanations ................................. 23
   4.1 Reclassification Risk .................................... 24
   4.2 Time Inconsistency ...................................... 25
   4.3 Fixed Costs ............................................... 26

5 Conclusion ..................................................... 28

References ....................................................... 28

Appendix: Survey Results ....................................... 32

Online Appendices ................................................. A-1
   A. Evidence on Lapse-Based Pricing ....................... A-1
   B. Extensions ................................................ A-2
   C. Description of MetLife and SBLI Data ................. A-5
   D. TIAA-CREF Survey Questions ........................... A-6
   E. Opposition to Secondary Markets ....................... A-8
   F. Competing Models ....................................... A-12
   G. Health Transition Probabilities by Age ............... A-30
   H. Allowing for a Mix of Rational Consumers ............. A-31
   I. Proofs ...................................................... A-33
“I don’t have to be an insurance salesman!” – Tom Brady, NFL quarterback, describing the relief that he felt after finally being selected in Round 6, pick No. 199, of the 2000 NFL draft.

1 Introduction

Life insurance is both a large industry and the most valuable method for individuals to financially protect their loved ones upon death. Over 70 percent of U.S. families own life insurance (LIMRA 2014). About $30.8 trillion in individual life insurance coverage was issued between 1990 and 2010 (ACLI 2015, Table 7.8). In 2011, households paid over $101 billion in premiums for life insurance policies in the individual market. The average size of an individual life insurance policy stands at $267,300, roughly four times the median net worth of a household (LIMRA 2010 and U.S. Census Bureau 2011).

Virtually all life insurance policies are front loaded, as policyholders pay more than the actuarial cost of their contemporaneous mortality risk early into the policy in exchange for paying less than their actuarial cost later on. The majority of individual policies, however, never reach their maximum term or pay a death benefit. Instead, policyholders voluntarily terminate them, thereby losing their front load. Specifically, most term policies, which offer coverage for a fixed number of years, lapse prior to the end of the term, as about one in every 14 customers stop paying premiums each year. Similarly, most permanent policies are surrendered (i.e., lapsed and a cash value is paid) before death or their expiration at age 100 or older.

A vast empirical literature, starting as far back as Linton (1932), has documented the relationship between life insurance policy terminations and other variables. But a large puzzle remains. Theoretically, the conventional view is that insurers should use loads to reduce lapses (Hendel and Lizzeri, 2003). Without income shocks, the optimal load will be designed to prevent any lapse, thereby enforcing continued participation in an insurance pool as policyholders learn more about their mortality likelihood over time (“risk reclassification”). With income shocks, some lapses may occur in equilibrium since rational policyholders value the option to lapse after a large shock. Quantitatively, Hambel et al. (2015) simulate

---

1 http://profootballtalk.nbcsports.com/2011/04/13/bradys-perceived-slap-against-insurance-salesmen-makes-waves/
2 Life insurance is sometimes also provided as an employer-based voluntary group benefit. Group policies are generally not portable across employers and, therefore, are priced differently. This paper focuses on individual (non-group) policies. 44% of American households have individual life policies, whereas 49% have group policies. Individual policies tend to be substantially larger than group policies, which have an average coverage of $165,300 (LIMRA, 2010).
3 Front-loaded policies take on many forms, including level premiums, single premiums, limited-pay whole life, and decreasing term insurance policies. Life insurance policies with back loads are essentially non-existent: no related sales information is tracked by any major trade organization, and we could not find a single insurer offering back-loaded policies.
4 With many permanent policies, premiums are often collected only for part of a person’s life. As a result, for the same death benefit, permanent policies are typically much more expensive than term policies. This premium difference adds savings to a policyholders “cash value,” after front loads are deducted. The cash value typically increases for a while and eventually declines as the payment of the death benefit approaches. Upon surrender, the cash value is returned, but the presence of front loading means that the cash value is smaller than the premiums paid after adjustments for the cost of insurance (mortality risk). If the permanent policy is not surrendered, the death benefit is paid upon death or when the policyholder reaches age 100, 105, 110, 120, or 121. In Subsection, we show that our model generates policy loan provisions (i.e., partial lapses), as in most permanent policies. However, most of our formal analysis does not distinguish between term and permanent policies.
life insurance demand in a calibrated rational-expectations lifecycle model with income shocks, health shocks, liquidity constraints, reclassification risk, and industry-average markups. They find lapse rates that are much lower than found in the data. Rather than face a substantial risk of lapsing and losing the front load, rational households facing potential income constraints buy less or even no life insurance. This finding is also consistent with the results in Krebs, Kuhn, and Wright (2015), who model endogenously binding borrowing constraints in the context of life insurance purchases and macroeconomic shocks.

As we show later, life insurance companies earn large profits on clients who terminate their policies, since policies are often terminated before mortality increases sufficiently above the premium paid. But insurers lose money on those who keep their policies. Therefore, insurers do not earn extra-ordinary profits. Rather, policyholders who lapse cross-subsidize those who keep their coverage.

Making a profit from policies that lapse is a taboo topic in the life insurance industry. It is informally discouraged by regulators and commonly referenced in a negative manner in public by insurance firm executives. As one of their main trade groups recently put it, “[t]he life insurance business vigorously seeks to minimize the lapsing of policies” (ACLI 2012: 64). However, as we show herein, competitive pressure not only forces insurers to compete on this margin; life insurers must endogenously adopt front loads to encourage lapses. This result is the opposite of the conventional view that insurers use front loads to reduce lapses.

We propose and test a model of “differential attention.” Consumers face two sources of risk: mortality risk that motivates the purchase of life insurance and a possible “background” shock that produces a subsequent demand for liquidity. Examples of background shocks include unemployment, medical expenses, stock market fluctuations, real estate prices, new consumption opportunities, and the needs of dependents. Consumers in our model correctly account for mortality risk when buying life insurance but fail to sufficiently account for uncorrelated background risks.

Previous work has documented the presence of differential attention in related settings. For health insurance, there is strong evidence that people weigh different contract features unevenly (Abaluck and Gruber, 2011; Ericson and Starc, 2012; Handel and Kosltd, 2015; and Bhargava, Loewenstein, and Sydnor, 2015). More generally, the theory of narrow framing states that when an individual evaluates a risky prospect “she does not fully merge it with her preexisting risk but, rather, thinks about it in isolation, to some extent; in other words, she frames the gamble narrowly” (Barberis, Huang, and Thaler, 2006). It is reasonable, therefore, to consider whether differential attention might also play a significant role in the sizable life insurance market. Indeed, a large empirical literature reviewed later documents the strong effect that income and unemployment shocks have on life insurance lapses.

Since firms and consumers disagree over the likelihood of lapses in our model, they effectively engage in speculation. Of course, speculative trading with different priors is not novel. But we demonstrate that this speculation causes firms to offer insurance contracts that are seemingly cheap over the life

---

^5See Baicker, Mullainathan, and Schwartzstein (2015) for an insurance model where buyers make behavioral mistakes.
^6See Read, Loewenstein, and Rabin (1999) for a survey on narrow framing, and Rabin and Weitzsäcker (2009) for theoretical and empirical results on how narrow framing causes violations of stochastic dominance.
of the contract – that is, if consumers hold onto their policies – in exchange for being front-loaded. Front loading, in turn, reduces the policyholder’s current resources, magnifying the increase in marginal utility if the household suffers a background shock. A front-loaded policy, therefore, encourages the policyholder to lapse after a background shock, increasing the insurer’s profits. Policies produce cross-subsidies from consumers who lapse to those who do not. These policies are offered even if some of the consumers have correct expectations about all shocks. Moreover, no firm can profit from educating biased consumers about their failure to account for background shocks. These policies even survive the presence of paternalistic not-for-profit firms.

We test our model both indirectly and directly. For the indirect test, we show that the general pattern of premiums observed in practice is consistent with the comparative statics of our model but inconsistent with alternative explanations. These competing theories include reclassification risk, either naive or sophisticated time inconsistency, and the presence of fixed costs. For additional robustness of the indirect test, we collect policy data from two national life insurers to test a key prediction from our model that also allows us to directly distinguish it from other potential explanations. The data strongly supports our differential attention model and is generally inconsistent with the competing models.

We also directly test our hypothesis that consumers underestimate the probability of lapsing. We implemented a survey with the universe of customers from TIAA-CREF who purchased life insurance in the previous two years. Along with several other questions, we asked them about their expectations about lapsing and reasons they might lapse. Only 2.8% said that they planned to stop their policy before its expiration. In contrast, based on TIAA-CREF’s actual historical experience with these same type of policies, approximately 60% will likely lapse. Of course, this big mismatch between perceived and likely lapses is also potentially consistent with people being overconfident (or optimistic) about the safety of their future income, a behavior that is prevalent in other markets. So, to disentangle between differential attention and biased beliefs, we asked additional questions about expected future income shocks. Interestingly, survey respondents anticipate a high chance of negative income shocks: 27.2% reported an income loss in the last five years and 25.2% expect an income loss during the next five years. However, their beliefs about income shocks are essentially uncorrelated with beliefs about the chance of lapsing. Therefore, our results tend to favor the differential attention explanation for the life insurance market as opposed to overconfidence/optimism about future income.

In addition to the literature noted above, our paper is related to an emerging literature in behavioral industrial organization, which studies how firms respond to consumer biases. For example, Squintani and Sandroni (2007), Eliaz and Spiegler (2008), and Grubb (2009) study firms who face overconfident consumers, DellaVigna and Malmendier (2004), Eliaz and Spiegler (2006) and Heidhues and Kőszegi (2010) consider consumers who underestimate their degree of time inconsistency, and Eliaz and Spiegler

---

7 In the context of unemployment insurance, Spinnewijn (2015) finds that the unemployed vastly overestimate how quickly they will find work. Grubb (2009) shows that overconfidence accounts for the prevalence of three-part tariffs in cellular phone plans, Malmendier and Tate (2005) show that managerial overconfidence can account for investment distortions, and, in a political economy context, Ortoleva and Snowberg (2015) find that overconfidence can explain ideology and voter turnout. Bénabou and Tirole (2002) study endogenously optimistic beliefs.
When consumers are time-inconsistent, competition can also undermine the effectiveness of commitment devices (Kőszegi, 2005, and Gottlieb, 2008). For surveys of the behavioral industrial organization and behavioral contract theory literatures, see Ellison (2005) and Kőszegi (2014), and Grubb (2015).
2.1 Lapsing is the Norm

The Society of Actuaries and LIMRA, a large trade association representing major life insurers, define an insurance policy lapse as “termination for nonpayment of premium, insufficient cash value or full surrender of a policy, transfer to reduced paid-up or extended term status, and in most cases, terminations for unknown reason” (LIMRA 2011A, P. 7). About 4.2% of all life insurance policies lapse each year, representing about 5.2% of the face value actually insured (“in force”). For term policies, which contractually expire after a fixed number of years if death does not occur, about 6.4% lapse each year. For permanent policies, the lapse rate varies from 3.0% per year (3.7% on a face amount-weighted basis) for traditional whole life policies to 4.6% for universal life policies. So-called variable life and variable universal life types of permanent policies lapse at an even higher rate, equal to around 5.0% per year (LIMRA 2011A). While the majority of policies issued are permanent, the majority of face value now takes the term form (LIMRA 2011A, P. 10; ACLI 2011, P. 64).

These annualized rates lead to substantial lapsing over the multi-year life of the policies. About $30.8 trillion of new individual life insurance coverage was issued in the United States between 1990 and 2010 (ACLI, 2015), and around $24 trillion of in-force coverage was dropped during this same period.

As Figure 1(a) shows, almost 25% of permanent insurance policyholders lapse within just three years of first purchasing the policies; within 10 years, 40% have lapsed. According to Milliam USA (2004), almost 85% of term policies fail to pay a death claim; nearly 88% of universal life policies ultimately do not terminate with a death benefit claim. In fact, 74% of term policies and 76% of universal life policies sold to seniors at age 65 never pay a claim.

Why do people let their life insurance policies lapse? Starting as far back as Linton (1932), a vast insurance literature has established that income and unemployment shocks are key determinants of policy lapses. For example, Liebenberg, Carson, and Dumm (2012) find that households are twice more likely to surrender their policy after a spouse becomes unemployed. Fier and Liebenberg (2012) find that the probability of voluntarily lapsing a policy increases after large negative income shocks, especially for those with higher debt. As Figure 1(b) shows, lapses are more prevalent for smaller policies, which are typically purchased by lower-income households who are more exposed to liquidity shocks. Moreover, younger households are also more likely to experience liquidity shocks and lapse more. As shown in Figure 2 which shows lapse rates for eleven major life insurers in Canada, young policyholders lapse almost three times more often than older policyholders.

---

9 Drops include coverage issued before 1990. In some cases, policies were dropped based on other factors other than failure to pay (lapses), for example, if the insurer believes that the policy terms were not satisfied.
10 While term policies have a larger annual lapse rate, permanent policies are usually more likely to lapse over the actual life of the policy due to their longer duration.
11 Hoyt (1994) and Kim (2005) document the importance of unemployment for surrendering decisions using firm-level data. Jiang (2010) finds that both lapsing and policy loans are more likely after policyholders become unemployed. Using detailed socio-demographic data from Germany, Inderst and Sirak (2014) find that income and unemployment shocks are leading causes of lapses. They also find that the correlation between age and lapses disappears once one controls for income shocks and wealth.
Figure 2: Annual lapse rates per policy year by age; 15-year duration policies. Source: Canadian Institute of Actuaries (2007).

The macroeconomic evidence also broadly supports the role of income and unemployment shocks. Lapse rates spike during times of recessions, high unemployment, and increased poverty. For example, while $600B of coverage was dropped in 1993, almost $1 trillion was dropped in 1994 (a year with record poverty) before returning to around $600B per year through the remainder of the decade. After the 2000 stock market bubble burst, over $1.5 trillion in coverage was forfeited, more than double the previous year (ACLI 2011).

As we describe in detail in Section 3.6, we collected historical and prospective survey data from life insurance policies sold by TIAA-CREF. Their historical data is in line with these industry-wide findings. Lapse rates nearly doubled during the recessions of 2002 and 2009. Moreover, lapses are positively correlated with changes in the unemployment rate and negatively correlated with real GDP growth.

2.2 Lapse-Supported Pricing

Insurers profit from policyholders who lapse and lose money on those who do not lapse. Policyholders over-pay relative to their mortality risk early into the life of the policy in exchange for receiving a discount later on. When a policy is dropped, the amount paid in excess of the actuarially fair price is not fully repaid to consumers. Hence, insurers make money when policies are dropped.

There is substantial anecdotal evidence that insurers take subsequent profits from lapses into account when setting their premiums. For example, in explaining the rise in secondary markets (discussed in Online Appendix E), the National Underwriter Company writes: “Policy lapse arbitrage results because of assumptions made by life insurance companies. Policies were priced lower by insurance companies

\footnote{For studies using aggregate data from the United States, see Dar and Dodds (1989) for Great Britain, and Outreville (1990) and Kuo, Tsai, and Chen (2003).}

\footnote{As noted in the introduction, premiums for permanent insurance are larger than for term, thereby allowing the policyholder to build up some additional “cash value.” Upon surrendering these contracts prior to death, the cash value paid to the policyholder is much smaller in present value than the premiums paid to date in excess of actuarially fair premiums.}
on the assumption that a given number of policies would lapse.” (NUC 2008, p.88)

Dominique LeBel, actuary at Towers Perrin Tillinghast, defines a lapse-supported product as a “product where there would be a material decrease in profitability if, in the pricing calculation, the ultimate lapse rates were set to zero (assuming all other pricing parameters remain the same).” (Society of Actuaries 2006) Precisely measuring the extent to which life insurance policies are lapse-supported is challenging since insurers do not report the underlying numbers. One reason is regulatory: for determining the insurer’s reserve requirements, the historic NAIC “Model Regulation XXX” discouraged reliance on significant income from lapses for those policies surviving a certain threshold of time.\footnote{Most recently, principles-based regulations (PBR) have emerged, which are widely regarded to allow for more consideration of policy lapses for purposes of reserve calculations.}

A second motivation is competitive: insurers are naturally tight-lipped about their pricing strategies.

Nonetheless, various sources confirm the widespread use of lapse-supported pricing. First, like economists, actuaries employed by major insurers give seminars to their peers. The Society of Actuaries 2006 Annual Meetings held a session on lapse-supported pricing that included presentations from actuaries employed by several leading insurance companies and consultants. Kevin Howard, Vice President of Protective Life Insurance Company, for example, demonstrated the impact of lapses on profit margins for a representative male client who bought a level-premium secondary guarantee universal life policy, with the premium set equal to the average amount paid by such males in August 2006 in the company’s sample. Assuming a zero lapse rate, the insurer projected a substantial negative profit margin, equal to -12.8%. However, at a typical four percent lapse rate, the insurer’s projected profit margin was +13.6%, or a 26.4% increase relative to no lapsing.\footnote{For less popular single-premium policies, the swing was lower, from -6.5% to +8.7%.}

Similarly, at the 1998 Society of Actuaries meeting, Mark Mahony, marketing actuary at Transamerica Reinsurance, presented calculations for a large 30-year term insurance policy often sold by the company. The insurer stood to gain $103,000 in present value using historical standard lapse rate patterns over time. But, if there were no lapses, the insurer was projected to lose $942,000 in present value. He noted: “I would highly recommend that in pricing this type of product, you do a lot of sensitivity testing.” (Society of Actuaries 1998, p. 11)

In Canada, life insurance policies are also supported by lapsing.\footnote{See, for example, Canadian Institute of Actuaries (2007).} As A. David Pelletier, Executive Vice President of RGA Life Reinsurance Company, argues:

What companies were doing to get a competitive advantage was taking into account these higher projected future lapses to essentially discount the premiums to arrive at a much more competitive premium initially because of all the profits that would occur later when people lapsed. (Society of Actuaries 1998, P. 12)

In order to evaluate the importance of lapse-supported pricing with a more representative sample, we gathered data from Compulife, a quotation system for insurance brokers that contains policy data for over 100 American life insurance companies. In calculating insurance profits, we used the most recent...
Figure 3: Expected profit by time before consumer drops policy (a) and insurance loads in current dollars under a projected inflation rate of 3% (b). Source: Authors’ calculation for 20-year term policies with $500k coverage.

Society of Actuaries mortality table (2008). These tables, which are based on actual mortality experience of insured pools in order to correct for selection, are used by insurers for regulatory reporting purposes. Our calculations are discussed in more detail in Online Appendix A. The results confirm an enormous reliance on lapse income.

Consider, for example, a standard 20-year term policy with $500,000 in coverage for a 35-year old male in good health (“preferred plus” category). Figure 3 (a) shows the projected actuarial profits for all such policies available in February 2013 in the state of California (56 policies). These life insurers are projected to earn between $177 and $1,486 in present value if the consumer surrenders between the fifth and the tenth years of purchasing insurance. However, they are projected to lose between $304 and $2,464 if the consumer never surrenders.

Incidentally, a third source of evidence for lapse-based pricing comes from bankruptcy proceedings, which often force a public disclosure of pricing strategies in order to determine the fair distribution of remaining assets between permanent life policyholders with cash values and other claimants. For example, the insurer Conseco relied extensively on lapse-based income for their pricing; they also bet that interest rates earned by their reserves would persist throughout their projected period. Prior to filing for bankruptcy, they tried to increase required premiums – in fact, tripling the amounts on many existing customers – in an attempt to effectively reduce the cash values for their universal life policies (and, hence, reduce their liabilities). In bankruptcy court, they rationalized their price spikes based on two large blocks of policies that experienced lower-than-expected lapse rates (InvestmentNews 2011).

---

17 We chose California because it is the state with the largest number of available policies. The coverage level was set to the Compulife software’s default level ($500,000). The extent of lapse-based pricing, however, is extremely robust to different terms, ages, coverage levels, and states.

18 Premiums for universal life permanent policies can be adjusted under conditions outlined in the insurance contract, usually pertaining to changes in mortality projections. However, in this case, the bankruptcy court ruled that the Conseco contract did not include provisions for adjusting prices based on lower interest rates or lapse rates. Conseco, therefore, was forced into bankruptcy.
ceedings have revealed substantial lapse-based pricing in the long-term care insurance market as well (Wall Street Journal 2000); most recently, several large U.S. long-term care insurers dropped their coverage without declaring bankruptcy, citing lower-than-expected lapse rates, which they originally estimated from the life insurance market (InvestmentNews 2012).

2.3 Front Loading

As noted earlier, virtually all term and permanent policies are effectively “front loaded” since the initial premium exceeds the actuarially fair prices implied by the mortality probability at the time of purchase. This wedge between the premium and the actuarially fair price decreases over time, as mortality increases with age. The presence of inflation, in fact, reinforces the front-loading feature since the premium is often constant in nominal terms. Loads, therefore, start high and decrease over time. Figure 3(b) presents the insurance loads for the California policies described previously.

3 The Model

We consider a competitive life insurance market where consumers pay more attention to the mortality risk they are insuring than to other “background” or “liquidity” shocks. To highlight the underlying mechanism in the most transparent way possible, this section focuses on a very simplified model.

There are $N \geq 2$ insurance firms indexed by $j = 1, \ldots, N$ and a continuum of households. Each household consists of one head and at least one heir. Because household heads make all decisions, we refer to them as “the consumers.”

3.1 Timing

There are three periods: 0, 1, and 2. Period 0 is the contracting stage. At that stage, each firm offers an insurance policy and consumers decide which one, if any, to purchase. Consumption occurs in periods 1 and 2.

In period 1, consumers have an initial wealth $W$ and lose $L > 0$ dollars with probability $l \in (0, 1)$. Firms do not observe income losses. In period 2, each consumer dies with probability $\alpha \in (0, 1)$ and earns income $I > 0$ if alive. The assumption that individuals are not subject to mortality risk in period 1 is only made to simplify notation. Our results would remain unchanged if we assumed that mortality shocks happened in both periods. Ruling out income shocks in period 2 also helps the analysis, but can be substantially generalized. The key assumption for the front-loading of equilibrium policies is that income shocks are more prevalent earlier in the policy.

To examine the role that surrendering plays in providing liquidity, we assume that any other assets that consumers may have are fully illiquid and, therefore, cannot be rebalanced after an income shock. While this extreme assumption greatly simplifies the exposition, our results still go through if part of
the assets could be reallocated. All we require is some liquidity motivation for surrendering, which is consistent with the empirical evidence noted earlier.\footnote{Similarly, none of our results change if we allow for consumption in period 0. In line with our illiquidity assumption, Daily, Lizzeri and Hendel (2008) and Fang and Kung (2010) assume that no credit markets exist in order to generate lapses.} To simplify notation, we assume that there is no discounting.

An insurance contract is a vector of (possibly negative) state-contingent payments

\[ T_j \equiv \left( t_{1,j}^S, t_{1,j}^{NS}, t_{A,j}^S, t_{D,j}^A, t_{A,j}^{NS}, t_{D,j}^{NS} \right) \in \mathbb{R}^6, \]

where \( t_{1,j}^S \) and \( t_{1,j}^{NS} \) are payments in period 1 when the consumer does and does not suffer the income shock. The variables \( t_{A,j}^S, t_{D,j}^A, t_{A,j}^{NS}, \) and \( t_{D,j}^{NS} \) denote the payments in period 2 when the consumer is alive (A) or dead (D) conditional on whether (S) or not (NS) he suffered an income shock in period 1.

A natural interpretation of these state-contingent payments is as follows. Consumers pay a premium \( t_{1,j}^{NS} \) for insurance when they buy a policy in period 0. In period 1, they choose whether or not to surrender the policy. If they do not surrender, the insurance company repays \(-t_{A,j}^{NS}\) if they survive and \(-t_{D,j}^{NS}\) if they die at \( t=2 \). If they surrender the policy, the insurance company pays a cash value of \( t_{1,j}^S - t_{1,j}^{NS} \) in period 1. Then, at \( t=2 \), they get paid \(-t_{A,j}^S\) if they survive and \(-t_{D,j}^S\) if they die.

Therefore, the timing of the game is as follows:

\( t=0 \): Each firm \( j = 1, \ldots, N \) offers an insurance contract \( T_j \). Consumers decide which contract, if any, to accept. Those who are indifferent between more than one contract randomize between them with strictly positive probabilities.

\( t=1 \): Consumers lose \( L \) dollars with probability \( l \) and choose whether to “surrender the policy” (i.e., report a loss to the insurance company). They pay \( t_{1,j}^{NS} \) if they do not surrender and \( t_{1,j}^S \) if they do.

\( t=2 \): Consumers die with probability \( \alpha \). Those who survive earn income \( I > 0 \), whereas those who die make no income. The household of a consumer who purchased insurance from firm \( j \) and surrendered at \( t=1 \) receives \(-t_{A,j}^S\) if he survives and \(-t_{D,j}^S\) if he dies. If the consumer did not surrender at \( t=1 \), his household instead receives \(-t_{A,j}^{NS}\) if he survives and \(-t_{D,j}^{NS}\) if he dies.

We follow the standard approach in contract theory by not imposing any exogenous restrictions on the space of contracts. Since we are interested in explaining the pattern of life insurance contracts observed in practice, it is important that front loading and lapse fees emerge endogenously in equilibrium, rather than through exogenous restrictions on the contract space. The only constraints that firms face in our model...
when designing their contracts are informational constraints, which arise because they cannot observe each household’s need for money (modeled through the income shock).20

Our three-period model allows us to study the key properties of life insurance policies described in Section 2, namely, the cross-subsidization between lapsers and non-lapsers and the front-loading of policies. In this model, lapsing or surrendering a policy corresponds to the change in premiums and coverage that follows an income shock. However, the model does not distinguish between term and permanent policies. We will, therefore, use the words lapsing and surrendering interchangeably in the context of the model. Moreover, we will say that a consumer “reports a loss to the insurance firm” with the understanding that such a direct mechanism is equivalent to more realistic indirect mechanisms where consumers lapse or surrender their policies following an unobserved need for money.

For expositional simplicity, the current version of the model makes two assumptions that we later relax. First, the model assumes that all consumers are subject to the same loss. While this assumption simplifies the analysis, it is not important for our main results. In Subsection 3.8, we allow for a continuum of possible losses and show how it generalizes the results and naturally leads to policy loans, as observed in practice with most permanent policies. Second, we do not include health shocks in the model. While health shocks are important, especially at older ages, our intent is to show how inattention towards income shocks alone can explain the key stylized facts in the life insurance market. Therefore, we do not want to complicate the analysis by adding other shocks, which would not overturn our main conclusions. We consider a rational expectations model of health shocks in Section 4 and demonstrate that health shocks alone cannot explain the pattern of life insurance policies described previously.

### 3.2 Consumer Utility

The utility of household consumption when the consumer is alive and dead is represented by the strictly increasing, strictly concave, and twice differentiable functions $u_A(c)$ and $u_D(c)$, satisfying the following Inada conditions: $\lim_{c \downarrow 0} u_A'(c) = +\infty$ and $\lim_{c \downarrow 0} u_D'(c) = +\infty$. The utility received in the dead state corresponds the “joy of giving” resources to survivors.

Since other assets are illiquid, there is a one-to-one mapping between state-contingent payments and state-contingent consumption $C_j \equiv (c_{1,j}^S, c_{1,j}^{NS}, c_{A,j}^S, c_{D,j}^S, c_{A,j}^{NS}, c_{D,j}^{NS})$, so there is no loss of generality in assuming that a contract specifies a vector of state-contingent consumption rather than state-contingent payments.21

Because firms do not observe income shocks, insurance contracts have to induce consumers to report

---

20 We also assume two-sided commitment, although, in reality, consumers are allowed to drop their policies. This is assumed for simplicity only, as our results persist if we assume instead that only insurers are able to commit. Our model also assumes that policies are exclusive. The equilibrium of our model would remain unchanged if we assumed that life insurance policies were non-exclusive (as they are in practice). Furthermore, allowing for positive liquidity shocks would not qualitatively affect our results if policies are non-exclusive. In that case, consumers would buy additional policies at actuarially fair prices following an unexpected positive liquidity shock.

21 The vector of state-contingent consumption is determined by $c_{1,j}^{NS} = W - r_{1,j}$, $c_{A,j}^{NS} = I - r_{A,j}$, $c_{D,j}^{NS} = W - L - r_{D,j}$, $c_{A,j}^S = I - r_{A,j}$, and $c_{D,j}^S = -r_{D,j}$. We can interpret $c_{D,j}^S$ and $c_{NS,D,j}$ as “bequest consumption.”
them truthfully in period 1. Those who experience the shock report it truthfully if the following incentive-compatibility constraint is satisfied:

\[ u_A(c^S_{1,j}) + \alpha u_D(c^S_{D,j}) + (1 - \alpha)u_A(c^S_{A,j}) \geq u_A(c^{NS}_{1,j} - L) + \alpha u_D(c^{NS}_{D,j}) + (1 - \alpha)u_A(c^{NS}_{A,j}). \] (1)

In words: Surrendering the policy must give consumers a higher utility than absorbing the loss. Similarly, those who do not experience the income shock do not report one if the following incentive-compatibility constraint holds:

\[ u_A(c^{NS}_{1,j}) + \alpha u_D(c^{NS}_{D,j}) + (1 - \alpha)u_A(c^{NS}_{A,j}) \geq u_A(c^S_{1,j} + L) + \alpha u_D(c^S_{D,j}) + (1 - \alpha)u_A(c^S_{A,j}). \] (2)

Our key assumption is that consumers do not take background risk into account when buying life insurance in period 0. Formally, they attribute probability zero to suffering an income shock at the contracting stage. Consumers, therefore, evaluate contracts in period 0 according to the following expected utility function that only includes states in which background shocks do not occur:

\[ u_A(c^{NS}_{1,j}) + \alpha u_D(c^{NS}_{D,j}) + (1 - \alpha)u_A(c^{NS}_{A,j}). \]

We refer to this expression as the consumer’s “perceived utility.”

### 3.3 Firm Profits

A firm’s expected profit from an insurance policy is the expected net payments it gets from the consumer, which, expressing in terms of consumption, equals the sum of expected income minus the sum of expected consumption. Conditional on not surrendering, expected income is \( W + (1 - \alpha)I \), whereas expected consumption is \( c^{NS}_{1,j} + \alpha c^{NS}_{D,j} + (1 - \alpha)c^{NS}_{A,j} \). Similarly, conditional on surrendering, expected income equals \( W - L + (1 - \alpha)I \) and expected consumption equals \( c^S_{1,j} + \alpha c^S_{D,j} + (1 - \alpha)c^S_{A,j} \).

### 3.4 Equilibrium

An equilibrium of the game is a vector of policies offered by each firm, a consumer acceptance decision, and a consumer surrender decision conditional on the policy and on the liquidity shock with the following
properties. 

1. Each firm’s policy maximizes the firm’s expected profits,

2. Each consumer chooses a policy that maximizes his/her perceived utility, randomizing with strictly positive probabilities when multiple policies give the same perceived utility.

3. Each consumer chooses whether or not to report an income shock to maximize his/her period-2 continuation utility.

The equilibrium can be calculated by solving two nested programs. First, we consider a lower-level program, which determines consumption conditional on the income shock. Because consumers do not incorporate the possibility of an income shock when choosing a policy, any accepted offer must maximize the firm’s expected profits following an income shock subject to consumers not misreporting the shock. Formally, for any fixed profile of consumption in the absence of an income shock \((c_{NS1}^{NS}, c_{NAj}^{NS}, c_{NSDj}^{NS})\), firms offer policies that maximize profits subject to the incentive-compatibility constraints (1) and (2). Let \(\Pi\) denote the maximum profit a firm can obtain conditional on the income shock:

\[
\Pi \left( c_{NS1}^{NS}, c_{NAj}^{NS}, c_{NSDj}^{NS} \right) \equiv \max_{c_{S1j}^{S}, c_{SAj}^{S}, c_{SDj}^{S}} W - L - c_{S1j}^{S} - \alpha c_{SDj}^{S} - (1 - \alpha) \left( c_{SAj}^{S} - I \right)
\]

subject to (1) and (2).

We will initially ignore the non-binding constraint (2) and verify that it is satisfied in the solution. Intuitively, the relevant incentive problem consists of inducing consumers to surrender a policy after a shock, rather than preventing those who did not suffer a shock from misreporting one.

Second, we consider a higher-level program, which determines consumption in the states where there is no income shock, taking into account how they affect consumption when there is an income shock. Recall that contracting happens before income shocks are realized. Firms offer an insurance policy as long as they obtain non-negative expected profits. Price competition between firms forces them to offer policies that maximize the consumer’s perceived utility among policies that give zero profits:

\[
\max_{c_{NS1}^{NS}, c_{NSDj}^{NS}} u_A(c_{NS1j}) + \alpha u_D \left( c_{NSDj}^{NS} \right) + (1 - \alpha) u_A \left( c_{NSA}^{NS} \right)
\]

subject to \(l \Pi \left( c_{NS1}^{NS}, c_{NAj}^{NS}, c_{NSDj}^{NS} \right) + (1 - l) \left[ W - c_{NS1j}^{NS} - \alpha c_{NSDj}^{NS} - (1 - \alpha) \left( c_{NSA}^{NS} - I \right) \right] = 0.\)

Lemma 1 establishes this result formally:

\(^24\)This corresponds to pure-strategy subgame-perfect Bayesian Nash equilibria of the game with a tie breaking rule that specifies that, when indifferent, consumers randomize with strictly positive probabilities. It turns out that, in this particular game, the restriction to pure strategies is without loss of generality.

\(^25\)There is no renegotiation of contracts in our model. The separation of the firm’s contract design program into two parts – before and after an income shock – does not correspond to a renegotiation between the parties. Instead, the program are nested because consumers attribute zero weight to a liquidity shock, so, at the contracting stage, firms will always offer contracts that maximize their profits conditional on such a shock.
Lemma 1. A set of state-dependent consumption \( \{C_j\}_{j=1,...,N} \) and a set of acceptance decisions is an equilibrium of the game if and only if:

1. At least two offers are accepted with positive probability,
2. All offers accepted with positive probability solve Program (3), and
3. All offers that are not accepted give consumers a perceived utility lower than the solutions of Program (3).

Since firms get zero profits in equilibrium, when there are more than two firms, there exist equilibria in which some firms offer “unreasonable” contracts that are never accepted. An equilibrium of the game is essentially unique if the set of contracts accepted with positive probability is the same in all equilibria. An equilibrium of the game is symmetric if all contracts accepted with positive probability are equal: if \( C_j \) and \( C'_j \) are accepted with positive probability, then \( C_j = C'_j \). The next lemma establishes existence, uniqueness, and symmetry of the equilibrium:

Lemma 2. There exists an equilibrium. Moreover, the equilibrium is essentially unique and symmetric.

Because the equilibrium is symmetric, we omit the index \( j \) from contracts that are accepted with positive probability. We now present the main properties of the equilibrium contracts:

Proposition 1. In the essentially unique equilibrium, any contract accepted with positive probability has the following properties:

1. \( u'_A(c^S_A) = u'_D(c^S_D) = u'_D(c^S_D) \),
2. \( u'_D(c^{NS}_D) = u'_A(c^{NS}_A) < u'_A(c^{NS}_1) \),
3. \( \pi^S > 0 > \pi^{NS} \), and
4. \( c^{NS}_1 - L < c^S_1 \).

Condition 1 states that there is full insurance conditional on the income shock. Since insurance companies maximize profits conditional on the income shock subject to leaving consumers with a fixed utility level (incentive compatibility), the solution must be on the Pareto frontier conditional on the shock, thereby equating the marginal utility of consumption in all states after the income shock.

The equality of Condition 2 states that consumers are fully insured against mortality risk conditional on not suffering an income shock. Because risk-averse consumers and risk-neutral firms are fully aware of the risk of death, firms fully insure consumers against mortality risk.

The inequality of Condition 2, however, shows that the insurance policy also induces excessive saving relative to efficient consumption smoothing, which equates the marginal utility of consumption across periods. Intuitively, shifting consumption away from period 1 increases the harm of the income loss if it were to occur, thereby encouraging consumers to surrender their policies and produce more profits.
for firms after an income shock. More formally, the excessive savings result follows from incentive compatibility after an income shock: shifting consumption from period 1 to period 2 increases the cost of absorbing the liquidity shock. Consumers are fully aware of the intertemporal wedge induced by the equilibrium policy. Nevertheless, this wedge persists in equilibrium because, since consumers do not believe they will surrender their policies in period 1, any firm that attempts to offer a contract that smooths inter-temporal consumption would be unable to price it competitively.

Condition 3 states that firms obtain positive profits if the consumer surrenders the policy and negative profits if he does not. Since, in expectation, insurance companies make zero expected profits, the profits obtained after an income shock are competed away by charging a lower price from policyholders who do not experience an income shock and, therefore, hold their policies to term.

Condition 4 states that if, after an income shock, a consumer decided not to surrender the policy and, instead, reduce consumption to absorb the loss, his consumption in that period would decrease by more than if he surrendered the policy. This can be interpreted as paying a positive cash value (such as with permanent policies), thereby allowing the policyholder to increase consumption in that period. It can also be interpreted as paying a positive load into the policy in order to keep coverage, such as with term policies.

Let $\bar{C}_s^s \equiv c_s^1 + \alpha c_s^D + (1 - \alpha) c_s^A$ denote the expected consumption conditional on each shock $s \in \{S, NS\}$. The next proposition determines how changes in the probability of lapsing $l$ affect the equilibrium policies.

**Proposition 2.** In the essentially unique equilibrium, any contract accepted with positive probability has the following properties:

1. $c_s^A$ and $c_s^D$ are strictly increasing functions of $l$,
2. $c_s^1$, $c_s^S$, $c_s^A$, and $c_s^D$ are strictly decreasing functions of $l$,
3. $\bar{C}_s^{NS}$ is strictly increasing and $\bar{C}_s^{S}$ is strictly decreasing in $l$.

Conditions 1 and 2 imply that, for consumers who do not lapse, the difference between premiums paid in the first period and in the second period if alive is increasing in the probability of a liquidity shock, $l$. Therefore, the policy becomes more front loaded as the probability of the liquidity shock increases. Condition 3 states that the expected consumption if the policyholder does not suffer a liquidity shock increases in $l$, whereas the expected consumption in case of a liquidity shock decreases. Thus, the model predicts that surrender fees increase in the probability of lapsing, where the surrender fee is defined as the amount the consumer loses after the liquidity shock.

Because firms and consumers disagree on the probability of lapsing, they speculate by trading a policy with high premiums conditional on a liquidity shock and cheap premiums otherwise. However, the difference between first- and second-period premiums is $r_1^{NS} - r_A^{NS} = W - I - e_1^{NS} + c_A^{NS}$, which, by Proposition 2, is increasing in $l$. 

---

26The difference between first- and second-period premiums is $r_1^{NS} - r_A^{NS} = W - I - e_1^{NS} + c_A^{NS}$, which, by Proposition 2, is increasing in $l$. 

15
because the firm does not observe the liquidity shock, the policy has to induce the consumer to report it truthfully. This is achieved by front loading the premium payments, which disproportionately increases the cost of a premium after a liquidity shock. But, because consumers value smooth consumption, front-loaded premiums are costly. Then, the front load balances the “benefit” from speculation (i.e., the firm’s benefit from exploiting consumer bias) against the cost of a less balanced consumption when there is no shock. Since a higher probability of the liquidity shock raises the value of exploiting the consumer bias, it increases the front load.

Notice that our definition of surrender fee includes both explicit and implicit fees. Therefore, it includes not only the explicit fees that are standard in permanent insurance but also the implicit fees that are substantial in term insurance. More precisely, because term insurance policies typically have no cash value, all previously-paid insurance loads are implicit surrender fees.

To summarize, consumers in our model “lapse” or “surrender” after suffering an unexpected background shock. Consistent with the empirical evidence reported in Subsection 2.2, although insurance companies do not get extraordinary profits, there is cross subsidization: They make positive profits on consumers who lapse and negative profits on those who do not (Condition 3 from Proposition 1). The equilibrium policy shifts consumption into the future (Condition 2 of Proposition 1), so initial premiums are high and later premiums are low. Shifting consumption into the future magnifies the impact of an income shock and induces consumers to surrender their policies, thereby raising the firm’s profits. Of course, these profits are competed away in equilibrium.

### 3.5 Testing the Comparative Statics of the Model

As we examine in Section 4, even consumers with rational expectations may demand lapse-based policies. In these cases, lapse fees balance the benefit of being “locked into” a policy against the cost of being unable to smooth consumption after a liquidity shock. Since this expected cost is increasing in the probability of a liquidity shock, these rational expectations models predict that surrender fees should decrease in the probability of facing a liquidity shock. In contrast, Proposition 2 predicts that surrender fees should increase in the probability of facing liquidity shocks in our model. Hence, the empirical relationship between surrender fees and the probability of liquidity shocks allows us to test our model and distinguish it from alternatives.

In order to evaluate the relationship between surrender fees and the probability of liquidity shocks, we hand-collected detailed whole life insurance data from two national insurance companies, MetLife and SBLI, for both genders, across all American states except for New York.\(^{27}\) MetLife is the largest U.S. life insurer with over $2 trillion in total life insurance coverage in force while SBLI has about $125 billion

\(^{27}\)See Online Appendix C for a more detailed description of the data. Unfortunately, these two companies did not sell this type of policies in the state New York. Whole-life policies are typically used differently than Universal Life (UL) policies, as UL policies are often used a tax-preferred investment vehicle in addition to insurance. In order to verify the robustness of our findings to other companies, we also collected data from other insurance firms for the state of California and obtained the same results.
of coverage in force. The data set consists of policies for ages between 20 and 70 in five-year increments and for face values of $100,000, $250,000, $500,000, $750,000 and $1,000,000, adding up to a total of 10,738 policies. The surrender fee for each policy corresponds to the proportion of the discounted sum of insurance loads (i.e., present value of premiums paid in excess of the actuarially fair price) that cannot be recovered as cash surrender value. Thus, the surrender fee is the fraction of pre-paid premiums that cannot be recovered if the policy is surrendered. To ensure the comparability of the policies, we kept the terms of each policy constant except for our controls (coverage, ages, and genders). We, therefore, focused on policies for the “preferred plus” health category that require a health exam.

Proposition 2 implies that surrender fees should increase in the probability of liquidity shocks. In order to test this prediction, we need observable measures of the probability of liquidity shocks. Since we have detailed policy data but no information about the individuals who buy each policy, we need to proxy for the probability of liquidity shocks using the terms of the policies. We use two different proxies: age and coverage. It is widely documented that younger individuals are more likely to be liquidity...
constrained, and age is a frequently-used proxy for the presence of liquidity constraints. Moreover, individuals who purchase smaller policies tend to be less wealthy and more likely to be liquidity constrained. In fact, consistently with these proxies, lapse rates are decreasing in both age and coverage (Section 2). The model therefore predicts that surrender fees should decrease in the age of the policyholder and in the level of coverage.

The data is strongly consistent with these predictions. Consider first the role of age. Figure 5 shows the mean surrender fees as a function of policy duration at each age along with their associated 95% confidence intervals, where standard errors are clustered by policy. Because whole life policies do not have a cash surrender value during the first few years after purchasing, surrender fees start at 100% for each age. As policies mature, they accumulate cash value, reducing the surrender fee. Our interest, however, is in the difference in surrender fees for policies sold to individuals of different ages. For both MetLife and SBLI, notice that the surrender fees decrease in age, at each duration. Thus, as predicted by the model, younger individuals face higher surrender fees.

Figure 6 shows the mean surrender fees for different coverage amounts and their associated 95% confidence intervals. Standard errors are clustered by policy.

---

28 See, for example, Jappelli (1990), Jappelli, Pischke, and Souleles (1998), Besley, Meads, and Surico (2010), and Zhang (2014).
confidence intervals. As predicted by the model, surrender fees for both MetLife and SBLI policies are decreasing in coverage. However, while for MetLife the difference is always statistically significant, for SBLI policies with coverage above $100,000 are not statistically significant at 5% level.\footnote{The lack of statistical significance for SBLI policies with more than $100,000 coverage could be due to the fact that, while lapse probabilities are much higher for smaller policies, the difference is not very large for policies with coverage above $200,000 (see Figure 2).}

The differences by age are not only statistically significant; they are also economically large. For example, while a 20-year-old policyholder who surrenders after 5 years would not collect any cash value, a 70 year old collects about 30% of the amount paid in excess of the actuarily fair prices. Differences by coverage levels are slightly smaller. Nevertheless, the surrender fee on a $100,000 policy is, on average, between 5 and 10 percentage points larger than the surrender fee on a $1,000,000 policy.

### 3.6 Testing the Mechanism

While Section 4 shows that policy data is consistent with the comparative statics from the model, it does not directly test the mechanism underlying our model, namely, whether consumers underestimate the probability of lapsing. To elicit policyholder beliefs about the probability of lapses, we developed and implemented a survey with the universe of customers from TIAA-CREF who purchased term life insurance in the previous two years.\footnote{We also have data from those who bought permanent policies. Our qualitative results do not change when we include permanent policies. However, we exclude permanent policies because TIAA-CREF permanent policies include a unique type of universal life policy that is being used by TIAA-CREF’s wealth management group as a tax-efficient investment vehicle for people who have maxed out their tax preferred (e.g., 401(k), 403(b), or IRA) contributions. Our data from TIAA-CREF, however, does not include identifiers that would allow us to separate these investment products from more traditional types of permanent policies. (Officials at TIAA-CREF verified – in fact, they were the first to point out – that their unique investment-focused permanent policies are quite different from their other permanent policies.) In contrast, all TIAA-CREF’s term policies provide a pure form of life insurance, representing a direct test of our model.}

Unlike its retirement accounts, TIAA-CREF life insurance is widely marketed to the general U.S. population. Along with MassMutual, MetLife, NYLife, and State Farm, TIAA-CREF is one of just a handful of large U.S. life insurers that regularly receives an A.M. Best Company rating of A++. In this section, we briefly summarize the main result from our survey. The Appendix presents a detailed description of the survey and formal regression results. The survey asked customers up to 15 questions regarding their reasons for buying life insurance, the channel through which they bought it, their beliefs about the chances and reasons for lapsing, and their beliefs about future income shocks. We also have detailed information about the customers and their policies, including age, risk class, marital status, education, job type, job tenure, and the type and size of their policies.

The survey was accessible through an e-mail sent to all customers who purchased life insurance in the previous two years. Those who didn’t respond were sent two reminders, the first one a week after the original email was sent and the second one a day before the survey deadline.\footnote{Ideally, one would elicit customers’ beliefs prior to buying insurance. However, concerns about how asking these questions might affect their purchasing decision prevented us from being able to implement this approach. We, therefore, focused on customers who bought insurance recently.} The response rate –
approximately 15% – was slightly above TIAA-CREF’s typical rate.

For our purposes, the most important question was question 6, which asks:

*Your life insurance policy has about n years left on it. Do you plan to stop your policy (sometimes called lapsing) before then?*

(The value \( n \) was set equal to the actual value for that customer.) 97.2% of respondents answered “No”, whereas 2.8% said that they planned to stop their policy before its expiration, indicating that most policyholders do not think there is a considerable chance of lapsing. In contrast, TIAA-CREF’s historical experience with these policies suggests that lapses are very common, similar to the industry as a whole. The average lapse rate on these TIAA-CREF policies for the past 15 years was 5.2% per year. As a back-of-the-envelope calculation, suppose policyholders face a constant lapse rate of 5.2% per year. Since policies in our sample have, on average, 16.8 years left, approximately 60% of these policies will lapse.\(^{32}\)

We asked respondents who indicated that they thought they were going to lapse the possible reasons for doing so. The most common answers were “*Insurance premiums are too costly*” (43%) and “*My needs changed, and I don’t need life insurance anymore*” (24%). In contrast, none of the respondents selected the option, “*I feel healthier than I expected, and so I am planning to buy a different policy*”, an explanation that would have been consistent with reclassification risk.

In order to understand the reason for underweighting the probability of lapsing, we asked customers about income fluctuations. Out of the 751 survey respondents, 204 (27.2%) reported an income loss in the last 5 years, whereas 189 (25.2%) expected an income loss in the next 5 years. Despite the prevalence of income losses, they did not translate into beliefs about lapsing. The correlation between expecting an income loss and expecting to lapse was 0.050 and is not statistically significant. This result suggests that overconfidence or optimism about future income is not the key explanation for underweighting lapses.

### 3.7 Inefficiency and the Effect of Secondary Markets

We believe that the appropriate efficiency criterion in our model evaluates consumer welfare according to the correct distribution of income shocks. Therefore, we say that an allocation is *efficient* if there is no other allocation that increases the expected utility of consumers (evaluated according to the true probability distribution over states of the world) without decreasing the expected profit of any firm. Because consumers are risk averse and insurance companies are risk neutral, any efficient allocation should produce constant marginal utility of consumption across all states (full insurance).

The equilibrium of the model, therefore, is inefficient in two ways. First, because the marginal utility of consumption increases after the shock, there is incomplete insurance with respect to the income shock. Of course, this source of inefficiency is standard in models with unobservable income shocks. However,\(^{32}\)
differential attention further exacerbates the effect of income shocks by transferring consumption from the shock state (where marginal utility is high) to the no-shock state (where marginal utility is low). Second, because consumption is increasing over time when there is no income shock, there is incomplete intertemporal consumption smoothing. This second source of inefficiency is not standard and is produced by differential attention where consumers fail to account for background shocks.

In Appendix E, we formally study the effects from introducing a competitive secondary market for life insurance policies in our model. In a secondary market, individuals may resell their policies to risk-neutral firms, who then become the beneficiaries of such policies. We consider the short- and long-run effects. The “short-run equilibrium” takes the primary market policies obtained previously (i.e., in the model in which there is no secondary market) as given. The “long-run equilibrium” allows primary market firms to anticipate contracts that will be offered in the secondary market.

The equilibrium policies in our model produce two sources of profitable trade between consumers and firms in the secondary market. First, policies generate a cross-subsidy from policyholders who lapse to those who do not lapse. Therefore, firms in the secondary market can profit by buying policies from consumers who would lapse, splitting the primary firm’s profits with the policyholder. In turn, this renegotiation reduces the profits of firms in the primary market, who are then left only with policies that do not lapse. Second, policies are front-loaded relative to the prices consistent with an optimal inter-temporal consumption smoothing. By renegotiating on the secondary market, consumers are able to obtain a smoother consumption stream. Therefore, in the short run, the introduction of a secondary market makes consumers better off and primary market insurers worse off. Firms in the secondary market obtain zero profits by our perfect competition assumption.

In the long run, primary market firms anticipate that any source of profitable ex-post renegotiation will be arbitrated away in the secondary market. As a result, they offer policies that are neither front-loaded nor lapse-based. Nevertheless, because consumers do not anticipate background shocks, there is still imperfect consumption smoothing as consumption falls after the shock. Firms earn zero profits in both primary and secondary markets, while consumers are better off.33

Taking into account both short- and long-run effects, it is clear that primary insurers would oppose the rise of secondary markets despite the improvement in efficiency.

3.8 Heterogeneous Shocks

So far, we have assumed that the possible background loss $L$ could only take one possible value, which was known by insurance firms. In practice, insurance firms do not know the size of the possible loss, both because consumers are heterogeneous in unobservable ways and because consumers are subject to

---

33Perhaps surprisingly, consumers who do not take background risk into account would not ex-ante favor a regulation that allows insurance to be sold at a secondary market. Therefore, in our model, the same behavioral trait that introduces inefficiency in the competitive equilibrium also prevents majority voting from implementing an efficiency-enhancing regulation. See, for example, Bisin, Lizzeri, and Yariv (2015) and Warren and Wood (2014) for interesting analyses of political economy based on behavioral economics models. They would, of course, favor such a regulation ex-post.
multiple losses. We now relax the assumption that firms know the size of the possible loss. All the main previous results remain unchanged, except that now firms will offer policies with “coverage reduction” terms. These terms can be interpreted as policy loans, which allow the policyholder to borrow from the policy at a fee. The loan is either repaid in a future period or subtracted from the face value of the policy. These loans are common in practice. According to LIMRA (2015, P. 10), life insurer loans to policyholders against the cash value of their life insurance policies amount to $133 billion by year-end 2014.

More formally, we now assume that firms believe that, with probability \( l \), consumers face a loss \( L \) that is distributed according to a density function \( f \) with full support on the interval \([L, \bar{L}]\subset \mathbb{R}_+\). With probability \( 1-l \), the consumer does not suffer an income loss \((L = 0)\).

The model is otherwise unchanged from Section 3 except that firms now offer policies with a menu of payments conditional on each possible realization of the income shock. Therefore, the only difference is that the lower-level program now involves a continuum of possible losses. This program can be written as a screening model with a type-dependent participation constraint. Despite the non-transferability of utility in our context, we can characterize the solution using standard methods by working with the promised continuation utility, which enters the utility function linearly and, therefore, plays the same role as transfers in quasi-linear environments. Because the liquidity shock \( L \) is the consumer’s private information, we refer to \( L \) as the consumer’s type.

There are two sets of incentive-compatibility constraints. First, types have to report their income losses truthfully rather than absorb them and pretend not to have suffered any loss:

\[
\begin{align*}
    u_A(c_A^S(L)) + \alpha u_D(c_D^S(L)) + (1-\alpha)u_A(c_A^1(L)) \\
    \geq u_A(c_A^{NS} - L) + \alpha u_D(c_D^{NS}) + (1-\alpha)u_A(c_A^{NS}) \quad \forall L.
\end{align*}
\]

Second, types have to report their true income loss instead of claiming a different loss amount:

\[
\begin{align*}
    u_A(c_A^S(L)) + \alpha u_D(c_D^S(L)) + (1-\alpha)u_A(c_A^1(L)) \\
    \geq u_A(c_A^S(\hat{L}) - L + \hat{L}) + \alpha u_D(c_D^S(\hat{L})) + (1-\alpha)u_A(c_A^S(\hat{L})) \quad \forall L, \hat{L}.
\end{align*}
\]

As in Subsection 3.4, any equilibrium must maximize the firm’s expected profit conditional on an income shock \((L \neq 0)\) subject to the incentive-compatibility constraints above.

The following result, characterizing the equilibrium of the model, is proven in Online Appendix I:

**Proposition 3.** In the equilibrium of the model with a continuum of losses, any contract accepted with positive probability has the following properties:

1. \( u_D'(c_{DS}^{NS}) = u_A'(c_{SA}^{NS}) < u_A'(c_{SI}^{NS}) \),

2. \( u_D'(c_{DS}(L)) = u_A'(c_{SA}(L)) < u_A'(c_{SI}(L)) \) for all \( L < \hat{L} \),
3. \( u'_A(c_S^1(L)) = u'_D(c_S^L(L)) = u'_D(c_A(L)) \).

4. \( \dot{c}_S^1(L) \geq 1, \dot{c}_S^A(L) \leq 0 \) and \( \dot{c}_S^D(L) \leq 0 \).

5. \( \pi^S(L) \geq 0 > \pi^{NS}, \) and \( \pi^S(L) \) is strictly increasing.

As in the model with a single possible loss, insurance premiums are front loaded for those who do not suffer a liquidity shock (Part 1). Among policyholders who suffer a liquidity shock, all types except for the one with the highest shock \( (L = \bar{L}) \) also get front-loaded premiums (Parts 2 and 3). Thus, lapse fees induce all but the types with the highest need for liquidity to have incomplete intertemporal smoothing.

Part 4 implies that the premium paid to the insurance company in the first period net of the liquidity shock \( (W - c_S^1(L) - L) \) — is decreasing in the loss \( L \), whereas the payments received from the insurance company in the second period \( (c_A(L) - I) \) and \( c_D(L) \) are increasing in \( L \). Hence, as with permanent policies in practice, the equilibrium policies allow the policyholder to reduce the premiums paid in the first period in exchange for lower coverage in period 2.

It may seem counterintuitive that firms would allow consumers to borrow from their policies rather than try to induce them to lapse. However, offering policyholders with intermediate losses policy loans allows firms to extract higher rents from those with greater liquidity needs. Since types would like to pretend to claim a lower liquidity need, firms screen a consumer’s need for money by charging different fees for different policy loans. The higher the shock, the higher the fee. If, instead, the firm tried to induce an intermediate type \( L^* \) to lapse, it would need to provide a larger cash value, which would entail leaving higher information rents to all types above \( L^* \), who benefit from having better surrender conditions.

Only the policyholders with the highest shock get an efficient allocation, which completely smooths their consumption. We can interpret their action as surrendering an old policy and replacing it with a new one with a lower face value. All other types get front-loaded policies (i.e., policies that induce inefficiently low consumption in period 1).

Part 5 states that, as in the model with a single loss, firms make positive profits when consumers suffer an income shock and negative profits when they do not. Moreover, profits are increasing in the size of the loss.

4 Other Potential Explanations

Our goal in this paper is to provide a model that simultaneously explains both the demand and the supply side of the life insurance market. There are many possible explanations for why a consumer may prefer a front-loaded life insurance policy, holding the design of policies fixed, and for why a life insurer may offer a front-loaded and lapse-based policy, taking consumers’ decisions as fixed. It is much harder to provide a unified account of both consumers’ and the life insurers’ decisions. In this section, we discuss other potential explanations that also account for both the demand and the supply side of the life insurance market. Some of them produce back-loaded rather than front-loaded premiums. Moreover, most
alternative models produce comparative statics that are the opposite of those from Proposition 2, which were tested in Subsection 3.5. Finally, each competing model produces some additional counterfactual pieces of evidence.

4.1 Reclassification Risk

The conventional view is that policy loads help enforce continued participation in an insurance pool when policyholders learn more about their mortality likelihood over time (“risk reclassification”). Without a load, policyholders who enjoy a favorable health shock – that is, an increase in conditional life expectancy – will want to drop from the existing risk pool and re-contract with a new pool, thereby undermining much of the benefit from intertemporal risk pooling. Ex-ante identical policyholders, therefore, contract on a dynamic load that punishes those who leave the pool. If reclassification is the only relevant risk and consumers can borrow, then the load will be constructed to be sufficiently large to prevent any lapsing. With a second “background” risk, such as a liquidity shock, some lapses may occur in equilibrium since rational policyholders now value ex-ante the option to lapse after a sufficiently negative liquidity shock.

It is, of course, impossible to reject every conceivable source of informational asymmetry as a motivation for lapse-based pricing. But the risk reclassification model faces several challenges for being the primary explanation of the patterns observed in the life insurance market.

First, as noted in Section 1, the reclassification model with rational agents produces a much lower level of lapses than found in the data. With rational expectations, most households avoid lapsing except after fairly extreme shocks. Second, the empirical support for the actual mechanism of the risk reclassification model is much weaker in the context of life insurance relative to other markets, such as health insurance (Handel, Hendel, and Whinston 2015). Using the comparatively old population in the Health and Retirement Study, where health shocks are likely to be more prevalent, Fang and Kung (2012, P.11), for example, show that people who lapse after a health shock tend to be less healthy than those who keep their policies, more consistent with the need for liquidity to cover medical expenses.

Third, a plausibly calibrated risk reclassification model counter-factually predicts that lapse fees should increase with the age of a contract, that is, be relatively more back-loaded than front-loaded.

The third point requires more elaboration. In Online Appendix F, we extend the models of Hendel

---

34 In the context of a car insurance market, Dionne and Doherty (1994) consider a two-period model with persistent risk types and show that, with one-sided commitment, firms make positive rents in the second period. Since most life insurance policies require a health exam and health status typically evolves over the consumer’s life, the assumption of privately-known persistent risk may not be well suited for the life insurance market. Accordingly, in their seminal paper, Hendel and Lizzeri (2003) present a model in which risks are common knowledge at the contracting stage and consumers are subject to health shocks. They show that, in the absence of credit markets, front loads are set according to a trade-off between reducing reclassification risk and enhancing consumption smoothing. Daily, Hendel, and Lizzeri (2008) and Fang and Kung (2010) extend this model by incorporating a bequest shock, according to which policyholders lose all their bequest motives.

35 See, in particular, their Table 6 (PP. 11), which shows the determinants of lapses in a multinomial logit regression. As they argue, “individuals who have experienced an increase in the number of health conditions are somewhat more likely to lapse all coverage, though the effect is not statistically significant.” In their structural model, which assumes that individuals choose coverage rationally, they find that younger individuals (among the relatively old population in the HRS) mostly lapse due to i.i.d. shocks. As individuals age, however, the importance of health shocks grows.
and Lizzeri (2003), Daily, Hendel, and Lizzeri (2008), and Fang and Kung (2010) by adding an initial period in which consumers are subject to an unobservable liquidity shock. Consumers are then subject to liquidity shocks in the first period, health shocks in the second period, and mortality risk in the third period — a stylized representation of the fact that health shocks are more important later in life.\footnote{In Daily, Hendel, and Lizzeri (2008) and Fang and Kung (2010), individuals live for two periods and are subject to both a health shock and a bequest shock in the first period. In Online Appendix F, we study the temporal separation of shocks, capturing the idea that non-health shocks are relatively more important earlier in life and health shocks are more important later in life.} We discuss the main results here.

Recall that the optimal surrender fee in the risk reclassification model balances the benefit from pooling (discouraging lapsing after positive health shocks) against the cost of preventing the consumer from obtaining a smoother consumption stream after a background shock. However, as Online Appendix G documents, younger people are quite likely to remain healthy; health shocks only become material at older ages.\footnote{See also Jung (2008).} The reclassification risk model then predicts that policies should not charge a positive lapse fee if the individual decides to lapse early on. The reason is that lapse fees exist to penalize agents who drop out due to favorable health shocks, thereby ensuring that the pool remains balanced. Charging a lapse fee for non-health related shocks is inefficient, as it exacerbates the consumer’s demand for money and undermines the amount of insurance provision. So, consumers who are more likely to suffer liquidity shocks — e.g., younger consumers and those who buy smaller policies — should be offered lower surrender fees, contrary to practice (Figures 5 and 6). More generally, since the importance of health-related shocks increases with age, the risk reclassification model predicts that lapse fees will increase (in real value) as people age, contrary to the observed decreasing pattern (Figures 5 and 6). Moreover, insurance companies would not profit from policies that lapse early on. In contrast, the empirical evidence presented earlier shows that insurers make considerable profits on policies that lapse.

\subsection{4.2 Time Inconsistency}

A large literature in behavioral economics has established that illiquid assets may be valuable to time-inconsistent individuals because they serve as commitment devices. Since front-loaded premiums reduce the incentive to drop the policy, time inconsistency may, at first glance, explain why insurance policies are front loaded.

DellaVigna and Malmendier (2004), for example, study a market where firms sell an indivisible good to time-inconsistent consumers. Heidhues and Kőszegi (2010) embed their framework in a model of credit cards. When consumers are sophisticated, firms offer a contract that corrects for time inconsistency, implementing the efficient level of savings. In a context of life insurance without background shocks, this model corresponds to a front-loaded policy that equates marginal utilities across periods, producing no lapses in equilibrium. Because zero lapsing is counterfactual, we need to introduce a motive for lapses to occur in equilibrium. There are two natural sources: partial naiveté or background shocks. We study
each of them formally in Online Appendix F and summarize the main results here.

Partially naive consumers underestimate their time inconsistency. Heidhues and Kőszegi (2010) show that competitive equilibrium contracts for partially naive consumers have front-loaded repayments and the option to postpone the client’s payments in exchange for a large future fee. Because consumers underestimate their time-inconsistency, they believe that they will repay the debt up front but end up refinancing it, effectively using back-loaded contracts. As mentioned previously, back-loaded life insurance policies do not exist. We could prevent back loads within this model by assuming that consumers cannot commit to keeping their policies. In this case, however, only actuarially fair policies, which are also not front loaded, are accepted in the competitive equilibrium. In sum, partial naivety cannot account for the front-loaded insurance policies observed in practice.

Alternatively, we could introduce a background shock that motivates lapsation in the sophisticated model. The equilibrium surrender fee then balances the benefit from providing commitment against the cost of precluding efficient lapses after such a shock. Policies have to be designed to prevent time-inconsistent policyholders from pretending to have experienced an income shock in order to increase their present consumption. Then, because the binding incentive constraint is now the one preventing consumers from pretending to have suffered an income shock (which is the non-binding constraint in our model), policies are distorted in the opposite direction. That is, policies are actually back loaded. In fact, policies designed for individuals who suffer a liquidity shock are even more back loaded than the time-inconsistent self would prefer.

In addition, firms would typically charge negative surrender fees in the sophisticated model. More formally, the equilibrium policies generally produce a cross-subsidy from consumers who have not suffered an income shock to those who have, just the opposite direction from our model and the pattern observed in practice. Intuitively, recall that equilibrium contracts feature a trade-off between providing commitment and insuring consumers against liquidity shocks. A surrender fee transfers resources away from consumers after an income shock, precisely when their marginal utility is the highest. Thus, only when the commitment problem is sufficiently intense relative to the benefit from consumption smoothing does it make sense to charge positive surrender fees. As we show numerically in Online Appendix F, the equilibrium only features positive surrender fees if the commitment problem is quite severe and consumers are fairly risk tolerant; otherwise, surrender fees are negative.

Summarizing, the model of time-inconsistency with liquidity shocks yields back-loaded policies. Moreover, whenever the commitment problem is “not too intense” relative to the policyholder’s risk aversion, it predicts negative surrender fees. In practice, policies are front-loaded and have large surrender fees.

4.3 Fixed Costs

Insurance companies may also charge surrender fees in order to recover sales commissions paid to brokers. But there are two problems with this explanation for the life insurance pricing observed in practice.
First, commissions are endogenous; companies choose how to structure their sales commissions. An explanation for front-loaded premiums that is based on the fact that sales commissions are front loaded needs to justify why commissions are front loaded in the first place. In fact, commissions paid to insurance brokers highly encourage selling to shorter-term consumers. While their commissions may last several years, the bulk of the payment is typically made in the first or second year. However, commissions are often not paid if the policy is surrendered in the first year since then the insurer could lose money.\footnote{For example, Genworth Life’s (2011) commission schedule reads: “In the event a withdrawal or partial surrender (above any applicable penalty-free amount) is granted or a policy or contract is surrendered or canceled within the first twelve (12) months after the date specified in paragraph (c) of this Section 2, compensations will be charged back to you as follows: 100% of compensations paid during that twelve (12) month period.”} In contrast, commissions paid to wealth managers, for example, are fairly proportional to the actual fee revenue collected from clients, thereby encouraging the wealth manager to keep the relationship active.\footnote{With broker-dealers, the client typically pays an initial fee along with a trailer fee that is proportional to ongoing assets under management. With fiduciary financial advisors, clients typically pay just a fee that is proportional to their assets being managed. In both cases, the wealth advisor collects a proportion of the revenue collected from clients, and so the product provider does not actually lose money if the client leaves. Moreover, all wealth managers are incentivized to keep clients active because of the potential to collect ongoing fees.}

Our model suggests that front-loaded commissions may be used to incentivize insurance brokers to find clients without concern for whether they will hold their policies for very long.

Second, since the bulk of commissions are paid in the first year, lapse fees should be constant after the first few years, when commissions are no longer paid. That is, according to the fixed cost story, insurance firms should not obtain different actuarial profits if consumers lapse after 5, 10, or 20 years since they do not have to pay any additional commissions after the first few years. Empirically, however, actuarial profits are substantially different if policies lapse after 5, 10, or 20 years (see Figure 3).\footnote{The decreasing relationship between surrender fees and coverage can be explained by the need to recover some fixed costs. This explanation, however, cannot account for the strong decreasing relationship between age and surrender fees.}

Alternatively, insurance companies may charge surrender fees in order to discourage lapses, reducing the insurance company’s needs for liquid assets, allowing it to obtain higher returns on its portfolio by making more illiquid investments. If consumers have rational expectations about their liquidity needs, the optimal surrender fee should, therefore, balance the gains to the insurance company’s portfolio against the costs of preventing policyholders from adjusting their consumption after liquidity shocks. This theory, however, also predicts patterns of surrender fees that are inconsistent with actual practice. Because younger individuals tend to be more liquidity constrained, the cost of preventing them from adjusting their consumption after a liquidity shock is relatively high. Thus, firms should offer them relative lower surrender fees (at higher premiums to compensate for the lower surrender fees). In practice, we observe the opposite pattern (Figure 5). Moreover, because larger policies require more liquid assets to be held by insurers in order to repay those who surrender, and because larger policies are typically purchased by wealthier individuals with lower liquidity needs, we should expect surrender fees to increase with policy size. In practice, surrender fees weakly decrease with policy size (Figure 6).\footnote{The decreasing relationship between surrender fees and coverage can be explained by the need to recover some fixed costs. This explanation, however, cannot account for the strong decreasing relationship between age and surrender fees.}
5 Conclusion

This paper documents three stylized facts in the life insurance market — substantial lapsation, lapsed-based pricing, and front-loading of premiums — and shows how a simple model with differential attention can explain them. We also showed that the front-loading pattern of premiums observed in practice is not consistent with other explanations, including the standard model of insurance against reclassification risk, either naive or sophisticated time inconsistency, or the presence of fixed costs. Moreover, using actual policy data from two national life insurers, we indirectly test a new comparative static prediction of our model that also provides a clear discriminatory test against other potential explanations. The data strongly supports our model. We also developed and implemented a survey of customers who recently purchased life insurance from TIAA-CREF in order to directly test the central mechanism of our model. Those results are strongly consistent with differential attention and not consistent with other behavioral mechanisms, such as optimism.

References


LIMRA, 2011B. “Person-Level Trends in U.S. Life Insurance Ownership.” Windsor, CT.


Appendix: Survey Results

This Appendix describes in detail the survey we developed and implemented. A link to the survey was sent by email to the universe of customers from TIAA-CREF who had purchased life insurance in the previous two years. The survey was available for three weeks. In addition to the original email, two reminders were sent to customers who had not yet completed the survey. The first reminder was sent a week after the original email. The second was sent a day before the survey deadline.

Following a standard TIAA-CREF procedure, customers were randomly split into two waves. Participants in the first wave were sent the email to participate on the survey on September 1, 2015. Those in the second wave were sent the email on September 23, 2015. The same exact survey was implemented in both waves. The response rate was 14.63%, which is slightly above the TIAA-CREF average for surveys conducted by email. As explained in footnote 30, we focus on consumers who purchased term policies. This leaves us with a total of 751 respondents.

Figure 7 shows the characteristics of consumers who did and did not respond to the survey. Respondents and non-respondents are statistically indistinguishable in terms of gender and marital status. However, consumers who purchased their policies more recently are slightly more likely to respond. Respondents are about 2.3 years older and slightly healthier than non-respondents. They also buy very similar policies: policy terms and additional features of respondents and non-respondents are statistically insignificant from each other, although respondents tend to have lower death benefits. Finally, a slightly higher proportion of college employees responded to our survey, perhaps because we identified ourselves in the opening page and stated that the survey would be used for academic research.

Online Appendix D presents all questions in our survey. After the first page, which informs subjects of the purpose of the survey and gives them our contact information if they have any concerns, subjects were asked up to 15 questions. Questions 1-5 are related to the consumer’s purchasing decision: 1, 2, and 3 ask about what influenced the customer to buy insurance (insurance broker, financial advisor, online calculators, etc.); question 4 asks about the reasons for buying life insurance; whereas question 5 asks how many different price quotes the customer got before buying his or her policy.

Questions 6-9 were the main questions. In question 6, customers were asked:

*Your life insurance policy has about n years left on it. Do you plan to stop your policy (sometimes called lapsing) before then?*

The parameter $n$ was set equal to the number of years left on that customer’s policy. Only 2.8% answered said that they planned to stop their policy before its expiration, with the remaining 97.2% answering that they did not plan to stop it. In contrast, the average lapse rate on these policies in the last 15 years was 5.2% per year. Assuming a constant annual lapse rate of 5.2%, 57.5% of these policies would lapse.
Figure 7: Descriptive statistics for respondents (blue) and non-respondents (red). Bars represent 95% confidence intervals. Starting from the top, the figures represent marital status; gender; age; employer type; health status when the policy was purchased; policy term; death benefit; and whether the policy includes a waiver of premium (which allows the policyholder not to pay premiums in case of serious illness or disability) and an accelerated death benefit (which entitles the policyholder to receive cash advances against the death benefit in the case of serious illness).
Those who answered “yes” in question 6 were then asked questions 7-9. Question 7 asks: *In how many years?* Figure 8 contrasts policyholders’ plans policy cancellations and predictions assuming a constant annual lapse rate of 5.2%. As can be seen in the picture, respondents vastly underestimate the chance that they will lapse on their policies.

Those who stated that they were planning on lapsing were then asked about the possible reasons for doing so (values in brackets are the total number and the proportion of respondents who pick each option).

**Question 8: What would be the main reason for stopping your policy?**

- Insurance premiums are too costly [9; 43%]
- My needs changed, and I don’t need life insurance anymore [5; 24%]
- Unhappy with TIAA-CREF [0; 0%]
- Other (State) [7; 33%][41]

Those who selected the second option were then asked:

**Question 9: What changed about your needs?**

- My dependents are now capable of providing for themselves [2; 40%]
- My family situation changed, and so I don’t need to protect my dependents anymore [2; 40%]
- My investment opportunities have changed [1; 20%]
- I feel healthier than I expected, and so I am planning to buy a different policy [0; 0%]
- Other: (State) [0, 0%]

---

[41] Most answers in “other” involve a change in insurance needs (“I have an IRA that would provide amply for my survivors”, “Grandchildren will have finished college”, “Premiums are costly, and my 401K should have some left for family”).
These questions should be taken with caution since only 21 people answered that they were planning to lapse, so the sample is very small. However, we note that no one answered “I feel healthier than expected, and so I am planning to buy a different policy,” which is what one would expect if reclassification risk was a major issue.

Next, we ask all subjects about income fluctuations. This is important to disentangle between different reasons why people may underweight the probability of lapsing. Question 10 asks whether they had an income loss in the previous 5 years, whereas question 11 asks whether subjects expected an income loss in the next 5 years. Out of the 751 respondents, 204 (27.2%) had an income loss in the last 5 years whereas 189 (25.2%) expected an income loss in the next 5 years. Despite how prevalent income losses were, they do not seem to translate into the beliefs about lapsing. The (unconditional) correlation between expecting an income loss and expecting to lapse was 0.050, which is not statistically significant.\(^{42}\) Therefore, our results do not support an explanation based on overconfidence or optimism about future income.

In question 12, subjects who expected a future income loss were asked the possible reason for such an income loss. The most common answer, mentioned by 95 of the 189 subjects (50.3%) was a job separation by the respondent or the respondent’s spouse. The second most common answer, mentioned by 42 subjects (22.2%) was retirement, followed by fluctuations in commissions and bonuses (8 subjects, 4.23%).

Questions 13 and 14 asked the subject’s occupation and employer. Question 15 was an open question, suggested by TIAA-CREF, about other aspects that influenced them in buying life insurance.

In a linear regression framework, most variables are statistically insignificant in explaining whether respondents think they will lapse. The only robustly statistically significant variable is age, with older people reporting a higher chance of lapsing. People in healthier categories are slightly less likely to lapse, although only the “Preferred Plus” category was statistically significant at the 10% level. In addition, people who bought following the recommendation of a friend or a family member were slightly less likely to lapse, whereas those who bought life insurance to protect their charitable giving or to pay for estate taxes reported being slightly more likely to lapse.

\(^{42}\)The correlation between having had an income loss in the past 5 years and expecting to lapse was 0.006.